

60  
2e1



# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

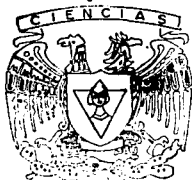
## FACULTAD DE CIENCIAS

### "ANALISIS DE REGRESION PERIODICA:

El modelo más simple y su aplicación a fenómenos  
cíclicos y a fenómenos de direcciones"

T E S I S  
Que para obtener el Título de:  
A C T U A R I O  
P r e s e n t a:

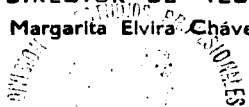
MIGUEL MACIAS RODRIGUEZ



Facultad de Ciencias  
UNAM

DIRECTOR DE TESIS:

Mat. Margarita Elvira Chávez Cano



1997  
FACULTAD DE CIENCIAS  
SECRETARÍA GENERAL

TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

**PAGINACION VARIA**

**COMPLETA LA INFORMACION**



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

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

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"ANÁLISIS DE REGRESIÓN PERIÓDICA: El modelo más simple y su  
aplicación a fenómenos cíclicos y a fenómenos de direcciones",  
realizado por MIGUEL MACÍAS RODRÍGUEZ,

con número de cuenta 8922857-3 , pasante de la carrera de ACTUARIA.

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	MAT. MARGARITA ELVIRA CHAVEZ CANO.
Propietario	DRA. ASUNCION BEGONA FERNANDEZ FERNANDEZ
Propietario	DR. JOSE RODOLFO MENDOZA BLANCO.
Suplente	ACT. CLAUDIA CARRILLO GUTIEROZ.
Suplente	MAT. VINICIO PEREZ ROMERO.

*M. Elfrán*  
*Begona Fernández*  
*Jose Rodolfo Mendoza Blanco.*  
*Claudia Carrillo G.*  
*Vinicio Perez Romero*

Consejo Departamental de Matemáticas  
CONSEJO DEPARTAMENTAL  
DE  
MATEMÁTICAS

**A la memoria de mi abuelo Benjamín...**

**A la memoria de mi tía Luz María...**

**A todos y cada uno de los miembros que integran mi familia, especialmente a mi abuelita Guadalupe y a mi Madre. Gracias...**

**A la Universidad Nacional Autónoma de México...**

**A la Facultad de Ciencias...**

**A mis maestros, con especial agradecimiento a la Profesora Margarita Elvira Chávez Cano...**

**A mis compañeros...**

**A mis amigos.**

**“El Universo se desarrolla mediante ciclos,  
las condiciones que se presentan en un momento  
dado se repetirán tarde o temprano,  
una vez que se cumpla el ciclo al  
que pertenecen...”**

**LUIS GONZÁLEZ REIMANN.**

# **ÍNDICE**

<b>INTRODUCCIÓN</b>	<b>ix</b>
<b>CAPÍTULO 1. Conceptos Fundamentales</b>	<b>3</b>
1.1 Introducción	3
1.2 Representación gráfica	4
1.3 Vector medio	7
1.4 Funciones periódicas	16
<b>CAPÍTULO 2. El Modelo más Simple de Regresión Periódica</b>	<b>21</b>
2.1 Introducción	21
2.2 Estimación por mínimos cuadrados	26
2.3 Propiedades de los estimadores por mínimos cuadrados	29
2.4 Estimación por máxima verosimilitud	38
2.5 Estimadores para $\phi_A$ y $A$	42
2.6 Intervalos de confianza	45
2.7 Coeficiente de correlación	51
2.8 Una prueba de hipótesis de gran importancia	57
2.9 Predicción	64
2.10 Inferencia simultánea sobre los parámetros del modelo	68

<b>CAPÍTULO 3 . Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones</b>	<b>73</b>
3.1 Introducción	73
3.2 Viento dominante y velocidad media	74
3.3 Viento dominante y velocidad máxima	81
3.4 Consumo de energía eléctrica a nivel nacional ( Sector Doméstico )	89
3.5 Consumo de energía eléctrica a nivel nacional ( Sector Industrial )	99
<b>CONCLUSIONES</b>	<b>109</b>
<b>BIBLIOGRAFÍA</b>	<b>111</b>
<b>APÉNDICE. Símbolos</b>	<b>115</b>



## ***INTRODUCCIÓN***

El análisis de regresión al igual que otras teorías estadísticas, juega un papel muy importante en casi todos los aspectos de la vida cotidiana. A partir de la recopilación, clasificación y análisis de una muestra de datos, se pueden realizar inferencias sobre la población involucrada en un determinado fenómeno, perteneciente a diversos campos del conocimiento como la Medicina, Biología, Economía, Política, Psicología, Agricultura, Meteorología, Física, Química o Ingeniería.

Existen fenómenos en los cuales las variables que en ellos intervienen, presentan una relación lineal. Para estos casos el ANÁLISIS DE REGRESIÓN LINEAL y el ANÁLISIS DE CORRELACIÓN LINEAL ayudan tanto a modelar el fenómeno como a realizar inferencias de la población involucrada en el mismo.

En este sentido, es importante señalar que las teorías de REGRESIÓN y CORRELACIÓN LINEALES, han sido desarrolladas y difundidas de una forma muy amplia.

Por el contrario, la parte de REGRESIÓN y CORRELACIÓN que puede aplicarse a fenómenos en donde sus variables presentan una relación periódica o circular, ha sido poco desarrollada y difundida.

El objetivo de este trabajo es presentar la teoría del modelo más simple de regresión periódica, además de mostrar cómo y a qué tipo de fenómenos se puede aplicar. Para cumplir tal objetivo se pretende por un lado, incluir resultados presentados en diversos artículos de

**ANÁLISIS DE REGRESIÓN PERIÓDICA : El Modelo más Simple y su Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

---

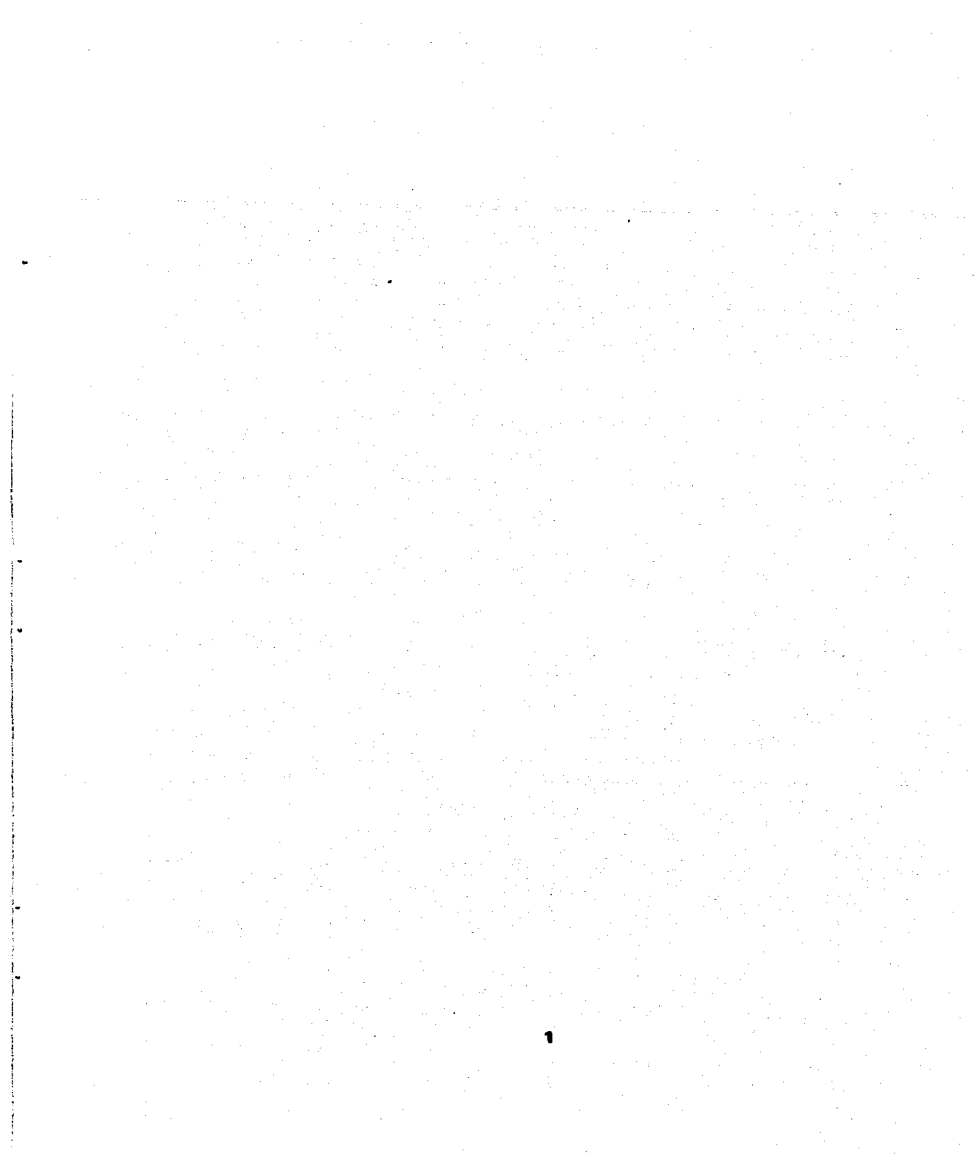
investigación y por otro, proponer algunos resultados que pudieran ser útiles para el desarrollo de esta teoría.

En el primer capítulo se presentan algunos conceptos fundamentales de la ESTADÍSTICA CIRCULAR que son necesarios para comprender la teoría de la REGRESIÓN PERIÓDICA. Es importante señalar que para este capítulo nos apoyamos en diversos artículos de investigación, así como en la bibliografía que existe al respecto.

En el segundo capítulo se presenta la teoría del modelo más simple de regresión periódica. Aquí se proponen estimadores e intervalos de confianza para los parámetros del modelo, así como pruebas de hipótesis que son útiles para validar ó rechazar el modelo en cuestión. En este capítulo también se presenta la teoría de CORRELACIÓN LINEAL CIRCULAR, desarrollada principalmente en la década de los años setentas por K.V. Mardia.

El tercer capítulo presenta cuatro aplicaciones de toda la teoría contenida en los capítulos anteriores. De los cuatro fenómenos analizados, dos son fenómenos de direcciones y los dos restantes son de los llamados fenómenos cíclicos. Aquí es importante señalar que no se pretende realizar un estudio profundo de investigación acerca de los fenómenos presentados; lo que se pretende en este capítulo es el mostrar cómo se aplica la teoría del modelo más simple de regresión periódica, a los fenómenos cíclicos y a los fenómenos de direcciones.

Finalmente el trabajo contiene algunas conclusiones derivadas de la recopilación, el desarrollo y la aplicación de los resultados presentados a lo largo de este documento.



## **CAPÍTULO 1. CONCEPTOS FUNDAMENTALES**

### **1.1 Introducción**

Un conjunto de datos que se obtiene al observar : a) direcciones en un plano o b) fenómenos cíclicos, puede ser representado por puntos distribuidos sobre un círculo. Tales observaciones son interpretadas como OBSERVACIONES CIRCULARES.

Este tipo de datos son analizados por la ESTADÍSTICA CIRCULAR y surgen en muchas aplicaciones de Biología, Medicina, Geología, Geografía, Meteorología, Paleontología, Astronomía, Física y Economía; por ejemplo en Biología existen áreas como la Orientación y Navegación Animal, donde se observan y se analizan direcciones.

En ESTADÍSTICA CIRCULAR, a cada observación se le asocia un ángulo que fluctúa entre  $0^\circ$  y  $360^\circ$  (o su equivalente entre 0 y  $2\pi$  radianes). Un claro ejemplo de la utilización de ángulos para el estudio de observaciones circulares, se presenta en el área de ritmos biológicos, ya que cualquier periodo de un evento cíclico puede ser representado por un ángulo entre  $0^\circ$  y  $360^\circ$ .

En este sentido, se considera que el ángulo es una VARIABLE CIRCULAR debido a que mide observaciones circulares; es importante diferenciar a una VARIABLE CIRCULAR de VARIABLES LINEALES tales como la longitud, peso, temperatura y voltaje, que miden observaciones que pueden ser representadas en una línea recta, es decir, miden OBSERVACIONES LINEALES.

## 1.2 Representación Gráfica

Para la representación gráfica de observaciones circulares, comúnmente se utiliza un diagrama de dispersión, en donde las observaciones son puntos alrededor de un círculo. Estos puntos tienen asignada una dirección la cual puede ser medida de la siguiente forma :

a) Se mide el ángulo formado entre el Norte y el punto que representa la observación, siempre en el sentido en que giran las manecillas del reloj. A este ángulo se le llama AZIMUTH. En la figura (1.2.1) se muestran las escalas de medición más comunes tomando en cuenta el ángulo AZIMUTH.

b) Otra forma es la que se utiliza comúnmente en matemáticas : Tomar el ángulo polar medio desde el eje de las abscisas hasta el punto que representa la observación, siempre en sentido contrario al que giran las manecillas del reloj. En la figura (1.2.2) se muestra un diagrama de dispersión tomando en cuenta el ángulo polar y un diagrama de dispersión tomando en cuenta el ángulo AZIMUTH.

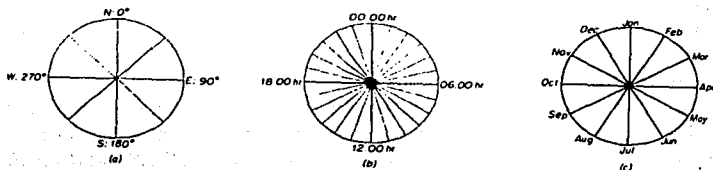


Figura 1.2.1 Escalas de medición más comunes tomando en cuenta el ángulo AZIMUTH. (a) Compás de direcciones (b) Horas del día (c) Días del año, el primero de cada mes. Tomado de Zar (1974).

## Conceptos Fundamentales.

---

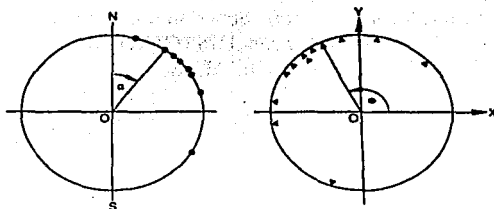


Figura 1.2.2 Diagramas de dispersión. Tomado de Batschelet (1981).

Otra forma de representar observaciones circulares es mediante el HISTOGRAMA CIRCULAR, este se utiliza para un gran número de observaciones, para “condensar” información las observaciones se agrupan. En la figura (1.2.3) se muestra un ejemplo de un HISTOGRAMA CIRCULAR.

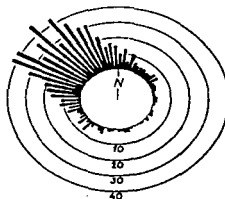


Figura 1.2.3 Histograma Circular : 714 aves fueron puestas en libertad y se desplazaron de 30 a 250 km. en todas direcciones. En el diagrama de frecuencia el intervalo de grupo es de  $5^\circ$ . Tomado de Schmidt-Koenig(1964).

## Conceptos Fundamentales.

Una variante del HISTOGRAMA CIRCULAR es cuando se registran observaciones individuales, es decir, observaciones no agrupadas. En las figuras (1.2.4) y (1.2.5) se muestran HISTOGRAMAS CIRCULARES PARA OBSERVACIONES INDIVIDUALES.

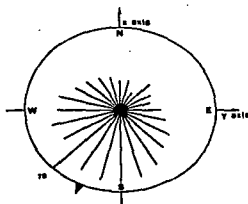


Figura 1.2.4 Histograma Circular para observaciones individuales. Tomado de Batschelet(1981).

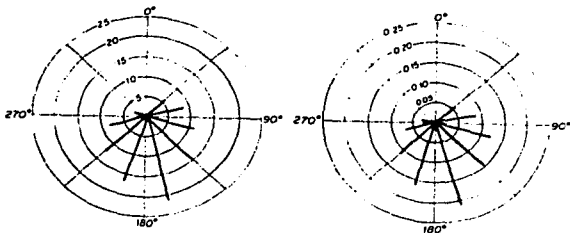


Figura 1.2.5 Histograma Circular para observaciones individuales. Tomado de Zar(1974).

Existe otra representación propuesta por Landreth y Ferguson en 1967, en la que se unen los extremos del histograma circular para observaciones individuales, como se muestra en la figura (1.2.6) :

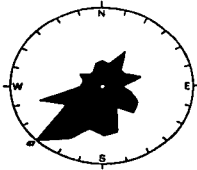


Figura 1.2.6 Representación propuesta por Landreth y Ferguson en 1967. Tomado de Batschelet(1981).

### 1.3 Vector Medio

Suponga que se tienen tres direcciones representadas por los ángulos

$$\phi_1=80^\circ, \quad \phi_2=350^\circ, \quad \phi_3=50^\circ,$$

una inquietud que surge casi inmediatamente es el conocer la dirección promedio( ÁNGULO MEDIO ). Si se elige la media aritmética se tendría que la dirección promedio es  $160^\circ$ , pero qué pasa si se reemplaza  $350^\circ$  por un ángulo equivalente,  $-10^\circ$  por ejemplo, resulta que ahora la dirección promedio es  $40^\circ$ ; de seguir haciendo este tipo de reemplazos la dirección promedio seguramente continuará cambiando. Por lo tanto, la media aritmética no resulta conveniente (más adelante se dará un ejemplo de la inconveniencia de utilizar la media aritmética de ángulos).



Para encontrar la estadística apropiada se hará lo siguiente :

Considere un diagrama de dispersión en donde el círculo sea unitario, a cada punto se le asigna una masa de igual valor,  $m$ , y se encuentra el centro de masa,  $C$ . Si este es diferente del origen,  $O$ , entonces el segmento de recta  $\overline{OC}$  define una dirección la cual se llamará DIRECCIÓN MEDIA DE LA MUESTRA.

Existen dos formas diferentes de encontrar el centro de masa, la primera utilizando álgebra vectorial y la segunda, utilizando funciones trigonométricas.

### 1. Aplicando Álgebra Vectorial.

Cada punto en el diagrama de dispersión se puede ver como un vector unitario (recuerde que se está considerando un círculo unitario para el diagrama de dispersión). Sean los vectores unitarios

$$\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$$

que representan a la muestra; entonces el vector

$$\mathbf{m} = \frac{1}{\left(\sum_{i=1}^n m_i\right)} \left(\sum_{i=1}^n m_i \vec{e}_i\right) = \frac{1}{n} \left(\sum_{i=1}^n \vec{e}_i\right) \quad (1.3.1)$$

## Conceptos Fundamentales.

coincide con el centro de masa,  $\bar{c}$ . El vector  $\mathbf{m}$  es conocido como el VECTOR MEDIO DE LA MUESTRA. En la figura (1.3.1) se muestra la construcción del centro de masa para tres y cuatro puntos.

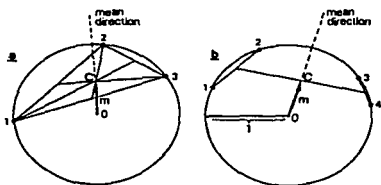


Figura 1.3.1 Construcción del Centro de Masa,  $\bar{c}$ . (a) Centro de Masa para tres puntos (b) Centro de Masa para cuatro puntos. Tomado de Batschelet(1981).

Sea  $R$  la longitud del vector resultante y  $r$  la longitud del vector medio, esto es

$$\left| \sum_{i=1}^n \bar{\mathbf{e}}_i \right| = R \quad , \quad \left| \mathbf{m} \right| = r \quad ,$$

entonces

$$r=R/n \tag{1.3.2}$$

## 2. Aplicando Funciones Trigonómicas.

Sean  $x_i$  y  $y_i$  las componentes para cada vector unitario  $\bar{\mathbf{e}}_i$  (Ver figuras (1.3.2) y (1.3.3)). Por definición de seno y coseno se tiene que

$$x_i = \text{Cos } \phi_i \quad , \quad y_i = \text{Sen } \phi_i \tag{1.3.3}$$

por otra parte, las componentes del centro de masa son

$$(\bar{x}, \bar{y}) = \left( \frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i \right) \quad (1.3.4)$$

utilizando lo anterior en la ecuación (1.3.2) se obtiene lo siguiente

$$\begin{aligned} r &= \sqrt{\bar{x}^2 + \bar{y}^2} \\ R &= \sqrt{(\sum x_i)^2 + (\sum y_i)^2} \end{aligned} \quad (1.3.5)$$

otra forma de expresar a  $r$ , utilizando las ecuaciones (1.3.3), es la siguiente

$$r = \frac{1}{n} \sqrt{(\sum \text{Cos}\phi_i)^2 + (\sum \text{Sen}\phi_i)^2} \quad (1.3.6)$$

Al vector medio se le puede asociar un ángulo que es medido desde el eje X, a éste se le conoce como el **ÁNGULO MEDIO DE LA MUESTRA** y está denotado por  $\bar{\phi}$ .

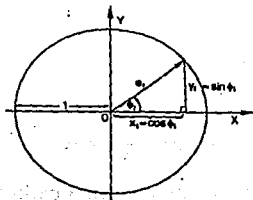


Figura 1.3.2 Componentes rectangulares de un vector unitario. Tomado de Batschelet(1981).

## Conceptos Fundamentales.

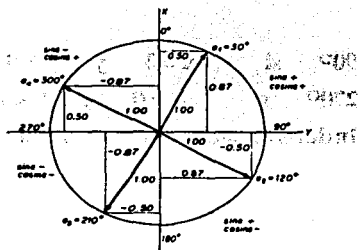


Figura 1.3.3 En un círculo unitario se muestran las coordenadas polares y rectangulares para cuatro puntos. Tomado de Zar(1974).

Para calcular  $\bar{\phi}$ , se utilizan las ecuaciones (1.3.5) y la identidad trigonométrica

$$\tan \phi = \frac{\text{sen } \phi}{\text{cos } \phi} \quad (1.3.7)$$

de esta forma se obtiene

$$\bar{\phi} = \begin{cases} \arctan \frac{\bar{y}}{\bar{x}} & \text{si } \bar{x} > 0 \\ 180^\circ + \arctan \frac{\bar{y}}{\bar{x}} & \text{si } \bar{x} < 0 \end{cases} \quad (1.3.8)$$

con los siguientes casos

$$\bar{\phi} = \begin{cases} 90^\circ & \text{si } \bar{x} = 0 \text{ y } \bar{y} > 0 \\ 270^\circ & \text{si } \bar{x} = 0 \text{ y } \bar{y} < 0 \\ \text{Indeterminado} & \text{si } \bar{x} = 0 \text{ y } \bar{y} = 0 \end{cases} \quad (1.3.9)$$

Nótese que para obtener la ecuación anterior, las coordenadas rectangulares se definen como en la figura (1.3.3). En adelante se utilizará esta forma de graficar coordenadas rectangulares.

En las figuras (1.3.4) y (1.3.5) se representan dos muestras de observaciones circulares y su respectivo **ÁNGULO MEDIO**.

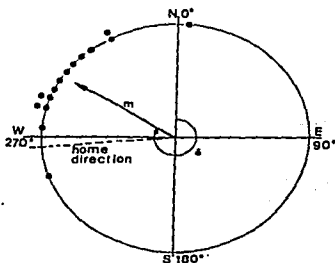
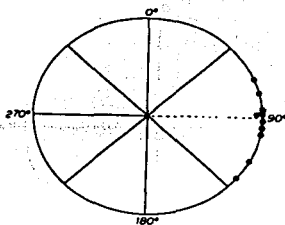


Figura 1.3.4 Experimento de HOMING en pichones en donde se muestra el vector medio y el ángulo medio. Tomado de Batschelet(1981).

## Conceptos Fundamentales.

---



*Figura 1.3.5 Diagrama de dispersión que representa a la siguiente muestra :  $66^\circ$ ,  $75^\circ$ ,  $86^\circ$ ,  $88^\circ$ ,  $88^\circ$ ,  $93^\circ$ ,  $97^\circ$ ,  $101^\circ$ ,  $118^\circ$ ,  $130^\circ$ ; en donde la línea punteada indica el **ÁNGULO MEDIO DE LA MUESTRA**,  $94^\circ$ .*

Como se mencionó anteriormente, un diagrama de dispersión no solamente se puede utilizar para representar direcciones en un plano; también se utiliza para representar cualquier fenómeno cíclico. Por ejemplo :

- La hora en que ocurre algún accidente automovilístico en determinada ciudad, durante varios días.
- La hora en que se registra un nacimiento en los hospitales de cierta zona, durante distintos días.

En los ejemplos anteriores, una hora determinada se puede asociar con una y solo una dirección. Por lo tanto, tiene sentido hablar de una **HORA MEDIA** la cual corresponde a la noción de dirección media. La hora media también es conocida como **HORA PICO** o **ACROFASE** (Ver figura (1.3.6)).

## Conceptos Fundamentales.

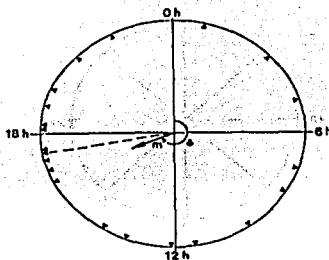


Figura 1.3.6 Hora en que ocurrió algún accidente de tránsito en una ciudad, durante distintos días. La HORA PICO o ACROFASE está determinada por el vector medio,  $m$ . Tomado de Batschelet (1981).

Un ejemplo que muestra las desventajas en el uso de la media aritmética de ángulos se muestra a continuación :

En un cruce de la Ciudad de México, se registraron 14 choques automovilísticos durante cinco semanas. Las horas en que ocurrieron los choques se presentan en el siguiente cuadro

HORA	DIRECCIÓN ASOCIADA
5:00	75°
5:20	80°
5:40	85°
6:00	90°
6:20	95°
6:40	100°

### Conceptos Fundamentales.

7:00	105°
20:00	255°
20:20	260°
20:40	265°
21:00	270°
21:20	275°
21:40	280°
22:00	285°

La media aritmética de ángulos da como dirección promedio a 180°. Esto indicaría que la hora promedio en que ocurre un choque en este cruce, es a las 12:00 hrs, lo cual es completamente falso.

Por otro lado, la longitud del vector medio,  $r$ , sirve como una medida de concentración. Esto es, cuando los puntos se concentran alrededor de alguna dirección,  $r$  es muy cercano a 1; cuando los puntos están uniformemente distribuidos la longitud del vector medio es cero.

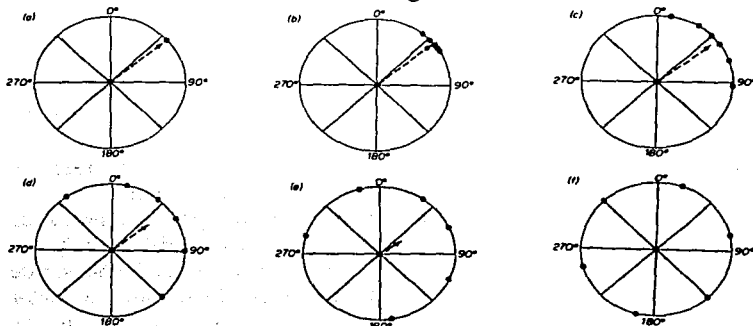


Figura 1.3.7 La longitud del vector medio,  $r$ , como medida de concentración. Tomado de Zar (1974).



## 1.4 Funciones Periódicas

Sea **P** un punto en el plano con coordenadas polares  $r$  y  $\phi$ ; entonces el Coseno y Seno de  $\phi$  están definidos de la siguiente forma

$$\begin{aligned}\text{Cos } \phi &= x \quad , \\ \text{Sen } \phi &= y \quad ,\end{aligned}\tag{1.4.1}$$

en donde  $(x,y)$  son las coordenadas rectangulares del punto **P**.

El  $\text{Sen } \phi$  y  $\text{Cos } \phi$  son funciones de  $\phi$ ; si el dominio de  $\phi$  son todos los números reales, tanto  $x$  como  $y$  son FUNCIONES PERIÓDICAS de  $\phi$ , ya que cada nueva rotación alrededor del círculo unitario genera los mismos valores de  $(x,y)$ .

Cuando se tiene un conjunto de datos pertenecientes a algún fenómeno cíclico o de direcciones, es muy común ajustarles una función periódica para modelar dicho fenómeno.

En el presente trabajo se utilizará la función

$$Y = M + A \text{Cos}[\omega(t - t_A)] \quad ,\tag{1.4.2}$$

para modelar fenómenos de direcciones y fenómenos cíclicos. A continuación se presentan los pasos para obtener esta función :

1) Se toma como variable independiente al tiempo,  $t$ , además se considera que el periodo  $T$  (intervalo de tiempo en el cual se repite un fenómeno cíclico o intervalo angular en el cual se repite un fenómeno de direcciones) es conocido. De esta forma se puede definir un factor escalar, que introduce una nueva unidad para el eje  $X$ , conocido como la

FRECUENCIA ANGULAR,  $\omega$ . Esto indica el número de veces que está contenido un periodo en una rotación completa

$$\omega = \frac{360^\circ}{T} \quad \text{ó} \quad \omega = \frac{2\pi}{T} \quad (1.4.3)$$

de esta manera se obtiene la función

$$Y = \text{Cos}(\omega t) \quad (1.4.4)$$

una gráfica de esta función se muestra en la figura (1.4.1) parte (a).

2) Suponga que la curva no alcanza un máximo en  $t=0$  sino que en  $t_A$ , donde  $0 < t_A < T$ . Con lo anterior lo único que se hace es sumar una constante,  $t_A$ , a la antigua variable

$$t' = t + t_A \quad (1.4.5)$$

de esta forma la ecuación (1.4.4) se puede expresar así

$$Y = \text{Cos} [\omega(t' - t_A)] \quad (1.4.6)$$

pero por simplicidad se expresará de la siguiente manera :

$$Y = \text{Cos} [\omega(t - t_A)] \quad (1.4.7)$$

una gráfica de esta función se muestra en la figura (1.4.1) parte (b).

## Conceptos Fundamentales.

---

3) A continuación se tiene que ajustar la AMPLITUD de la función. Se llamará AMPLITUD a la máxima desviación que tiene una oscilación desde su centro (Ver figura 1.4.1 parte (c)). Las ecuaciones de (1.4.4) y (1.4.7) tienen amplitud 1. Si se desea que la amplitud sea diferente de 1, se tiene que multiplicar a (1.4.7) por una constante positiva (esta constante es considerada como un factor escalar en el eje Y);

$$Y = A \text{ Cos } [\omega(t - t_A)] \quad (1.4.8)$$

una gráfica de esta función se muestra en la figura (1.4.1) parte (c).

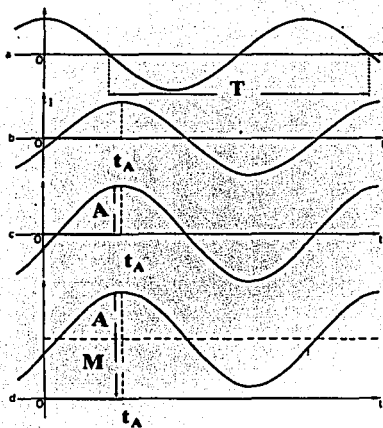
4) Finalmente, Y puede oscilar alrededor de un valor M (en las ecuaciones de (1.4.4), (1.4.7) y (1.4.8) M es igual a cero). Así se obtiene la función deseada

$$Y = M + A \text{ Cos } [\omega(t - t_A)] \quad (1.4.9)$$

una gráfica de esta función se muestra en la figura (1.4.1) parte (d).

**Conceptos Fundamentales.**

---



*Figura 1.4.1 Cuatro diferentes pasos para obtener la función  $Y = M + A \cos [\omega(t - t_A)]$ . Tomado de Batschelet (1975).*

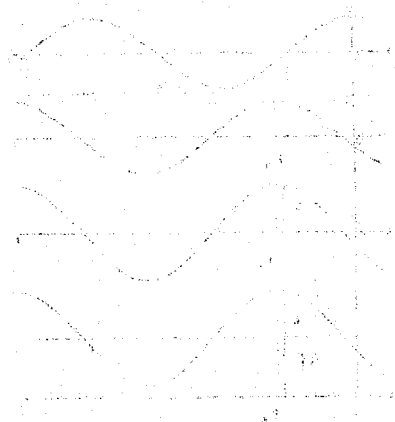
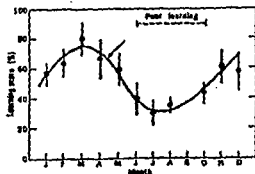


Figure 1. Comparison of the experimental results with the theoretical results. The curves are plotted for  $\omega = 100$  and  $\nu = 10$ . The solid line is the experimental result, the dashed line is the theoretical result, and the dotted line is the theoretical result for  $\nu = 1$ .

## ***CAPÍTULO 2. EL MODELO MÁS SIMPLE DE REGRESIÓN PERIÓDICA***

### **2.1 Introducción**

En el ANÁLISIS DE REGRESIÓN LINEAL, tanto la variable dependiente como la independiente son variables LINEALES (miden observaciones que pueden ser representadas en una línea recta); en este caso la variable independiente es CIRCULAR y la variable dependiente es lineal. Así, al representar un fenómeno de direcciones o un fenómeno cíclico en un plano cartesiano, se observa que una función periódica puede ser un buen modelo, de ahí el nombre de REGRESIÓN PERIÓDICA; algunos autores la llaman también REGRESIÓN LINEAL-CIRCULAR (Ver figura (2.1.1)).



*Figura 2.1.1 Cambios en la destreza que tiene la carpa dorada al nadar. La carpa muestra una gran destreza durante el invierno, mientras que en verano sucede lo contrario. Tomado de Shashoua, V.E. (1973).*

Un modelo simple para un fenómeno de direcciones o para un fenómeno cíclico puede ser el siguiente <sup>1</sup>

$$Y = M + A \text{ Cos } [\omega(t - t_A)] \quad (2.1.1)$$

en donde

- t** es la variable independiente (usualmente es el tiempo)
- M** es el nivel medio
- A** es la amplitud ( $A \geq 0$ )
- $\omega$  es la frecuencia angular
- $t_A$  es la acrofase

la frecuencia angular se puede obtener de la siguiente forma

$$\omega = \frac{360^\circ}{T} \quad \text{ó} \quad \omega = \frac{2\pi}{T} \quad (2.1.2)$$

en donde **T** es el periodo de la variable independiente, el cual se supone conocido (por ejemplo  $360^\circ$  ó 24 horas).

La frecuencia angular,  $\omega$ , indica el número de veces que se repite el periodo de la variable independiente, T, en una rotación completa. En la figura (2.1.2) se muestra la función propuesta para modelar fenómenos cíclicos o fenómenos de direcciones.

**1** Matemáticamente es equivalente utilizar  $Y = M + A \text{ sen}(\omega(t - t_A))$ .

## El Modelo más Simple de Regresión Periódica.

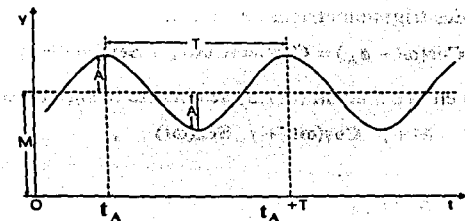


Figura 2.1.2 Gráfica de una función coseno con nivel medio  $M$ , amplitud  $A$ , acrofase  $t_A$  y periodo  $T=2\pi/\omega$ . Tomado de Batschelet, E.(1981).

Desarrollando el modelo (2.1.1) se obtiene lo siguiente

$$\begin{aligned} Y &= M + A \text{ Cos } [\omega(t - t_A)] \\ &= M + A \text{ Cos } [\omega t - \omega t_A] \\ &= M + A \text{ Cos } [\omega t - \phi_A] \end{aligned} \quad (2.1.3)$$

comúnmente  $\phi_A$  es llamado el **ÁNGULO DE LA ACROFASE** ya que es donde  $Y$  alcanza su punto más alto <sup>2</sup>

**2** En la escala del tiempo (desde 0 hasta T)  $t_A$  es la acrofase; en la correspondiente escala angular (de  $0^\circ$  a  $360^\circ$ ) la función alcanza su punto más alto en  $\phi_A$ .



### El Modelo más Simple de Regresión Periódica.

---

Por identidades trigonométricas se sabe que

$$\cos(\omega t - \phi_A) = \cos(\omega t)\cos\phi_A + \text{Sen}(\omega t)\text{Sen}\phi_A ,$$

sustituyendo en la ecuación (2.1.3) se obtiene lo siguiente :

$$Y = M + \beta \cos(\omega t) + \gamma \text{Sen}(\omega t) , \quad (2.1.4)$$

en donde

$$\begin{aligned} \beta &= A\cos\phi_A \\ \gamma &= A\text{Sen}\phi_A \end{aligned} \quad (2.1.5)$$

de las ecuaciones (2.1.5) se obtiene :

$$A = \sqrt{\beta^2 + \gamma^2} , \quad (2.1.6)$$

$$\phi_A = \begin{cases} \arctan \gamma/\beta & \text{si } \beta > 0 \\ 180^\circ + \arctan \gamma/\beta & \text{si } \beta < 0 \end{cases} , \quad (2.1.7)$$

con los siguientes casos :  $\phi_A = 90^\circ$  si  $\beta=0, \gamma > 0$ ;  $\phi_A = 270^\circ$  si  $\beta=0, \gamma < 0$ ;  
 $\phi_A$  indeterminado si  $\beta=0, \gamma=0$ .

Si se tienen  $n$  observaciones apareadas  $(t_1, Y_1) , \dots , (t_n, Y_n)$ , se puede proponer el siguiente modelo :

### El Modelo más Simple de Regresión Periódica.

---

$$\begin{aligned} Y_i &= M + \beta \cos(\omega t_i) + \gamma \operatorname{Sen}(\omega t_i) + \varepsilon_i \quad ; \quad i = 1, \dots, n \\ &= M + \beta \cos(\phi_i) + \gamma \operatorname{Sen}(\phi_i) + \varepsilon_i \quad ; \quad i = 1, \dots, n \end{aligned} \quad (2.1.8)$$

en donde la desviación de una observación  $Y_i$  de su media poblacional  $E(Y_i) = M + \beta \cos(\omega t_i) + \gamma \operatorname{Sen}(\omega t_i)$ , se toma en cuenta sumando el error aleatorio  $\varepsilon_i$ ; además se supone que las  $t_i$  son medidas sin error.

Para el modelo (2.1.8) se tienen dos tipos de hipótesis :

**Hipótesis Estructural** : El modelo es lineal en los parámetros  $M$ ,  $\beta$  y  $\gamma$ .

**Hipótesis Distribucional** : Los errores aleatorios son independientes y normalmente distribuidos; además la esperanza de los errores aleatorios es igual a cero y su varianza es constante, común y desconocida;

es decir ,

$\varepsilon_i$  se distribuye  $N(0, \sigma^2)$  ;  $i=1, \dots, n$

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

## 2.2 Estimación por mínimos cuadrados

Para el modelo más simple de regresión periódica

$$Y_i = M + \beta \text{Cos}(\omega t_i) + \gamma \text{Sen}(\omega t_i) + \varepsilon_i \quad ; \quad i = 1, \dots, n$$

se debe estimar sólo los parámetros  $M$ ,  $\beta$  y  $\gamma$ , ya que  $\omega$  se obtiene fácilmente a partir del *periodo de la variable independiente*,  $T$ , que se supone conocido. Para la estimación se utiliza el método de mínimos cuadrados.

Sean  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$  los estimadores para  $M$ ,  $\beta$  y  $\gamma$ . Se denota como

$$\hat{Y}_i = \hat{M} + \hat{\beta} \text{Cos}(\omega t_i) + \hat{\gamma} \text{Sen}(\omega t_i) \quad ; \quad i = 1, \dots, n \quad (2.2.1)$$

a la función estimada o ajustada.

Para estimar los parámetros, se debe minimizar la siguiente función

$$S = \sum_{i=1}^n \mathbf{e}_i^2 \quad (2.2.2)$$

donde

$$\mathbf{e}_i = Y_i - \hat{Y}_i \quad ; \quad i = 1, \dots, n.$$

Dado que  $\omega$  es conocida, se puede definir lo siguiente :

$$C_i = \text{Cos} \phi_i = \text{Cos} (\omega t_i) ,$$

$$S_i = \text{Sen} \phi_i = \text{Sen} (\omega t_i) , \quad (2.2.3)$$

y obtener la función

$$S = \sum_{i=1}^n (Y_i - \hat{M} - \hat{\beta}C_i - \hat{\gamma}S_i)^2 \quad , \quad (2.2.4)$$

a continuación se deriva (2.2.4) con respecto a  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$

$$\frac{\delta S}{\delta \hat{M}} = -2 \sum_{i=1}^n (Y_i - \hat{M} - \hat{\beta}C_i - \hat{\gamma}S_i) \quad , \quad (2.2.5)$$

$$\frac{\delta S}{\delta \hat{\beta}} = -2 \sum_{i=1}^n (Y_i - \hat{M} - \hat{\beta}C_i - \hat{\gamma}S_i) C_i \quad , \quad (2.2.6)$$

$$\frac{\delta S}{\delta \hat{\gamma}} = -2 \sum_{i=1}^n (Y_i - \hat{M} - \hat{\beta}C_i - \hat{\gamma}S_i) S_i \quad , \quad (2.2.7)$$

igualando a cero las derivadas parciales y desarrollando, se obtienen las ecuaciones normales

$$n\hat{M} + \hat{\beta} \sum_{i=1}^n C_i + \hat{\gamma} \sum_{i=1}^n S_i = \sum_{i=1}^n Y_i \quad ,$$

$$\hat{M} \sum_{i=1}^n C_i + \hat{\beta} \sum_{i=1}^n C_i^2 + \hat{\gamma} \sum_{i=1}^n S_i C_i = \sum_{i=1}^n Y_i C_i \quad , \quad (2.2.8)$$

$$\hat{M} \sum_{i=1}^n S_i + \hat{\beta} \sum_{i=1}^n S_i C_i + \hat{\gamma} \sum_{i=1}^n S_i^2 = \sum_{i=1}^n Y_i S_i \quad ,$$

resolviendo simultáneamente para  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$  se obtienen los estimadores de  $M$ ,  $\beta$  y  $\gamma$  :

$$\hat{\gamma} = \frac{SccSsy - ScsScy}{SccSss - (Scs)^2} \quad , \quad (2.2.9)$$

$$\hat{\beta} = \frac{SssScy - ScsSsy}{SccSss - (Scs)^2} \quad , \quad (2.2.10)$$

$$\hat{M} = \bar{Y} - \hat{\beta} \bar{C} - \hat{\gamma} \bar{S} \quad , \quad (2.2.11)$$

en donde

$$\bar{C} = \frac{\sum_{i=1}^n C_i}{n} \quad , \quad \bar{S} = \frac{\sum_{i=1}^n S_i}{n} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad ,$$

$$Scc = \sum_{i=1}^n (C_i - \bar{C})^2 \quad , \quad Sss = \sum_{i=1}^n (S_i - \bar{S})^2 \quad ,$$

$$Syy = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad , \quad Scs = \sum_{i=1}^n (C_i - \bar{C})(S_i - \bar{S}) \quad ,$$

$$Scy = \sum_{i=1}^n (C_i - \bar{C})(Y_i - \bar{Y}) \quad , \quad Ssy = \sum_{i=1}^n (S_i - \bar{S})(Y_i - \bar{Y}) \quad .$$

### 2.3 Propiedades de los estimadores por mínimos cuadrados

A continuación se obtendrá la esperanza de cada estimador :

$$\begin{aligned}
 E(\hat{\gamma}) &= E \left[ \frac{\text{SecSsy} - \text{ScsScy}}{\text{SecSss} - (\text{Scs})^2} \right] \\
 &= \frac{\text{Sec} \sum \{ (S_i - \bar{S}) E(Y_i - \bar{Y}) \} - \text{Scs} \sum \{ (C_i - \bar{C}) E(Y_i - \bar{Y}) \}}{\text{SecSss} - (\text{Scs})^2} \\
 &= \frac{\text{Sec} \sum \{ (S_i - \bar{S})(M + \beta C_i + \gamma S_i) \} - \bar{Y} \text{Sec} \sum (S_i - \bar{S})}{\text{SecSss} - (\text{Scs})^2} \\
 &\quad - \frac{\text{Scs} \sum \{ (C_i - \bar{C})(M + \beta C_i + \gamma S_i) \} - \bar{Y} \text{Scs} \sum (C_i - \bar{C})}{\text{SecSss} - (\text{Scs})^2}
 \end{aligned}$$

desarrollando se obtiene lo siguiente

$$\begin{aligned}
 E(\hat{\gamma}) &= \frac{\gamma (\text{SecSss} - (\text{Scs})^2)}{\text{SecSss} - (\text{Scs})^2} \\
 &= \gamma \tag{2.3.1}
 \end{aligned}$$

$$\begin{aligned}
 E(\hat{\beta}) &= E \left[ \frac{S_{ss}S_{cy} - S_{cs}S_{sy}}{S_{cc}S_{ss} - (S_{cs})^2} \right] \\
 &= \frac{S_{ss} \sum \{ (C_i - \bar{C}) E(Y_i - \bar{Y}) \} - S_{cs} \sum \{ (S_i - \bar{S}) E(Y_i - \bar{Y}) \}}{S_{cc}S_{ss} - (S_{cs})^2} \\
 &= \frac{S_{ss} \sum \{ (C_i - \bar{C})(M + \beta C_i + \gamma S_i) \} - \bar{Y} S_{ss} \sum (C_i - \bar{C})}{S_{cc}S_{ss} - (S_{cs})^2} \\
 &\quad - \frac{S_{cs} \sum \{ (S_i - \bar{S})(M + \beta C_i + \gamma S_i) \} - \bar{Y} S_{cs} \sum (S_i - \bar{S})}{S_{cc}S_{ss} - (S_{cs})^2}
 \end{aligned}$$

desarrollando se obtiene lo siguiente

$$\begin{aligned}
 E(\hat{\beta}) &= \frac{\beta(S_{cc}S_{ss} - (S_{cs})^2)}{S_{cc}S_{ss} - (S_{cs})^2} \\
 &= \beta \tag{2.3.2}
 \end{aligned}$$

$$E(\hat{M}) = E(\bar{Y} - \hat{\beta} \bar{C} - \hat{\gamma} \bar{S})$$

$$= \sum_{i=1}^n \frac{E(Y_i)}{n} - \beta \bar{C} - \gamma \bar{S}$$

$$= \sum_{i=1}^n \frac{M + \beta C_i + \gamma S_i}{n} - \beta \bar{C} - \gamma \bar{S}$$

$$= M$$

(2.3.3)

Por lo tanto  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$  son estimadores insesgados.

Antes de calcular la varianza de los estimadores  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$ ; se enunciará la siguiente proposición

**Proposición (2.3.1).**

$$\sum_{i=1}^n (u_i - \bar{u})(w_i - \bar{w}) = \sum_{i=1}^n (u_i - \bar{u})w_i = \sum_{i=1}^n (w_i - \bar{w})u_i$$

**Demostración .**

$$\sum_{i=1}^n (u_i - \bar{u})(w_i - \bar{w}) = \sum_{i=1}^n (u_i w_i - u_i \bar{w} - w_i \bar{u} + \bar{w} \bar{u})$$

$$= \sum_{i=1}^n u_i w_i - \frac{1}{n} \sum_{i=1}^n w_i \sum_{i=1}^n u_i - \frac{1}{n} \sum_{i=1}^n u_i \sum_{i=1}^n w_i + \frac{1}{n} \sum_{i=1}^n u_i \sum_{i=1}^n w_i$$



**El Modelo más Simple de Regresión Periódica.**

$$\begin{aligned} \sum_{i=1}^n (u_i - \bar{u})(w_i - \bar{w}) &= \sum_{i=1}^n u_i w_i - \bar{u} \sum_{i=1}^n w_i \\ &= \sum_{i=1}^n (u_i - \bar{u}) w_i \end{aligned}$$

(2.3.1)

$$\begin{aligned} \sum_{i=1}^n (u_i - \bar{u})(w_i - \bar{w}) &= \sum_{i=1}^n u_i w_i - \bar{u} \sum_{i=1}^n w_i \\ &= \sum_{i=1}^n u_i w_i - \bar{w} \sum_{i=1}^n u_i \\ &= \sum_{i=1}^n (w_i - \bar{w}) u_i \end{aligned}$$



A continuación se calculará la varianza de  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$  utilizando la proposición (2.3.1) :

$$\begin{aligned} \text{Var}(\hat{\gamma}) &= \text{Var} \left[ \frac{\text{SecSsy} - \text{ScsScy}}{\text{SecSss} - (\text{Scs})^2} \right] \\ &= \text{Var} \left[ \frac{\text{Scs} \sum (S_i - \bar{S}) Y_i - \text{Scs} \sum (C_i - \bar{C}) Y_i}{\text{SecSss} - (\text{Scs})^2} \right] \end{aligned}$$

### El Modelo más Simple de Regresión Periódica.

---

$$= \text{Var} \left[ \sum_{i=1}^n \left( \frac{(S_i - \bar{S})\text{Sc}c - (C_i - \bar{C})\text{Sc}s}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \right) Y_i \right]$$

Expresando a  $\hat{\gamma}$  como combinación lineal de las  $Y_i$ 's y con la hipótesis de que éstas son independientes se obtiene lo siguiente :

$$\text{Var}(\hat{\gamma}) = \sum_{i=1}^n \left( \frac{(S_i - \bar{S})\text{Sc}c - (C_i - \bar{C})\text{Sc}s}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \right)^2 \text{Var}[Y_i]$$

Desarrollando la igualdad anterior se obtiene la varianza de  $\hat{\gamma}$  :

$$\text{Var}(\hat{\gamma}) = \sigma^2 \frac{\text{Sc}c}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \quad (2.3.4)$$

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var} \left[ \frac{\text{S}ss\text{Sc}y - \text{Sc}s\text{S}sy}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \right] \\ &= \text{Var} \left[ \frac{\text{S}ss \sum (C_i - \bar{C}) Y_i - \text{Sc}s \sum (S_i - \bar{S}) Y_i}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \right] \end{aligned}$$

$$= \text{Var} \left[ \sum_{i=1}^n \left( \frac{(C_i - \bar{C})S_{ss} - (S_i - \bar{S})S_{cs}}{S_{cc}S_{ss} - (S_{cs})^2} \right) Y_i \right]$$

Expresando a  $\hat{\beta}$  como combinación lineal de las  $Y_i$ 's y con la hipótesis de que éstas son independientes se obtiene lo siguiente :

$$\text{Var}(\hat{\beta}) = \sum_{i=1}^n \left( \frac{(C_i - \bar{C})S_{ss} - (S_i - \bar{S})S_{cs}}{S_{cc}S_{ss} - (S_{cs})^2} \right)^2 \text{Var}(Y_i)$$

Desarrollando la igualdad anterior se obtiene la varianza de  $\hat{\beta}$  :

$$\text{Var}(\hat{\beta}) = \sigma^2 \frac{S_{ss}}{S_{cc}S_{ss} - (S_{cs})^2} \quad (2.3.5)$$

$$\text{Var}(\hat{M}) = \text{Var}(\bar{Y} - \hat{\beta}\bar{C} - \hat{\gamma}\bar{S})$$

$$= \text{Var} \left[ \sum_{i=1}^n \left( \frac{1}{n} - \frac{\bar{C}[(C_i - \bar{C})S_{ss} - (S_i - \bar{S})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} - \frac{\bar{S}[(S_i - \bar{S})S_{cc} - (C_i - \bar{C})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} \right) Y_i \right]$$

Expresando a  $\hat{M}$  como combinación lineal de las  $Y_i$ 's y con la hipótesis de que éstas son independientes se obtiene lo siguiente :

$$\text{Var}(\hat{M}) = \sum_{i=1}^n \left( \frac{1}{n} - \frac{\bar{C}[(C_i - \bar{C})S_{ss} - (S_i - \bar{S})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} - \frac{\bar{S}[(S_i - \bar{S})S_{cc} - (C_i - \bar{C})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} \right)^2 \text{Var}(Y_i)$$

Desarrollando la igualdad anterior se obtiene la varianza de  $\hat{M}$  :

$$\text{Var}(\hat{M}) = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{C}^2 S_{ss} - 2\bar{C}\bar{S}S_{cs} + \bar{S}^2 S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2} \right] \quad (2.3.6)$$

Después de haber obtenido la varianza para cada estimador, surge la siguiente pregunta :

¿Los estimadores son independientes?

Antes de dar respuesta al cuestionamiento anterior se enunciará la siguiente proposición :

**Proposición (2.3.2).** Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias independientes. Si  $U$  y  $W$  son dos combinaciones lineales de las  $Y_i$ 's, es decir

$$U = \sum_{i=1}^n u_i Y_i \quad \text{y} \quad W = \sum_{i=1}^n w_i Y_i \quad ,$$

### El Modelo más Simple de Regresión Periódica.

---

entonces

$$\text{Cov}(U, W) = \sum_{i=1}^n u_i w_i \text{Var}(Y_i)$$

#### Demostración .

$$\text{Cov}(U, W) = E[(U - E(U))(W - E(W))]$$

$$= E\left[\left(\sum u_i Y_i - \sum u_i E(Y_i)\right)\left(\sum w_i Y_i - \sum w_i E(Y_i)\right)\right]$$

$$= E\left[\left(\sum u_i (Y_i - E(Y_i))\right)\left(\sum w_i (Y_i - E(Y_i))\right)\right]$$

$$= \sum u_i w_i E\left[(Y_i - E(Y_i))^2\right]$$

$$+ \sum_{\substack{i, j \\ i \neq j}} u_i w_j E\left[(Y_i - E(Y_i))(Y_j - E(Y_j))\right]$$

$$= \sum u_i w_i \text{Var}(Y_i) + \sum_{\substack{i, j \\ i \neq j}} u_i w_j \text{Cov}(Y_i, Y_j)$$

como las  $Y_i$ 's son independientes, finalmente se obtiene la covarianza

$$\text{Cov}(U, W) = \sum u_i w_i \text{Var}(Y_i)$$



A continuación se obtendrá la covarianza de  $\hat{M}$  y  $\hat{\beta}$ . En primer lugar, se expresarán a estos estimadores como combinaciones lineales de las  $Y_i$ 's.

$$\text{Cov}(\hat{M}, \hat{\beta}) = \text{Cov} \left[ \sum_{i=1}^n \left( \frac{1}{n} \frac{\bar{C}[(C_i - \bar{C})S_{ss} - (S_i - \bar{S})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} - \frac{\bar{S}[(S_i - \bar{S})S_{cc} - (C_i - \bar{C})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} \right) Y_i, \sum_{i=1}^n \left( \frac{(C_i - \bar{C})S_{ss} - (S_i - \bar{S})S_{cs}}{S_{cc}S_{ss} - (S_{cs})^2} \right) Y_i \right]$$

Usando la proposición (2.3.2) y desarrollando se obtiene la covarianza de  $\hat{M}$  y  $\hat{\beta}$ .

$$\text{Cov}(\hat{M}, \hat{\beta}) = \sigma^2 \left( - \frac{\bar{C}S_{ss} - \bar{S}S_{cs}}{S_{cc}S_{ss} - (S_{cs})^2} \right) \quad (2.3.7)$$

Por lo tanto  $\hat{M}$  y  $\hat{\beta}$  no son en general independientes.

De forma análoga se obtiene la covarianza de  $\hat{M}$  y  $\hat{\gamma}$  :

$$\text{Cov}(\hat{M}, \hat{\gamma}) = \text{Cov} \left[ \sum_{i=1}^n \left( \frac{1}{n} \frac{\bar{C}[(C_i - \bar{C})S_{ss} - (S_i - \bar{S})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} - \frac{\bar{S}[(S_i - \bar{S})S_{cc} - (C_i - \bar{C})S_{cs}]}{S_{cc}S_{ss} - (S_{cs})^2} \right) Y_i, \sum_{i=1}^n \left( \frac{(S_i - \bar{S})S_{cc} - (C_i - \bar{C})S_{cs}}{S_{cc}S_{ss} - (S_{cs})^2} \right) Y_i \right]$$

Usando la proposición (2.3.2) y desarrollando se obtiene la covarianza de  $\hat{M}$  y  $\hat{\gamma}$ .

$$\text{Cov}(\hat{M}, \hat{\gamma}) = \sigma^2 \left( -\frac{\bar{S}\text{Sc}c - \bar{C}\text{Sc}s}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \right) \quad (2.3.8)$$

Por lo tanto  $\hat{M}$  y  $\hat{\gamma}$  no son en general independientes.

De forma análoga se obtiene la covarianza de  $\hat{\beta}$  y  $\hat{\gamma}$ :

$$\text{Cov}(\hat{\beta}, \hat{\gamma}) = \text{Cov} \left[ \sum_{i=1}^n \left( \frac{(C_i - \bar{C})\text{Sc}ss - (S_i - \bar{S})\text{Sc}s}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \right) Y_i, \sum_{i=1}^n \left( \frac{(S_i - \bar{S})\text{Sc}c - (C_i - \bar{C})\text{Sc}s}{\text{Sc}c\text{S}ss - (\text{Sc}s)^2} \right) Y_i \right]$$

Usando la proposición (2.3.2) y desarrollando se obtiene la covarianza de  $\hat{\beta}$  y  $\hat{\gamma}$ .

$$\text{Cov}(\hat{\beta}, \hat{\gamma}) = \sigma^2 (-\text{Sc}s) \quad (2.3.9)$$

Por lo tanto  $\hat{\beta}$  y  $\hat{\gamma}$  no son en general independientes.

## 2.4 Estimación por máxima verosimilitud

El método de estimación por máxima verosimilitud selecciona como estimador, a aquel valor del parámetro que tiene la propiedad de maximizar el valor de la probabilidad de la muestra aleatoria observada.

Los errores  $\varepsilon_i$  en el modelo

$$Y_i = M + \beta C_i + \gamma S_i + \varepsilon_i \quad ; \quad i = 1, \dots, n \quad (2.4.1)$$

tienen una distribución Normal, por lo tanto

$$Y_i \text{ se distribuye } N\left(M + \beta C_i + \gamma S_i, \sigma^2\right) \quad ; \quad i = 1, \dots, n \quad (2.4.2)$$

En este caso la función de verosimilitud es la siguiente :

$$\begin{aligned} L(Y_1, \dots, Y_n; M, \beta, \gamma, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y_i - M - \beta C_i - \gamma S_i)^2} \\ &= (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum (Y_i - M - \beta C_i - \gamma S_i)^2} \end{aligned} \quad (2.4.3)$$

como el logaritmo natural es una función monótona creciente, se pueden obtener los estimadores máximo verosímiles al maximizar la siguiente función :

$$\begin{aligned} \ln L(Y_1, \dots, Y_n; M, \beta, \gamma, \sigma^2) \\ = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (Y_i - M - \beta C_i - \gamma S_i)^2 \end{aligned} \quad (2.4.4)$$

a continuación se deriva con respecto a  $M, \beta, \gamma$  y  $\sigma^2$



a continuación se deriva con respecto a  $M$ ,  $\beta$ ,  $\gamma$  y  $\sigma^2$

$$\frac{\delta \mathcal{L}}{\delta M} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - M - \beta C_i - \gamma S_i) \quad ,$$

$$\frac{\delta \mathcal{L}}{\delta \beta} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - M - \beta C_i - \gamma S_i) C_i \quad , \quad (2.4.5)$$

$$\frac{\delta \mathcal{L}}{\delta \gamma} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - M - \beta C_i - \gamma S_i) S_i \quad ,$$

$$\frac{\delta \mathcal{L}}{\delta \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (Y_i - M - \beta C_i - \gamma S_i)^2 \quad ,$$

igualando a cero las derivadas parciales y desarrollando, se obtiene lo siguiente :

$$nM + \beta \sum_{i=1}^n C_i + \gamma \sum_{i=1}^n S_i = \sum_{i=1}^n Y_i \quad , \quad (2.4.6)$$

$$M \sum_{i=1}^n C_i + \beta \sum_{i=1}^n C_i^2 + \gamma \sum_{i=1}^n S_i C_i = \sum_{i=1}^n Y_i C_i \quad , \quad (2.4.7)$$

$$M \sum_{i=1}^n S_i + \beta \sum_{i=1}^n S_i C_i + \gamma \sum_{i=1}^n S_i^2 = \sum_{i=1}^n Y_i S_i \quad , \quad (2.4.8)$$

$$\frac{1}{n} \sum_{i=1}^n (Y_i - M - \beta C_i - \gamma S_i)^2 = \sigma^2 \quad (2.4.9)$$

resolviendo simultáneamente para  $M$ ,  $\beta$ ,  $\gamma$  y  $\sigma^2$  se obtienen los estimadores de estos parámetros

$$\hat{\gamma} = \frac{SccSsy - ScsScy}{SccSss - (Scs)^2}$$

$$\hat{\beta} = \frac{SssScy - ScsSsy}{SccSss - (Scs)^2}$$

$$\hat{M} = \bar{Y} - \hat{\beta} \bar{C} - \hat{\gamma} \bar{S}$$

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

En la sección anterior se mostró que los estimadores para  $M$ ,  $\beta$  y  $\gamma$  son estimadores insesgados. Esto no sucede con  $\hat{\sigma}^2$  ya que

$$E(\hat{\sigma}^2) = \frac{n-3}{n} \sigma^2$$

por lo tanto un estimador insesgado para  $\sigma^2$  es el siguiente :

$$\hat{\sigma}^2_{ins} = \frac{1}{n-3} \sum_{i=1}^n e_i^2 \quad (2.4.10)$$

## 2.5 Estimadores para $\phi_A$ y A

Al principio de este capítulo se propuso un modelo simple para fenómenos de direcciones o para fenómenos cíclicos

$$Y_i = M + A \cos [\omega t_i - \phi_A] + \varepsilon_i \quad ; \quad i = 1, \dots, n$$

posteriormente este modelo se desarrolló de tal forma que se convirtiera en un modelo lineal en los parámetros contenidos

$$Y_i = M + \beta C_i + \gamma S_i + \varepsilon_i \quad ; \quad i = 1, \dots, n$$

Hasta el momento se han encontrado estimadores para los parámetros  $M$ ,  $\beta$  y  $\gamma$  por dos métodos distintos, mínimos cuadrados y máxima verosimilitud, en ambos casos se obtuvo lo siguiente :

$$\hat{\gamma} = \frac{SccSsy - ScsScy}{SccSss - (Scs)^2}$$

$$\hat{\beta} = \frac{SssScy - ScsSsy}{SccSss - (Scs)^2}$$

$$\hat{M} = \bar{Y} - \hat{\beta} \bar{C} - \hat{\gamma} \bar{S}$$

Por el Teorema de Gauss-Marcov se sabe que  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$  son estimadores lineales insesgados de mínima varianza; además son también estimadores máximo verosímiles, por lo tanto, son los mejores

### El Modelo más Simple de Regresión Periódica.

---

estimadores insesgados de sus respectivos parámetros (mejores en el sentido de mínima varianza). Con esto se puede construir un modelo confiable que represente a un fenómeno de direcciones o a un fenómeno cíclico determinado, entonces ¿Por qué encontrar un estimador para la amplitud?, ¿Por qué encontrar un estimador para el ángulo de la acrofase?, ¿Qué importancia tendría el contar con estos estimadores?; el objetivo de esta sección es precisamente dar respuesta a los cuestionamientos anteriores.

Hasta el momento, se han estimado tres parámetros que permiten ajustar una función periódica a un conjunto de datos pertenecientes a un fenómeno de direcciones o a un fenómeno cíclico; esto es ya un avance en nuestro trabajo, pero creemos que aún se pueden aportar estimadores puntuales para el ángulo de la acrofase y para la amplitud (máxima desviación que tiene una oscilación desde el nivel medio). Por tal motivo a continuación se proponen estimadores para ambos parámetros.

Si se considera que tanto  $\hat{\beta}$  como  $\hat{\gamma}$  son estimadores máximo verosímiles y que en la sección 2.1 se obtuvo la igualdad (2.1.7), entonces el principio de invarianza para estimadores máximo verosímiles (Ver Mood, Alexander & Graybill, Franklin (1974) sección VII-2) se puede aplicar para proponer el siguiente estimador máximo verosímil para el ángulo de la acrofase :

$$\hat{\phi}_A = \begin{cases} \arctan \frac{\hat{\gamma}}{\hat{\beta}} & \text{si } \hat{\beta} > 0 \\ 180^\circ + \arctan \frac{\hat{\gamma}}{\hat{\beta}} & \text{si } \hat{\beta} < 0 \end{cases} \quad (2.5.1)$$

con los siguientes casos

$$\hat{\phi}_A = \begin{cases} 90^\circ & \text{si } \hat{\beta} = 0 \text{ y } \hat{\gamma} > 0 \\ 270^\circ & \text{si } \hat{\beta} = 0 \text{ y } \hat{\gamma} < 0 \\ \text{indeterminado} & \text{si } \hat{\beta} = 0 \text{ y } \hat{\gamma} = 0 \end{cases}$$

De forma análoga (sólo que ahora utilizando la igualdad (2.1.6)), se propone el siguiente estimador máximo verosímil para la amplitud :

$$\hat{A} = \sqrt{(\hat{\beta})^2 + (\hat{\gamma})^2} \quad (2.5.2)$$

otro estimador para la amplitud puede ser el que a continuación se propone :

$$\hat{A} = \sqrt{(\hat{\beta})^2 + (\hat{\gamma})^2 - \hat{\sigma}^2_{ins} \left[ \frac{S_{ss} + S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2} \right]} \quad (2.5.3)$$

en donde

$$E \left( (\hat{\beta})^2 + (\hat{\gamma})^2 - \hat{\sigma}^2_{ins} \left[ \frac{S_{ss} + S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2} \right] \right) = A^2$$

Por último, se harán los siguientes señalamientos para concluir esta sección :

1. El contar con los estimadores  $\hat{\phi}_A$ ,  $\hat{A}$  y  $\hat{\hat{A}}$  permitirá hacer más inferencias sobre la población involucrada en un fenómeno de direcciones o en un fenómeno cíclico.
2. No se puede asegurar que  $\hat{\phi}_A$ ,  $\hat{A}$  y  $\hat{\hat{A}}$  sean estimadores insesgados de mínima varianza.
3. Los estimadores  $\hat{\phi}_A$ ,  $\hat{A}$  y  $\hat{\hat{A}}$  tienen algunas propiedades óptimas que heredan del método de máxima verosimilitud, por ejemplo la propiedad *BAN* (Ver Mood, Alexander & Graybill, Franklin (1974) sección VII-9).

## **2.6 Intervalos de confianza**

En la sección 2.2, se obtuvieron estimadores puntuales para los parámetros del modelo más simple de regresión periódica

$$Y_i = M + \beta C_i + \gamma S_i + \varepsilon_i \quad ; \quad i = 1, \dots, n$$

en esta sección se obtendrán intervalos de confianza para estos mismos parámetros, así como un intervalo de confianza para  $\sigma^2$ .

### Intervalo de confianza para M

Se sabe que

$$\hat{M} \text{ se distribuye } N\left(M, \sigma^2 \left[ \frac{1}{n} + \frac{\bar{C}^2 S_{ss} - 2\bar{C}\bar{S}S_{cs} + \bar{S}^2 S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2} \right]\right)$$

⇒

$$\frac{\hat{M} - M}{\sigma \sqrt{\frac{1}{n} + \frac{\bar{C}^2 S_{ss} - 2\bar{C}\bar{S}S_{cs} + \bar{S}^2 S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2}}} \text{ se distribuye } N(0,1)$$

además

$$\frac{1}{\sigma^2} \sum_{i=1}^n e_i^2 \text{ se distribuye } \chi^2_{(n-3)} \text{ y es independiente de } \hat{M}$$

⇒

$$\frac{\hat{M} - M}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{C}^2 S_{ss} - 2\bar{C}\bar{S}S_{cs} + \bar{S}^2 S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2}}} \text{ se distribuye } t_{(n-3)}$$

en donde  $\hat{\sigma} = \sqrt{\frac{1}{n-3} \sum_{i=1}^n e_i^2}$

ahora se encuentra el intervalo de confianza :

$$P \left[ -t_{(1-\frac{\alpha}{2}, n-3)} < \frac{\hat{M} - M}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{C}^2 S_{ss} - 2\bar{C}\bar{S}S_{cs} + \bar{S}^2 S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2}}} < t_{(1-\frac{\alpha}{2}, n-3)} \right] = 1 - \alpha$$

donde  $-\mathbf{t}_{(1-\frac{\alpha}{2}, n-3)}$  y  $\mathbf{t}_{(1-\frac{\alpha}{2}, n-3)}$  son los cuantiles  $\frac{\alpha}{2}$  y  $1-\frac{\alpha}{2}$ , respectivamente, de una distribución  $\mathbf{t}$  con  $(n-3)$  grados de libertad. Por lo tanto un intervalo del  $(1-\alpha) \times 100\%$  de confianza para  $M$  es

$$\hat{M} \pm \mathbf{t}_{(1-\frac{\alpha}{2}, n-3)} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{C^2 S_{ss} - 2CS_{cs} + S^2 S_{cc}}{S_{cc} S_{ss} - (S_{cs})^2}} \quad (2.6.1)$$

### Intervalo de confianza para $\beta$

Se sabe que

$$\hat{\beta} \text{ se distribuye } N\left(\beta, \sigma^2 \left[ \frac{S_{ss}}{S_{cc} S_{ss} - (S_{cs})^2} \right] \right)$$

$\Rightarrow$

$$\frac{\hat{\beta} - \beta}{\sigma \sqrt{\frac{S_{ss}}{S_{cc} S_{ss} - (S_{cs})^2}}} \text{ se distribuye } N(0,1)$$

además

$$\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{e}_i^2 \text{ se distribuye } \chi_{(n-3)}^2 \text{ y es independiente de } \hat{\beta}$$

$\Rightarrow$

$$\frac{\hat{\beta} - \beta}{\hat{\sigma} \sqrt{\frac{S_{ss}}{S_{cc} S_{ss} - (S_{cs})^2}}} \text{ se distribuye } \mathbf{t}_{(n-3)}$$



en donde  $\hat{\sigma} = \sqrt{\frac{1}{n-3} \sum_{i=1}^n e_i^2}$

ahora se encuentra el intervalo de confianza :

$$P \left[ -t_{(1-\frac{\alpha}{2}, n-3)} < \frac{\hat{\beta} - \beta}{\hat{\sigma} \sqrt{\frac{S_{ss}}{S_{cc}S_{ss} - (S_{cs})^2}}} < t_{(1-\frac{\alpha}{2}, n-3)} \right] = 1 - \alpha$$

donde  $-t_{(1-\frac{\alpha}{2}, n-3)}$  y  $t_{(1-\frac{\alpha}{2}, n-3)}$  son los cuantiles  $\frac{\alpha}{2}$  y  $1-\frac{\alpha}{2}$ ,

respectivamente, de una distribución  $t$  con  $(n-3)$  grados de libertad.

Por lo tanto un intervalo del  $(1-\alpha) \times 100\%$  de confianza para  $\beta$  es

$$\hat{\beta} \pm t_{(1-\frac{\alpha}{2}, n-3)} \hat{\sigma} \sqrt{\frac{S_{ss}}{S_{cc}S_{ss} - (S_{cs})^2}} \quad (2.6.2)$$

### Intervalo de confianza para $\gamma$

Se sabe que

$$\hat{\gamma} \text{ se distribuye } N\left(\gamma, \sigma^2 \left[ \frac{S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2} \right] \right)$$

$\Rightarrow$

$$\frac{\hat{\gamma} - \gamma}{\sigma \sqrt{\frac{S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2}}} \text{ se distribuye } N(0,1)$$

además

$\Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^n e_i^2$  se distribuye  $\chi^2_{(n-3)}$  y es independiente de  $\hat{\gamma}$

$\Rightarrow$

$\frac{\hat{\gamma} - \gamma}{\hat{\sigma} \sqrt{\frac{S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2}}}$  se distribuye  $t_{(n-3)}$

en donde  $\hat{\sigma} = \sqrt{\frac{1}{n-3} \sum_{i=1}^n e_i^2}$

ahora se encuentra el intervalo de confianza :

$$P \left[ -t_{(1-\frac{\alpha}{2}, n-3)} < \frac{\hat{\gamma} - \gamma}{\hat{\sigma} \sqrt{\frac{S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2}}} < t_{(1-\frac{\alpha}{2}, n-3)} \right] = 1 - \alpha$$

donde  $-t_{(1-\frac{\alpha}{2}, n-3)}$  y  $t_{(1-\frac{\alpha}{2}, n-3)}$  son los cuantiles  $\frac{\alpha}{2}$  y  $1 - \frac{\alpha}{2}$ , respectivamente, de una distribución  $t$  con  $(n-3)$  grados de libertad.

Por lo tanto un intervalo del  $(1-\alpha) \times 100\%$  de confianza para  $\gamma$  es

$$\hat{\gamma} \pm t_{(1-\frac{\alpha}{2}, n-3)} \hat{\sigma} \sqrt{\frac{S_{cc}}{S_{cc}S_{ss} - (S_{cs})^2}} \quad (2.6.3)$$

### Intervalo de confianza para $\sigma^2$

Se sabe que

$$\frac{1}{\sigma^2} \sum_{i=1}^n e_i^2 \text{ se distribuye } \chi^2_{(n-3)}$$

en base a esto se puede construir un intervalo de confianza para  $\sigma^2$

$$P \left[ \chi^2_{\left(\frac{\alpha}{2}, n-3\right)} < \frac{1}{\sigma^2} \sum_{i=1}^n e_i^2 < \chi^2_{\left(1-\frac{\alpha}{2}, n-3\right)} \right] = 1 - \alpha$$

donde  $\chi^2_{\left(\frac{\alpha}{2}, n-3\right)}$  y  $\chi^2_{\left(1-\frac{\alpha}{2}, n-3\right)}$  son los cuantiles  $\frac{\alpha}{2}$  y  $1-\frac{\alpha}{2}$ , respectivamente, de una distribución  $\chi^2$  con  $(n-3)$  grados de libertad.

Por lo tanto un intervalo del  $(1-\alpha) \times 100\%$  de confianza para  $\sigma^2$  es

$$\left( \frac{\sum_{i=1}^n e_i^2}{\chi^2_{\left(1-\frac{\alpha}{2}, n-3\right)}}, \frac{\sum_{i=1}^n e_i^2}{\chi^2_{\left(\frac{\alpha}{2}, n-3\right)}} \right) \quad (2.6.4)$$

## 2.7 Coeficiente de correlación lineal-circular

El ANÁLISIS DE CORRELACIÓN LINEAL CIRCULAR es uno de los resultados más recientes que se han obtenido en la ESTADÍSTICA CIRCULAR, las primeras investigaciones fueron hechas en la década de los setentas y aunque esta teoría no ha tenido un gran desarrollo, los pocos estudios que existen son muy valiosos.

En 1976 K.V. Mardia propone un coeficiente de correlación lineal circular como una medida de asociación entre dos variables aleatorias  $Y$  y  $\Phi$ , donde  $Y$  es una variable lineal definida en  $(-\infty, \infty)$  y  $\Phi$  es una variable circular definida en  $[0, 2\pi)$ .

Este coeficiente se construye utilizando la teoría del ANÁLISIS MULTIVARIADO (Ver Srivastava and Carter (1983), sección 9)

correlacionando  $Y$  con  $\bar{X} = \begin{pmatrix} \text{Cos } \Phi \\ \text{Sen } \Phi \end{pmatrix}$ ; esto se define como la máxima

correlación entre  $Y$  y  $\lambda' \bar{X}$ , una combinación lineal de  $\text{Cos } \Phi$  y  $\text{Sen } \Phi$ .

$$\text{Sea} \quad G = \begin{pmatrix} Y \\ \bar{X} \end{pmatrix} \quad \text{Cov}(G) = \begin{pmatrix} \sigma^2 & \bar{\sigma}_{12}' \\ \bar{\sigma}_{12} & \Sigma_{22} \end{pmatrix} \quad (2.7.1)$$

en donde  $\text{Var}(Y) = \sigma^2$ ,  $\text{Cov}(\bar{X}) = \Sigma_{22}$  y  $\text{Cov}(Y, \bar{X}) = \bar{\sigma}_{12}$ .

Por lo tanto, el coeficiente de correlación entre  $Y$  y  $\lambda' \bar{X}$  es

$$\rho_{Y, \Phi} = \frac{\lambda' \bar{\sigma}_{12}}{\sqrt{\sigma^2 \lambda' \Sigma_{22} \lambda}} \quad (2.7.2)$$

ya que  $\text{Cov}(Y, \lambda' \bar{X}) = \lambda' \bar{\sigma}_{12}$  y  $\text{Var}(\lambda' \bar{X}) = \lambda' \Sigma_{22} \lambda$ .

El máximo valor (Ver Srivastava and Carter (1983), sección 9.5.1) está dado por :

$$\rho_{Y,\Phi} = \frac{\sqrt{\bar{\sigma}_{12}' \Sigma_{22}^{-1} \bar{\sigma}_{12}}}{\sqrt{\sigma^2}} \quad (2.7.3)$$

desarrollando se tiene lo siguiente :

$$\rho_{Y,\Phi} = \left( \frac{\text{Var}(\text{Cos}\Phi) \{ \text{Cov}(Y, \text{Sen}\Phi) \}^2 + \text{Var}(\text{Sen}\Phi) \{ \text{Cov}(Y, \text{Cos}\Phi) \}^2}{\sigma^2 \{ \text{Var}(\text{Sen}\Phi) \text{Var}(\text{Cos}\Phi) - \{ \text{Cov}(\text{Cos}\Phi, \text{Sen}\Phi) \}^2 \}} \right)^{\frac{1}{2}} - \frac{\text{Cov}(\text{Cos}\Phi, \text{Sen}\Phi) \{ \{ \text{Cov}(Y, \text{Sen}\Phi) \}^2 + \{ \text{Cov}(Y, \text{Cos}\Phi) \}^2 \}}{\sigma^2 \{ \text{Var}(\text{Sen}\Phi) \text{Var}(\text{Cos}\Phi) - \{ \text{Cov}(\text{Cos}\Phi, \text{Sen}\Phi) \}^2 \}} \quad (2.7.4)$$

si se supone que  $E(Y) = E(\text{Sen}\Phi) = 0$ , se pueden obtener las siguientes igualdades

$$E^2(Y \text{Sen}\Phi) = \{ \text{Cov}(Y, \text{Sen}\Phi) \}^2 \quad (2.7.5)$$

$$E^2(Y \text{Cos}\Phi) = \{ \text{Cov}(Y, \text{Cos}\Phi) \}^2 \quad (2.7.6)$$

$$E^2(\text{Sen}2\Phi) = 4 \{ \text{Cov}(\text{Cos}\Phi, \text{Sen}\Phi) \}^2 \quad (2.7.7)$$

$$E(\text{Sen}2\Phi)E(Y\text{Sen}\Phi)E(Y\text{Cos}\Phi) = \text{Cov}(\text{Cos}\Phi, \text{Sen}\Phi) \left\{ \left\{ \text{Cov}(Y, \text{Sen}\Phi) \right\}^2 + \left\{ \text{Cov}(Y, \text{Cos}\Phi) \right\}^2 \right\} \quad (2.7.8)$$

sustituyendo lo anterior en la ecuación (2.7.4) y desarrollando, se obtiene el cuadrado del coeficiente de correlación poblacional propuesto por K.V. Mardia :

$$\rho_{Y,\Phi}^2 = \frac{4(V_C^2 E_{Ys}^2 + V_S^2 E_{Yc}^2 - \eta_2 E_{Yc} E_{Ys})}{\sigma^2 (4V_C^2 V_S^2 - \eta_2^2)} \quad (2.7.9)$$

en donde

$$\begin{array}{ll} E(Y) = 0 & E(\text{Sen}\Phi) = 0 \\ E(Y\text{Sen}\Phi) = E_{Ys} & E(Y\text{Cos}\Phi) = E_{Yc} \\ \text{Var}(\text{Cos}\Phi) = V_C^2 & \text{Var}(\text{Sen}\Phi) = V_S^2 \\ \text{Var}(Y) = \sigma^2 & \eta_2 = E(\text{Sen}2\Phi) \end{array}$$

Cabe señalar que (2.7.9) es invariante bajo cambio de escala y de origen para Y, así como bajo cambio de dirección cero para  $\Phi$ ; obviamente (2.7.9) es igual a cero si y solo si Y es independiente de  $\Phi$  (Ver Jupp, P.E. and Mardia, K.V. (1980)).

De forma análoga se obtiene el coeficiente de correlación lineal circular muestral propuesto por K.V. Mardia :

Sea  $S$  la matriz de covarianza muestral  $S = \begin{pmatrix} s & \bar{s}_{12}' \\ \bar{s}_{12} & S_{22} \end{pmatrix}$

en donde

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

$$S_{22} = \begin{pmatrix} CC & CS \\ CS & SS \end{pmatrix}$$

$$CC = \frac{\sum_{i=1}^n \left( \text{Cos}\phi_i - \frac{\sum_{i=1}^n \text{Cos}\phi_i}{n} \right)^2}{n}$$

$$SS = \frac{\sum_{i=1}^n \left( \text{Sen}\phi_i - \frac{\sum_{i=1}^n \text{Sen}\phi_i}{n} \right)^2}{n}$$

$$CS = \frac{\sum_{i=1}^n \left( \text{Cos}\phi_i - \frac{\sum_{i=1}^n \text{Cos}\phi_i}{n} \right) \left( \text{Sen}\phi_i - \frac{\sum_{i=1}^n \text{Sen}\phi_i}{n} \right)}{n}$$

**El Modelo más Simple de Regresión Periódica.**

$$\bar{s}_{12} = \frac{\sum_{i=1}^n (Y_i - \bar{Y}) \left( \text{Cos}\phi_i - \frac{\sum_{i=1}^n \text{Cos}\phi_i}{n} \right)}{\sum_{i=1}^n (Y_i - \bar{Y}) \left( \text{Sen}\phi_i - \frac{\sum_{i=1}^n \text{Sen}\phi_i}{n} \right)}$$

entonces el coeficiente de correlación está definido de la siguiente forma (Ver Srivastava and Carter (1983), sección 9.5) :

$$R_{y,\phi} = \sqrt{\frac{\bar{s}_{12}' S_{22}^{-1} \bar{s}_{12}}{s}} \quad (2.7.10)$$

desarrollando la ecuación (2.7.10) se obtiene el cuadrado del coeficiente de correlación muestral propuesto por K.V. Mardia :

$$R_{y,\phi}^2 = \frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{23}^2} \quad (2.7.11)$$

en donde

$$r_{12} = \text{Corr}(Y, \text{Cos}\phi), \quad r_{13} = \text{Corr}(Y, \text{Sen}\phi), \quad r_{23} = \text{Corr}(\text{Cos}\phi, \text{Sen}\phi)$$

son coeficientes de correlación simples.

Debido a que  $R_{y,\phi}$  se construye utilizando la teoría del ANÁLISIS MULTIVARIADO, se puede decir que  $R_{y,\phi}^2$  es un coeficiente de



**determinación** (Ver Mardia, K.V. et al(1982) y Anderson, T.W. (1984)) es decir,  $R_{y,\phi}^2$  es la proporción de la suma de cuadrados de  $Y$  atribuible a la información obtenida de  $(\text{Cos}\phi, \text{Sen}\phi)$ .

Si  $R_{y,\phi}^2 = 1$ , se puede asegurar que todos los puntos están sobre la función ajustada; si  $R_{y,\phi}^2 = 0$ , la función ajustada es precisamente la recta  $\bar{Y}$  y por lo tanto no hay variación explicada por la regresión periódica.

Bajo la hipótesis de normalidad, la distribución de  $R_{y,\phi}^2$  cuando  $\rho_{Y,\phi} = 0$  es similar al coeficiente de correlación ordinario :

Si  $G = \begin{pmatrix} Y \\ \bar{X} \end{pmatrix} = \begin{pmatrix} Y \\ (\text{Cos}\Phi) \\ (\text{Sen}\Phi) \end{pmatrix}$  se distribuye  $N(\bar{\mu}, \Sigma)$

y además  $\rho_{Y,\phi} = 0$ , se puede obtener lo siguiente (Ver Liddell, I.G. and Ord, J.K. (1978)) :

$$U = \frac{\frac{n-3}{2} R_{y,\phi}^2}{1 - R_{y,\phi}^2} \quad \text{se distribuye} \quad F_{(2, n-3)} \quad (2.7.12)$$

por lo tanto, al realizar la prueba de hipótesis

$$\text{Ho: } \rho_{Y,\phi} = 0 \quad \text{vs} \quad \text{Ha: } \rho_{Y,\phi} \neq 0$$

se pueden utilizar tablas de la distribución  $F$ .

## 2.8 Una prueba de hipótesis de gran importancia

En lo concerniente a las pruebas de hipótesis, es de gran importancia la siguiente

$$H_0: \beta = \gamma = 0 \quad \text{vs}$$

**Ha:** Al menos uno de estos dos parámetros, es diferente de cero.

La hipótesis nula dice que la variable dependiente **Y**, no muestra ni incremento ni decremento cuando las variables independientes cambian .

Para realizar esta prueba, se aplica un caso especial de la prueba de hipótesis lineal general (Ver Montgomery, D.C. and Peck, E.A.(1992) sección 4.4.4 y Mardia, K.V. (1982) sección 6.3 ) que se lleva acabo en el ANÁLISIS DE REGRESIÓN LINEAL MÚLTIPLE : *La Prueba de Significancia Global* .

Se debe señalar que es posible aplicar la prueba de significancia global debido a que el modelo

$$Y_i = M + \beta \cos(\phi_i) + \gamma \text{Sen}(\phi_i) + \varepsilon_i \quad ; \quad i = 1, \dots, n$$

es lineal en los parámetros **M**,  $\beta$  y  $\gamma$  .

### La Hipótesis Lineal General

Sea el modelo lineal

$$\hat{Y} = X\beta + \bar{\varepsilon}$$

(2.8.1)

$\bar{\varepsilon}$  se distribuye  $N(\bar{0}, \sigma^2 I_n)$

**El Modelo más Simple de Regresión Periódica.**

---

en donde

$\bar{Y}$  es de tamaño  $(n \times 1)$   
 $\bar{\beta}$  es de tamaño  $(k \times 1)$   
 $X$  es de tamaño  $(n \times k)$   
 rango  $(X)$  es igual a  $k$

se desea probar en general

$$H_0: C\bar{\beta} = \bar{\gamma} \quad \text{vs} \quad H_a: C\bar{\beta} \neq \bar{\gamma} \quad (2.8.2)$$

en donde

$C$  es de tamaño  $(r \times k)$  con  $r \leq k$   
 Rango  $(C) = r$  es decir, rango completo  
 $\bar{\gamma}$  es de tamaño  $(r \times 1)$ ,

se sabe que

$\bar{Y}$  se distribuye  $N(X\bar{\beta}, \sigma^2 I_n)$

ya que

$\bar{e}$  se distribuye  $N(\bar{0}, \sigma^2 I_n)$ .

La región crítica utilizando la razón de verosimilitudes es

$$\left\{ \bar{Y} \mid \frac{\sup_{H_0} L(\bar{\beta}, \sigma^2 | \bar{Y})}{\sup_{H_0 \cup H_a} L(\bar{\beta}, \sigma^2 | \bar{Y})} \leq k_0 \right\} \quad (2.8.3)$$

donde

$$L(\bar{\beta}, \sigma^2 | \bar{Y}) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} (\bar{Y} - X\bar{\beta})' (\bar{Y} - X\bar{\beta})} \quad (2.8.4)$$

los estimadores máximo verosímiles para  $\bar{\beta}$  y  $\sigma^2$  son

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{Y}} \\ \hat{\sigma}^2 &= \frac{1}{n}(\bar{\mathbf{Y}} - \mathbf{X}\hat{\beta})'(\bar{\mathbf{Y}} - \mathbf{X}\hat{\beta})\end{aligned}\quad (2.8.5)$$

para obtener lo anterior se tuvo que maximizar  $\mathbf{L}(\bar{\beta}, \sigma^2 | \bar{\mathbf{Y}})$  sin restricciones, por lo tanto

$$\sup_{\mathbf{H}_0 \cup \mathbf{H}_a} \mathbf{L}(\bar{\beta}, \sigma^2 | \bar{\mathbf{Y}}) = \left(2\pi \hat{\sigma}^2\right)^{-n/2} e^{-n/2} \quad (2.8.6)$$

Para calcular el  $\sup_{\mathbf{H}_0} \mathbf{L}(\bar{\beta}, \sigma^2 | \bar{\mathbf{Y}})$  se tiene que maximizar  $\mathbf{L}(\bar{\beta}, \sigma^2 | \bar{\mathbf{Y}})$  sujeto a  $\mathbf{C}\bar{\beta} = \bar{\gamma}$ ; este problema puede convertirse en uno de minimización debido a que  $\mathbf{L}(\bar{\beta}, \sigma^2 | \bar{\mathbf{Y}})$  alcanza su máximo con respecto a  $\bar{\beta}$  cuando  $(\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})'(\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})$  es mínimo. Entonces basta con minimizar  $(\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})'(\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})$  sujeto a  $\mathbf{C}\bar{\beta} = \bar{\gamma}$ .

A continuación se utilizan multiplicadores de Lagrange :

Sea

$$\mathbf{J}(\bar{\beta}, \bar{\lambda}) = (\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})'(\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta}) + 2\bar{\lambda}'(\mathbf{C}\bar{\beta} - \bar{\gamma}) \quad (2.8.7)$$

en donde

$\bar{\lambda}$  es de tamaño  $(r \times 1)$ ;

se obtienen las derivadas parciales

$$\frac{\delta J}{\delta \beta} = -2\mathbf{X}'\bar{\mathbf{Y}} + 2\mathbf{X}'\mathbf{X}\bar{\beta} + 2\mathbf{C}'\bar{\lambda} \quad (2.8.8)$$

$$\frac{\delta J}{\delta \lambda} = -2(\mathbf{C}\bar{\beta} + \bar{\gamma})$$

igualando a cero las derivadas parciales y despejando se tiene que

$$\bar{\beta} = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}' \left[ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}' \right]^{-1} (\mathbf{C}\hat{\beta} - \bar{\gamma}) \quad (2.8.9)$$

minimiza a  $(\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})'(\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})$  sujeto a  $\mathbf{C}\bar{\beta} = \bar{\gamma}$ .

Ahora solo falta derivar  $L(\bar{\beta}, \sigma^2 | \bar{\mathbf{Y}})$  con respecto a  $\sigma^2$ . Sea

$$\mathcal{L}L = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})' (\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta}) \quad (2.8.10)$$

entonces

$$\frac{\delta \mathcal{L}L}{\delta \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})' (\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta}) \quad (2.8.11)$$

igualando a cero y despejando se obtiene

$$\hat{\sigma}^2_{\min} = \frac{1}{n} (\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta})' (\bar{\mathbf{Y}} - \mathbf{X}\bar{\beta}) \quad (2.8.12)$$

por lo tanto

$$\begin{aligned} \sup_{\text{Ho}} L(\bar{\beta}, \sigma^2 | \bar{\mathbf{Y}}) &= L(\bar{\beta}, \hat{\sigma}^2_{\min} | \bar{\mathbf{Y}}) \\ &= (2\pi \hat{\sigma}^2_{\min})^{-n/2} e^{-n/2} \end{aligned} \quad (2.8.13)$$

la razón de verosimilitudes queda de la siguiente forma

$$\frac{(\bar{Y} - \mathbf{X}\hat{\beta})'(\bar{Y} - \mathbf{X}\hat{\beta})}{(\bar{Y} - \mathbf{X}\tilde{\beta})'(\bar{Y} - \mathbf{X}\tilde{\beta})} = \frac{\text{SCE}_{\text{MC}}}{\text{SCE}_{\text{MR}}} \leq k_1 \quad (2.8.14)$$

por la descomposición de sumas de cuadrados se sabe que la SUMA DE CUADRADOS TOTAL es igual a la SUMA DE CUADRADOS NO EXPLICADA más la SUMA DE CUADRADOS EXPLICADA POR LA REGRESIÓN :

$$\text{SCT} = \text{SCE} + \text{SCR} \quad (2.8.15)$$

Por lo tanto

$$\text{SCE}_{\text{MR}} = \text{SCE}_{\text{MC}} + \text{SCE}_{\text{H}_0} \quad (2.8.16)$$

en donde  $\text{SCE}_{\text{MC}}$  es la suma de cuadrados del modelo completo y  $\text{SCE}_{\text{MR}}$  es la suma de cuadrados del modelo reducido; con lo anterior se obtiene lo siguiente :

$$\frac{\text{SCE}_{\text{H}_0}}{\text{SCE}_{\text{MC}}} \geq k_2 \quad (2.8.17)$$

Por otro lado se tiene que

$$\frac{\text{SCE}_{\text{MC}}}{\sigma^2} = \frac{(\bar{Y} - \mathbf{X}\hat{\beta})'(\bar{Y} - \mathbf{X}\hat{\beta})}{\sigma^2} \quad \text{se distribuye} \quad \chi^2_{(n-k)} \quad (2.8.18)$$

por otro lado y bajo  $H_0$  se tiene

$$\frac{SCE_{H_0}}{\sigma^2} = \frac{(C\hat{\beta} - \bar{y})' [C(X'X)^{-1}C']^{-1} (C\hat{\beta} - \bar{y})}{\sigma^2} \quad \text{se distribuye } \chi^2_{(r)}$$

ya que

$$(C\hat{\beta} - \bar{y}) \quad \text{se distribuye } N(C\bar{\beta} - \bar{y}, \sigma^2 C(X'X)^{-1}C') \quad (2.8.19)$$

como  $\frac{SCE_{MC}}{\sigma^2}$  y  $\frac{SCE_{H_0}}{\sigma^2}$  son independientes

$$\frac{SCE_{H_0}}{r} \Big/ \frac{SCE_{MC}}{n-k} \quad \text{se distribuye } F_{(r, n-k)} \quad (2.8.20)$$

por lo tanto la región crítica queda de la siguiente forma :

$$\left\{ \bar{Y} \mid \left( \frac{n-k}{r} \right) \left( \frac{SCE_{H_0}}{SCE_{MC}} \right) \geq F_{(r, n-k)}^{1-\alpha} \right\} \quad (2.8.21)$$

a continuación se presenta la tabla de análisis de varianza :

**El Modelo más Simple de Regresión Periódica.**

	Grados de libertad	Suma de cuadrados	Cuadrados medios	F
<b>H<sub>0</sub></b>	<b>r</b>	<b>SCE<sub>H<sub>0</sub></sub></b>	$\frac{SCE_{H_0}}{r} = CME_{H_0}$	$\frac{CME_{H_0}}{CME_{MC}}$
<b>Error (Modelo completo)</b>	<b>n-k</b>	<b>SCE<sub>MC</sub></b>	$\frac{SCE_{MC}}{n-k} = CME_{MC}$	
<b>Total</b>	<b>n-k+r</b>	<b>SCE<sub>MR</sub></b>		

Un caso especial de la forma general es la *Prueba de Significancia Global* :

**H<sub>0</sub>:  $\beta_1 = \beta_2 = \dots = \beta_{k-1} = 0$**  vs **H<sub>a</sub>:  $\beta_j \neq 0$**  para alguna **j**  
 en donde

$$C = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{r \times k}, \quad \tilde{\gamma} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{r \times 1}$$

En el caso del modelo más simple de regresión periódica, al realizar la prueba de hipótesis :

**H<sub>0</sub>:  $\beta = \gamma = 0$**  vs

**H<sub>a</sub>: Al menos uno de estos dos parámetros, es diferente de cero.**

se considera lo siguiente :

$$r=2, \quad k=3, \quad C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \tilde{\gamma} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$



$$\hat{\beta} = \begin{pmatrix} \hat{M} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & C_1 & S_1 \\ 1 & C_2 & S_2 \\ \vdots & \vdots & \vdots \\ 1 & C_n & S_n \end{pmatrix}, \quad \bar{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

## 2.9 Predicción

El análisis de regresión periódica tiene dos objetivos fundamentales : El primero consiste en ajustar una función periódica a un conjunto de datos pertenecientes a un fenómeno cíclico o a un fenómeno de direcciones; para que este ajuste sea confiable se debe contar con un conjunto de datos igualmente espaciados en el periodo  $T$ , si esto no sucede basta con que los datos sean lo suficientemente densos en el mismo periodo  $T$  (Ver figura (2.9.1)).

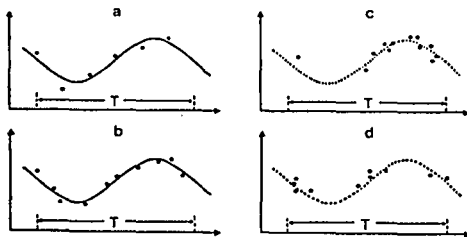


Figura 2.9.1 Ajustando una función periódica. (a) Los datos están igualmente espaciados. (b) Los datos no están igualmente espaciados, pero son lo suficientemente densos en el periodo. (c) Los datos están agrupados en una sola parte del periodo. El ajuste que se obtiene no es confiable. (d) Los datos están agrupados en ciertas partes, en este caso tampoco se obtiene un ajuste confiable Tomado de Batschelet, E.(1981).

### **El Modelo más Simple de Regresión Periódica.**

---

El segundo objetivo es de gran interés y consiste en hacer la predicción de respuestas futuras y la estimación de respuestas medias, así como la predicción de respuestas fuera del rango de los datos originales.

#### **Intervalo de confianza de la respuesta media**

Una vez obtenido el modelo de regresión periódica, es importante estimar la respuesta media  $E(Y)$  para algún valor particular de las variables regresoras ó independientes :

$$\text{Cos}(\omega t) = \text{Cos}\phi \quad , \quad \text{Sen}(\omega t) = \text{Sen}\phi$$

Sean  $C_0 = \text{Cos}\omega t_0 = \text{Cos}\phi_0$  y  $S_0 = \text{Sen}\omega t_0 = \text{Sen}\phi_0$  los valores de las variables regresoras para los cuales se desea estimar la respuesta media, es decir,  $E(Y|C_0, S_0)$ . Para este caso se debe suponer que  $C_0$  y  $S_0$  es cualquier valor de sus respectivas variables regresoras sobre el rango de los datos usados para ajustar el modelo.

Un estimador puntual insesgado para  $E(Y|C_0, S_0)$  es el siguiente

$$\hat{Y}_0 = \hat{M} + \hat{\beta}C_0 + \hat{\gamma}S_0 \quad (2.9.1)$$

donde  $\hat{M}$ ,  $\hat{\beta}$  y  $\hat{\gamma}$  son los estimadores por mínimos cuadrados y máxima verosimilitud obtenidos en secciones anteriores. La varianza de  $\hat{Y}_0$  se presenta a continuación :

**El Modelo más Simple de Regresión Periódica.**

$$\text{Var}(\hat{Y}_0) = \left[ \frac{1}{n} + \frac{(C_0 - \bar{C})^2 S_{SS} - (C_0 - \bar{C})(S_0 - \bar{S}) S_{CS}}{S_{CC} S_{SS} - (S_{CS})^2} + \frac{(S_0 - \bar{S})^2 S_{CC}}{S_{CC} S_{SS} - (S_{CS})^2} \right] \sigma^2 \quad (2.9.2)$$

Como  $\hat{Y}_0$  es una combinación lineal de las  $Y_i$ 's entonces también se distribuye Normal, por lo tanto

$$\frac{\hat{Y}_0 - E(Y|C_0, S_0)}{\sqrt{\left( \frac{1}{n} + \frac{(C_0 - \bar{C})^2 S_{SS} - (C_0 - \bar{C})(S_0 - \bar{S}) S_{CS} + (S_0 - \bar{S})^2 S_{CC}}{S_{CC} S_{SS} - (S_{CS})^2} \right) \sigma^2}}$$

se distribuye  $N(0,1)$ , además se sabe también que  $\frac{1}{\sigma^2} \sum_{i=1}^n e_i^2$  se distribuye  $\chi^2_{(n-3)}$  y es independiente de  $\hat{Y}_0$ , por lo tanto

$$\frac{\hat{Y}_0 - E(Y|C_0, S_0)}{\sqrt{\left( \frac{1}{n} + \frac{(C_0 - \bar{C})^2 S_{SS} - (C_0 - \bar{C})(S_0 - \bar{S}) S_{CS} + (S_0 - \bar{S})^2 S_{CC}}{S_{CC} S_{SS} - (S_{CS})^2} \right) \hat{\sigma}_{ins}^2}}$$

se distribuye  $t_{(n-3)}$ .

### El Modelo más Simple de Regresión Periódica.

---

Por lo tanto un intervalo del  $(1-\alpha) \times 100\%$  de confianza para la respuesta media en el punto  $(C_0, S_0)$  es

$$\hat{Y}_0 \pm t_{(1-\frac{\alpha}{2}, n-3)} \sqrt{\text{Var}(\hat{Y}_0) \frac{\hat{\sigma}^2_{ins}}{\sigma^2}} \quad (2.9.3)$$

donde  $-t_{(1-\frac{\alpha}{2}, n-3)}$  y  $t_{(1-\frac{\alpha}{2}, n-3)}$  son los cuantiles  $\frac{\alpha}{2}$  y  $1-\frac{\alpha}{2}$ , respectivamente, de una distribución  $t$  con  $(n-3)$  grados de libertad y además, el parámetro  $\sigma^2$  se elimina con el de la expresión de la  $\text{Var}(\hat{Y}_0)$ . Note que la amplitud mínima del intervalo ocurre en el punto  $(C_0, S_0) = (\bar{C}, \bar{S})$ .

### Predicción de observaciones nuevas

Si  $(C_0, S_0)$  es el valor de  $(\text{Cos}(\omega t), \text{Sen}(\omega t))$  entonces

$$\hat{Y}_0 = \hat{M} + \hat{\beta}C_0 + \hat{\gamma}S_0 \quad (2.9.4)$$

es el estimador puntual del nuevo valor de respuesta  $Y_0$ .

### Intervalo de predicción para la futura observación Y

Sea

$$\psi = (Y_0 - \hat{Y}_0) \quad (2.9.5)$$

cuya distribución es Normal

$$\psi \text{ se distribuye } N(0, \text{Var}(\psi)) \quad (2.9.6)$$

ahora se calcula lo siguiente :

$$\text{Var}(\psi) = \text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(Y_0) + \text{Var}(\hat{Y}_0)$$

esto es debido a que la observación futura  $Y_0$  es independiente de  $\hat{Y}_0$ , por lo tanto

$$\text{Var}(\psi) = \sigma^2 + \text{Var}(\hat{Y}_0) \quad (2.9.7)$$

Por lo tanto un intervalo de predicción del  $(1-\alpha) \times 100\%$  de confianza sobre una observación futura en el punto  $(C_0, S_0)$  es

$$\hat{Y}_0 \pm t_{(1-\frac{\alpha}{2}, n-3)} \sqrt{\text{Var}(\psi) \frac{\hat{\sigma}^2_{ins}}{\sigma^2}} \quad (2.9.8)$$

donde  $-t_{(1-\frac{\alpha}{2}, n-3)}$  y  $t_{(1-\frac{\alpha}{2}, n-3)}$  son los cuantiles  $\frac{\alpha}{2}$  y  $1-\frac{\alpha}{2}$ , respectivamente, de una distribución  $t$  con  $(n-3)$  grados de libertad y además, el parámetro  $\sigma^2$  se elimina con el de la expresión de la  $\text{Var}(\psi)$ .

## 2.10 Inferencia simultánea sobre los parámetros del modelo

En secciones anteriores se obtuvieron intervalos de confianza individuales para los parámetros del modelo; para esta sección se hará inferencia simultánea sobre estos parámetros.

Suponga que se tiene un intervalo del 95% de confianza para  $M$ , otro intervalo del 95% de confianza para  $\beta$  y un intervalo del 95% de confianza para  $\gamma$ ; si estas estimaciones por intervalo fueran independientes, la probabilidad asociada a los tres conjuntamente sería

**El Modelo más Simple de Regresión Periódica.**

---

$(0.95)^3 = 0.857375$ . Sin embargo, los intervalos están basados en estimadores que NO SON EN GENERAL INDEPENDIENTES.

Para encontrar una región simultánea del  $(1-\alpha) \times 100\%$  de confianza para  $M$ ,  $\beta$  y  $\gamma$ ; se utilizará la *Región General de Confianza* que se obtiene en el ANÁLISIS DE REGRESIÓN MÚLTIPLE (Ver Wonnacott, Thomas & Wonnacott, Ronald (1981) sección 12.5 y Montgomery, D.C. & Peck, E.A. (1992) sección 4.7):

$$\frac{(\mathbf{C}\bar{\beta} - \mathbf{C}\hat{\beta})'(\mathbf{X}'\mathbf{X})(\mathbf{C}\bar{\beta} - \mathbf{C}\hat{\beta})}{r\hat{\sigma}^2} \leq \mathbf{F}_{(r, n-k)}^{1-\alpha} \quad (2.10.1)$$

Se debe señalar que es posible aplicar este resultado, debido a que el modelo

$$Y_i = M + \beta \cos(\phi_i) + \gamma \sin(\phi_i) + \varepsilon_i \quad ; \quad i = 1, \dots, n$$

es lineal en los parámetros  $M$ ,  $\beta$  y  $\gamma$ .

En el caso del modelo más simple de regresión periódica, se tiene que

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \bar{\beta} = \begin{pmatrix} M \\ \beta \\ \gamma \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \hat{M} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix}$$

**El Modelo más Simple de Regresión Periódica.**

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^n C_i & \sum_{i=1}^n S_i \\ \sum_{i=1}^n C_i & \sum_{i=1}^n C_i^2 & \sum_{i=1}^n C_i S_i \\ \sum_{i=1}^n S_i & \sum_{i=1}^n C_i S_i & \sum_{i=1}^n S_i^2 \end{pmatrix}$$

$$\mathbf{k} = \mathbf{r} = 3$$

sustituyendo lo anterior en (2.10.1) y desarrollando se obtiene la región simultánea del  $(1-\alpha) \times 100\%$  de confianza para los parámetros del modelo :

$$\begin{aligned} n(\mathbf{M} - \hat{\mathbf{M}})^2 + \sum_{i=1}^n C_i^2 (\beta - \hat{\beta})^2 + \sum_{i=1}^n S_i^2 (\gamma - \hat{\gamma})^2 + 2 \sum_{i=1}^n C_i (\beta - \hat{\beta})(\mathbf{M} - \hat{\mathbf{M}}) \\ + 2 \sum_{i=1}^n S_i (\gamma - \hat{\gamma})(\mathbf{M} - \hat{\mathbf{M}}) + 2 \sum_{i=1}^n C_i S_i (\beta - \hat{\beta})(\gamma - \hat{\gamma}) \leq 3\hat{\sigma}^2 \mathbf{F}_{(3, n-3)}^{1-\alpha} \end{aligned}$$

(2.10.2)

la desigualdad anterior representa un elipsoide que con muestreo repetido contendrá a los parámetros  $\mathbf{M}$ ,  $\beta$  y  $\gamma$  simultáneamente, el  $(1-\alpha) \times 100\%$  de las veces.

De forma análoga se puede obtener una región simultánea del  $(1-\alpha) \times 100\%$  de confianza para  $\beta$  y  $\gamma$ . Para este caso se tiene lo siguiente :

**El Modelo más Simple de Regresión Periódica.**

---

$$C = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \bar{\beta} = \begin{pmatrix} M \\ \beta \\ \gamma \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \hat{M} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix}$$

$$X'X = \begin{pmatrix} n & \sum_{i=1}^n C_i & \sum_{i=1}^n S_i \\ \sum_{i=1}^n C_i & \sum_{i=1}^n C_i^2 & \sum_{i=1}^n C_i S_i \\ \sum_{i=1}^n S_i & \sum_{i=1}^n C_i S_i & \sum_{i=1}^n S_i^2 \end{pmatrix}$$

$$r = 2 \quad , \quad k = 3$$

sustituyendo lo anterior en (2.10.1) y desarrollando se obtiene la región simultánea del  $(1-\alpha) \times 100\%$  de confianza para  $\beta$  y  $\gamma$ :

$$\sum_{i=1}^n C_i^2 (\beta - \hat{\beta})^2 + \sum_{i=1}^n S_i^2 (\gamma - \hat{\gamma})^2 + 2 \sum_{i=1}^n C_i S_i (\beta - \hat{\beta})(\gamma - \hat{\gamma}) \leq 2\sigma^2 \mathbf{F}_{(2, n-3)}^{1-\alpha}$$

(2.10.3)

la desigualdad anterior representa una elipse que con muestreo repetido contendrá a los parámetros  $\beta$  y  $\gamma$  simultáneamente, el  $(1-\alpha) \times 100\%$  de las veces.



### ***CAPÍTULO 3. APLICACIÓN A FENÓMENOS CÍCLICOS Y A FENÓMENOS DE DIRECCIONES***

#### **3.1 Introducción**

En el presente capítulo se aplicará la teoría del ANÁLISIS DE REGRESIÓN PERIÓDICA a dos fenómenos de direcciones y a dos fenómenos cíclicos :

- a) La dirección del viento dominante y su velocidad media
- b) La dirección del viento dominante y su velocidad máxima
- c) El consumo de energía eléctrica a nivel nacional para el Sector Doméstico
- d) El consumo de energía eléctrica a nivel nacional para el Sector Industrial

Los datos estadísticos que se utilizaron en los fenómenos de direcciones, se obtuvieron de las TARJETAS DE RESUMEN MENSUAL, que maneja el SERVICIO METEOROLÓGICO NACIONAL, el cual se ubica en Av. Observatorio No. 192, México D.F.

Por otro lado, los datos estadísticos que se utilizaron en los fenómenos cíclicos se obtuvieron del BANCO DE DATOS DEL INEGI, el cual puede ser consultado en el CENTRO REGIONAL DE INFORMACIÓN, ubicado en Av. Patriotismo No. 711, México D.F.

### 3.2 Viento Dominante y Velocidad Media

Los datos que a continuación se presentan corresponden a la dirección del viento dominante y su velocidad media, registradas en la zona del Meteorológico de Tacubaya entre enero de 1992 y diciembre de 1995.

DIRECCIÓN	VELOCIDAD (m/seg)
N	3.07
NE	2.50
ENE	1.42
ESE	2.00
SSE	4.60
WSW	3.70
W	3.20
WNW	3.27
NW	2.92

si a cada una de estas direcciones se le asocia un ángulo que fluctúa entre 0° y 360°, se obtiene lo siguiente :

DIRECCIÓN	GRADOS
N	0
NE	45
ENE	67.5
ESE	112.5
SSE	157.5
WSW	247.5

### Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.

W	270
WNW	292.5
NW	315

de esta forma se puede construir un histograma circular (Ver figura (3.2.1)).

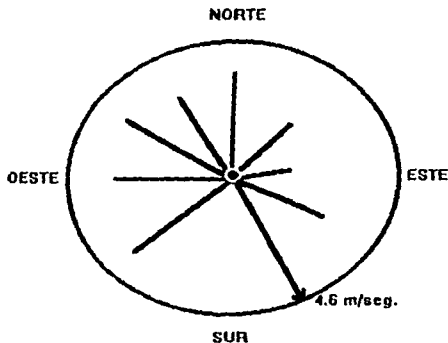


Figura 3.2.1 Histograma circular para la dirección del viento dominante y su velocidad media

La longitud del vector medio,  $r$ , para las direcciones que intervienen en este fenómeno es igual a **0.3216**, esto indica que las direcciones están lo suficientemente espaciadas en una rotación completa de  $360^\circ$ ; por lo tanto se puede ajustar una función periódica para modelar este fenómeno (Ver Figura (1.3.7)).

Si se representa este fenómeno de direcciones en un plano cartesiano, se observa que efectivamente, una función periódica podría ser un buen modelo (Ver figura (3.2.2)).

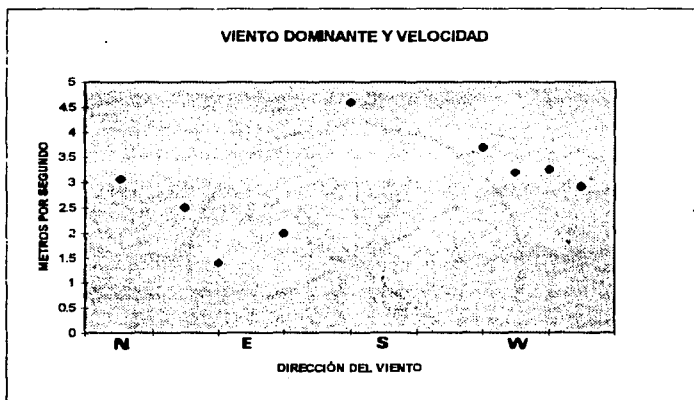


Figura 3.2.2 Dirección del viento dominante y su velocidad media (m/seg) .

Para este caso se tiene que el periodo,  $T$ , es igual a  $360^\circ$ , entonces  $\omega = 1$ , es decir, en el periodo sólo habrá un valle y sólo habrá una cresta. Con toda la información anterior se obtiene lo siguiente :

$$\begin{aligned}\bar{Y} &= 2.9644 \\ \bar{C} &= 0.1656 \\ \bar{S} &= -0.0686\end{aligned}$$

$$\begin{aligned}S_{ss} &= 5.5183 \\ S_{yy} &= 6.9092 \\ S_{sy} &= -3.1864\end{aligned}$$

**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

---

$$\sum_{i=1}^n C_i = 1.4903$$

$$Scs = - 0.2513$$

$$\sum_{i=1}^n S_i = - 0.6173$$

$$Scy = - 2.1518$$

$$Scc = 3.1926$$

$$\sum_{i=1}^n e_i^2 = 3.4107$$

A continuación se calculan los estimadores que son útiles para proponer una función que modele a este fenómeno de direcciones :

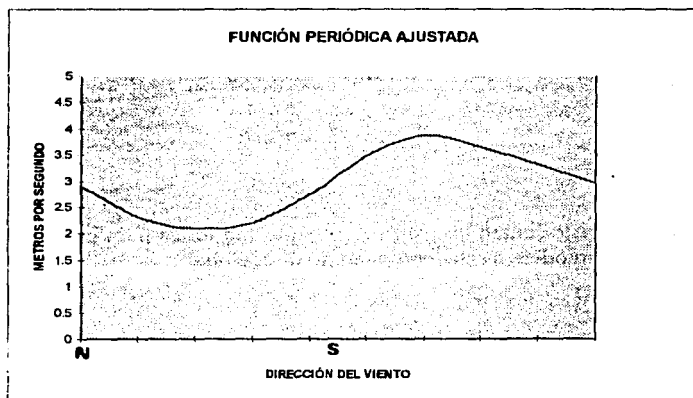
$$\begin{aligned} \hat{\gamma} &= - 0.6103 \\ \hat{\beta} &= - 0.7220 \\ \hat{M} &= 3.0422 \end{aligned} \quad (3.2.1)$$

por lo tanto la función periódica que se propone para modelar la dirección del viento dominante y su velocidad media es

$$\hat{Y} = 3.0422 - 0.7220 \text{ Cos}\phi - 0.6103 \text{ Sen}\phi \quad (3.2.2)$$

En la figura (3.2.3) se muestra la función periódica ajustada.

**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**



*Figura 3.2.3 Función Periódica Ajustada*

Los estimadores puntuales para el ángulo de la acrofase y para la amplitud se presentan a continuación :

$$\hat{\phi}_A = 220.2^\circ \quad (3.2.3)$$

$$\hat{A} = 0.9454$$

Los intervalos de confianza del 95% para  $M$ ,  $\beta$  y  $\gamma$ , son los siguientes :

$$\begin{aligned} 2.8292 < M < 3.2552 \\ -1.7145 < \beta < 0.2705 \\ -1.3653 < \gamma < 0.1447 \end{aligned} \quad (3.2.4)$$

de lo anterior se puede inferir que  $M \neq 0$ . También se observa que el cero está incluido en los intervalos de confianza de  $\beta$  y  $\gamma$ , por tal razón se hará una prueba de hipótesis para saber si en verdad el modelo tiene sentido.

El intervalo de confianza del 95% para  $\sigma^2$  es el siguiente :

$$0.2360 < \sigma^2 < 2.7565 \quad (3.2.5)$$

### Prueba de significancia global

**H<sub>0</sub>:**  $\beta = \gamma = 0$  vs

**H<sub>a</sub>:** Al menos uno de estos dos parámetros, es diferente de cero.

A continuación se presenta la tabla de análisis de varianza :

	Grados de libertad	Suma de cuadrados	Cuadrados medios	F
<b>H<sub>0</sub></b>	2	3.4981	1.7490	3.0768
<b>Error (Modelo completo)</b>	6	3.4107	0.5685	
<b>Total</b>	8	6.9088		

Tomando  $\alpha = 0.05$ , se tiene

$$3.0768 < 5.14 = F_{(2,6)}^{0.95} \quad (3.2.6)$$

por lo tanto no se rechaza  $H_0$ , es decir, el modelo que se ha propuesto no tiene sentido.

A continuación se obtiene el coeficiente de correlación lineal circular muestral :

$$R_{y,\phi} = 0.7115 \quad (3.2.7)$$

lo cual indica que en la muestra existe una cierta correlación, aunque no muy fuerte, entre la dirección del viento dominante y su velocidad media. Por lo tanto

$$R_{y,\phi}^2 = 0.5063 \quad (3.2.8)$$

es la proporción de la variación total de  $Y$  ( velocidad media ) que queda explicada por la regresión; en otras palabras, el 50.63% de la variación en la velocidad media es explicada por la variación en la dirección del viento dominante.

Con lo anterior se puede realizar una prueba de hipótesis para el coeficiente de correlación lineal circular poblacional

$$H_0: \rho_{Y,\phi} = 0 \quad \text{vs} \quad H_a: \rho_{Y,\phi} \neq 0$$



ESTADÍSTICA DE PRUEBA :

$$U = \frac{\frac{n-3}{2} R_{y,\phi}^2}{1 - R_{y,\phi}^2}$$

REGLA DE DECISIÓN :

Rechazar  $H_0$  si  $U > F_{(2,n-3)}^{1-\alpha}$

Sea  $\alpha = 0.05$

$$U = 3.0768 < 5.14 = F_{(2,6)}^{0.95} \quad (3.2.9)$$

por lo tanto  $\rho_{V,\phi} = 0$ , es decir, la velocidad media del viento dominante es independiente de su dirección; lo cual era esperado.

### 3.3 Viento Dominante y Velocidad Máxima

Los datos que a continuación se presentan corresponden a la dirección del viento dominante y su velocidad máxima, registradas en la zona del Meteorológico de Tacubaya entre enero de 1992 y diciembre de 1995.

DIRECCIÓN	VELOCIDAD (m/seg)
N	13.88
NE	14.50
ENE	16.50
E	16.55
ESE	12.40
SE	14.80
SSE	14.00

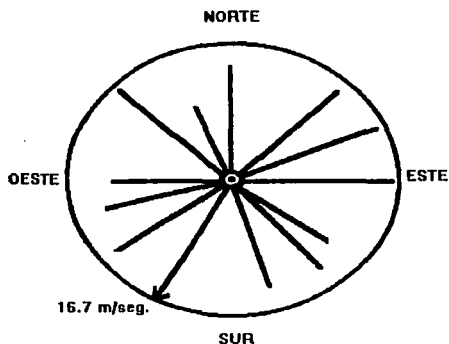
**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

SSW	16.70
SW	15.07
WSW	12.30
W	11.15
WNW	16.57
NW	10.45
NNW	14.77

si a cada una de estas direcciones se le asocia un ángulo que fluctúa entre  $0^\circ$  y  $360^\circ$ , se obtiene lo siguiente :

DIRECCIÓN	GRADOS
N	0
NE	45
ENE	67.5
E	90
ESE	112.5
SE	135
SSE	157.5
SSW	202.5
SW	225
WSW	247.5
W	270
WNW	292.5
NW	315
NNW	337.5

de esta forma se puede construir un histograma circular (Ver figura (3.3.1)).



*Figura 3.3.1 Histograma circular para la dirección del viento dominante y su velocidad máxima*

La longitud del vector medio,  $r$ , para las direcciones que intervienen en este fenómeno es igual a **0.1320**, esto indica que las direcciones están lo suficientemente espaciadas en una rotación completa de **360°**; por lo tanto se puede ajustar una función periódica para modelar este fenómeno ( Ver Figura (1.3.7) ).

Si se representa este fenómeno de direcciones en un plano cartesiano, se observa que efectivamente, una función periódica con dos crestas y dos valles, podría ser un buen modelo (Ver figura (3.3.2)).

**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

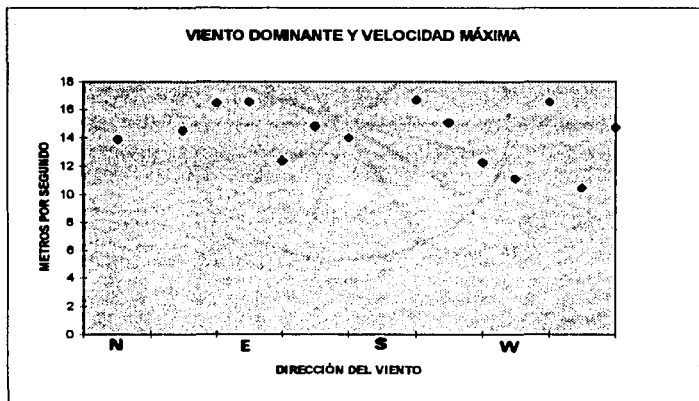


Figura 3.3.2 Dirección del viento dominante y su velocidad máxima (m/seg) .

para este caso se tiene que el periodo,  $T$ , es igual a  $180^\circ$ , entonces  $\omega = 2$ . Con toda la información anterior se obtiene lo siguiente :

$$\bar{Y} = 14.26$$

$$\bar{C} = -0.1219$$

$$\bar{S} = -0.0505$$

$$\sum_{i=1}^n C_i = -1.7071$$

$$S_{ss} = 7.4649$$

$$S_{yy} = 54.5182$$

$$S_{sy} = 5.7484$$

$$S_{cs} = -0.5862$$

$$\sum_{i=1}^n S_i = -0.7071$$
$$S_{cc} = 6.2918$$

$$S_{cy} = 1.8259$$

$$\sum_{i=1}^n e_i^2 = 49.2609$$

A continuación se calculan los estimadores que son útiles para proponer una función que modele a este fenómeno de direcciones :

$$\begin{aligned}\hat{\gamma} &= 0.7988 \\ \hat{\beta} &= 0.3646 \\ \hat{M} &= 14.3448\end{aligned}\tag{3.3.1}$$

por lo tanto la función periódica que se propone para modelar la dirección del viento dominante y su velocidad media es :

$$\hat{Y} = 14.3448 + 0.3646 \cos 2\phi + 0.7988 \operatorname{Sen} 2\phi\tag{3.3.2}$$

En la figura (3.3.3) se muestra la función periódica ajustada.

### Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.

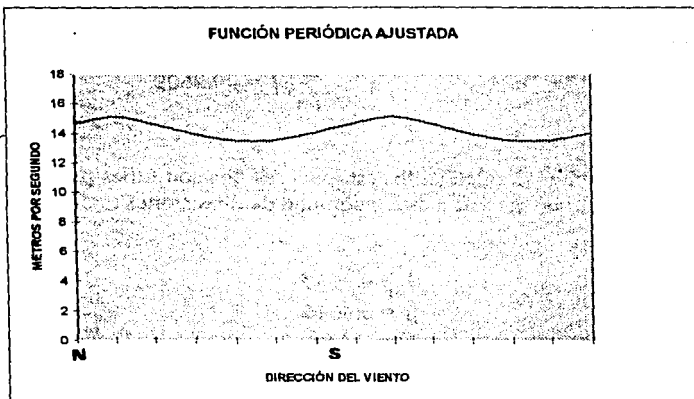


Figura 3.3.3 Función Periódica Ajustada

Los estimadores puntuales para el ángulo de la acrofase y para la amplitud, se presentan a continuación :

$$\hat{\phi}_A = 65.47^\circ, 130.93^\circ \quad (3.3.3)$$

$$\hat{A} = 0.8780$$

Los intervalos de confianza del 95% para  $M$ ,  $\beta$  y  $\gamma$ , son los siguientes :

$$\begin{aligned}
 13.0027 < M < 15.6869 \\
 -1.6097 < \beta < 2.3388 \\
 -1.0137 < \gamma < 2.6113
 \end{aligned}
 \tag{3.3.4}$$

de lo anterior se puede inferir que  $M \neq 0$ . También se observa que el cero está incluido en los intervalos de confianza de  $\beta$  y  $\gamma$ , por tal razón se hará una prueba de hipótesis para saber si en verdad el modelo tiene sentido.

El intervalo de confianza del 95% para  $\sigma^2$  es el siguiente :

$$3.4092 < \sigma^2 < 39.8117 \tag{3.3.5}$$

**Prueba de significancia global**

**H<sub>0</sub>:**  $\beta = \gamma = 0$  vs

**H<sub>a</sub>:** Al menos uno de estos dos parámetros , es diferente de cero .

A continuación se presenta la tabla de análisis de varianza :

	Grados de libertad	Suma de cuadrados	Cuadrados medios	F
<b>H<sub>0</sub></b>	2	5.2577	2.6288	<b>0.5870</b>
<b>Error (Modelo completo)</b>	11	49.2609	4.4783	
<b>Total</b>	13	54.5186		

Tomando  $\alpha = 0.05$ , se tiene

$$0.5870 < 3.98 = F_{(2,11)}^{0.95} \quad (3.3.6)$$

por lo tanto no se rechaza  $H_0$ , es decir, el modelo que se ha propuesto no tiene sentido.

A continuación se obtiene el coeficiente de correlación lineal circular muestral :

$$R_{y,\phi} = 0.3105 \quad (3.3.7)$$

lo cual indica que en la muestra casi no existe correlación entre la dirección del viento dominante y su velocidad máxima. Por lo tanto

$$R_{y,\phi}^2 = 0.0964 \quad (3.3.8)$$

es la proporción de la variación total de  $Y$  ( velocidad máxima ) que queda explicada por la regresión periódica. Esto indica una evidente mala calidad en el ajuste.

Finalmente se hará una prueba de hipótesis para el coeficiente de correlación lineal circular poblacional

$$H_0: \rho_{Y,\phi} = 0 \quad \text{vs} \quad H_a: \rho_{Y,\phi} \neq 0$$



ESTADÍSTICA DE PRUEBA :

$$U = \frac{\frac{n-3}{2} R_{y,\phi}^2}{1 - R_{y,\phi}^2}$$

REGLA DE DECISIÓN :

Rechazar  $H_0$  si  $U > F_{(2,n-3)}^{1-\alpha}$

Sea  $\alpha = 0.05$

$$U = 0.5870 < 3.98 = F_{(2,11)}^{0.95} \quad (3.3.9)$$

por lo tanto  $\rho_{Y,\phi} = 0$ , es decir, la velocidad máxima del viento dominante es independiente de su dirección, lo cual era esperado.

### 3.4 Consumo de Energía Eléctrica a Nivel Nacional ( Sector Doméstico )

Los datos que a continuación se presentan corresponden al consumo mensual promedio de energía eléctrica, a nivel nacional, para el sector doméstico durante el periodo de enero de 1990 a diciembre de 1995.

MES	CONSUMO DE ENERGÍA ELÉCTRICA ( Miles de millones de watts/hora )
ENE	1913.50
FEB	1830.83
MAR	1768.50
ABR	1800.67

**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

MAY	1884.67
JUN	2026.83
JUL	2153.00
AGO	2283.00
SEP	2308.00
OCT	2254.50
NOV	2119.83
DIC	1979.00

si a cada mes se le asocia un ángulo que fluctúa entre  $0^\circ$  y  $360^\circ$ , se tiene lo siguiente

<b>DIRECCIÓN</b>	<b>GRADOS</b>
ENE	0
FEB	30
MAR	60
ABR	90
MAY	120
JUN	150
JUL	180
AGO	210
SEP	240
OCT	270
NOV	300
DIC	330

de esta forma se puede construir un histograma circular (Ver figura (3.4.1)).

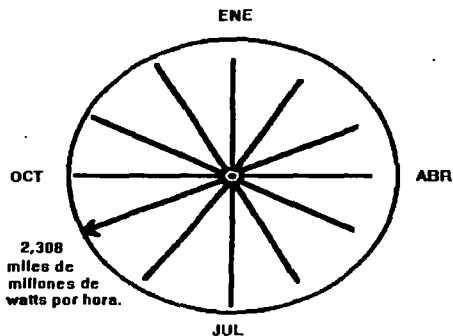


Figura 3.4.1 Histograma circular para el consumo mensual promedio de energía eléctrica (Sector Doméstico).

La longitud del vector medio,  $r$ , para los meses registrados en este fenómeno es igual a  $3.33 \times 10^{-17}$ , esto confirma que los ángulos asociados a cada mes están uniformemente distribuidos en una rotación completa de  $360^\circ$ ; por lo tanto se puede ajustar una función periódica para modelar este fenómeno (Ver Figura (1.3.7)).

Si se representa este fenómeno cíclico en un plano cartesiano, se observa que efectivamente, una función periódica puede ser un buen modelo (Ver figura (3.4.2)).

Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.

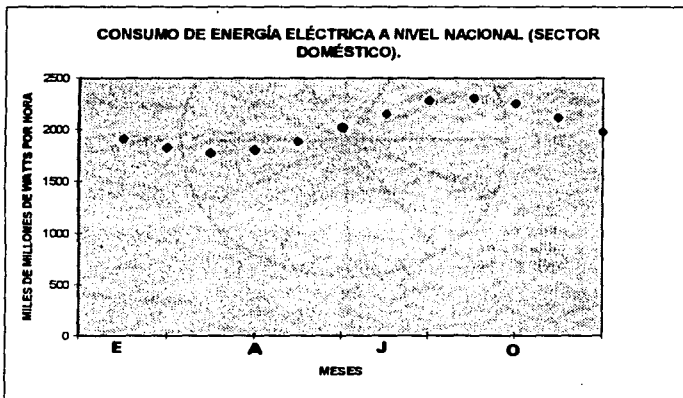


Figura 3.4.2 Consumo mensual promedio de energía eléctrica a nivel nacional (Sector Doméstico).

para este caso se tiene que el periodo,  $T$ , es igual a  $360^\circ$ , entonces  $\omega = 1$ , es decir, en el periodo solo habrá un valle y solo habrá una cresta. Con toda la información anterior se obtiene lo siguiente :

$$\bar{Y} = 2026.8608$$

$$S_{ss} = 6$$

$$\bar{C} = -2.7756 \times 10^{-17}$$

$$S_{yy} = 412720.72$$

**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

---

$$\bar{S} = -1.8504 \times 10^{-17}$$

$$\sum_{i=1}^n C_i = -3.3307 \times 10^{-16}$$

$$\sum_{i=1}^n S_i = -2.2204 \times 10^{-16}$$

$$S_{cc} = 6$$

$$S_{sy} = -1326.87$$

$$S_{cs} = 2.7756 \times 10^{-16}$$

$$S_{cy} = -824.68$$

$$\sum_{i=1}^n e_i^2 = 5937.47$$

A continuación se calculan los estimadores que son útiles para proponer una función que modele a este fenómeno cíclico :

$$\begin{aligned} \hat{\gamma} &= -221.1459 \\ \hat{\beta} &= -137.4472 \\ \hat{M} &= 2026.8608 \end{aligned} \quad (3.4.1)$$

por lo tanto la función periódica que se propone para modelar el consumo mensual promedio de energía eléctrica, a nivel nacional, para el sector doméstico es :

$$\hat{Y} = 2026.8608 - 137.4472 \text{ Cos}\phi - 221.1459 \text{ Sen}\phi \quad (3.4.2)$$

En la figura (3.4.3) se muestra la función periódica ajustada.

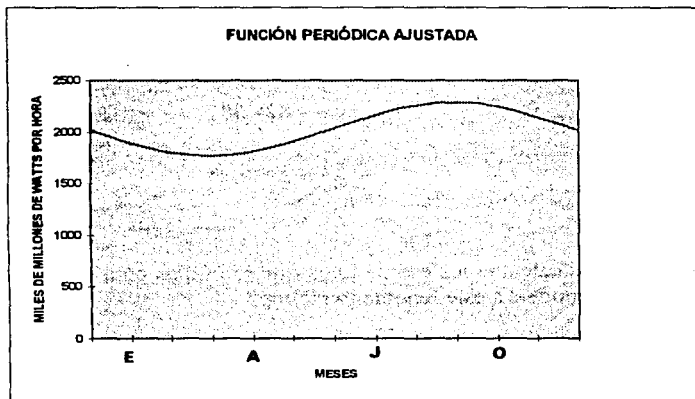


Figura 3.4.3 Función Periódica Ajustada

A continuación se presentan los estimadores puntuales para el ángulo de la acrofase y para la amplitud :

$$\hat{\phi}_A = 238.13^\circ \tag{3.4.3}$$

$$\hat{A} = 260.379$$

Los intervalos de confianza del 95% para los estimadores de los parámetros de la función ajustada, son los siguientes :

**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

$$\begin{aligned} 2022.1798 < M < 2031.5418 \\ -160.3794 < \beta < -114.5150 \\ -244.0779 < \gamma < -198.2138 \end{aligned} \quad (3.4.4)$$

de lo anterior se puede inferir que  $M \neq 0$ ,  $\beta \neq 0$  y  $\gamma \neq 0$ ; esto se confirmará al realizar la prueba de significancia global.

El intervalo de confianza del 95% para  $\sigma^2$  es el siguiente :

$$410.9146 < \sigma^2 < 4798.5488 \quad (3.4.5)$$

**Prueba de significancia global**

**H<sub>0</sub>:**  $\beta = \gamma = 0$  vs

**H<sub>a</sub>:** Al menos uno de estos dos parámetros, es diferente de cero.

A continuación se presenta la tabla de análisis de varianza :

	Grados de libertad	Suma de cuadrados	Cuadrados medios	F
H <sub>0</sub>	2	406783.28	203391.64	308.30
Error (Modelo completo)	9	5937.48	659.71	
Total	11	412720.76		

Tomando  $\alpha = 0.05$ , se tiene

$$308.30 > 4.26 = F_{(2,9)}^{0.95} \quad (3.4.6)$$

por lo tanto se rechaza  $H_0$ , es decir, el modelo que se ha propuesto tiene sentido. Se puede notar que la prueba es altamente significativa como se esperaba.

A continuación se obtiene el coeficiente de correlación lineal circular muestral

$$R_{y,\phi} = 0.9928 \quad (3.4.7)$$

lo cual indica que en la muestra existe una muy fuerte correlación entre un determinado mes del año y el consumo promedio de energía eléctrica que hay en él. Por lo tanto

$$R_{y,\phi}^2 = 0.9857 \quad (3.4.8)$$

es la proporción de la variación total de  $Y$  ( consumo mensual promedio de energía eléctrica ) que queda explicada por la regresión periódica; en otras palabras, el 98.57% de la variación en el consumo mensual promedio de energía eléctrica es explicada por la variación en el mes de consumo. Esto indica que se tiene un muy buen ajuste.

Con lo anterior se puede realizar una prueba de hipótesis para el coeficiente de correlación lineal circular poblacional

$$H_0: \rho_{Y,\Phi} = 0 \quad \text{vs} \quad H_a: \rho_{Y,\Phi} \neq 0$$



ESTADÍSTICA DE PRUEBA :

$$U = \frac{\frac{n-3}{2} R_{y,\phi}^2}{1 - R_{y,\phi}^2}$$

REGLA DE DECISIÓN :

Rechazar  $H_0$  si  $u > F_{(2,n-3)}^{1-\alpha}$

Sea  $\alpha = 0.05$

$$U = 308.30 > 4.26 = F_{(2,9)}^{0.95} \quad (3.4.9)$$

por lo tanto se rechaza  $H_0$ ; esta prueba es altamente significativa como se esperaba.

Después de haber ajustado una función periódica que modela el fenómeno cíclico en cuestión, el consumo mensual promedio de energía eléctrica, se continuará con la parte de predicción.

Suponga que se desea obtener un intervalo de confianza del 95% para la respuesta media del consumo promedio de energía eléctrica entre el 15 de enero y el 15 de febrero, es decir, se desea obtener un intervalo del 95% de confianza para la respuesta media en un punto  $(C_0, S_0)$ . El ángulo asociado con dicho periodo es  $45^\circ$ , por lo tanto

$$(C_0, S_0) = (0.7071, 0.7071) \quad (3.4.10)$$

recordemos que  $C_0$  y  $S_0$  son los valores de sus respectivas variables regresoras los cuales están sobre el rango de los datos usados para ajustar el modelo, entonces el intervalo buscado es el siguiente :

$$1701.4813 < E(Y|C_0, S_0) < 1845.1179 \quad (3.4.11)$$

Es importante señalar que para este fenómeno se pueden obtener intervalos de confianza para la respuesta media, sólo si se toman en cuenta periodos de 30 días; esto es debido a la forma en que se construyó el modelo.

También se debe señalar que debido a que los ángulos  $\phi_i$  están uniformemente distribuidos sobre una rotación completa de  $360^\circ$ , cualquier valor que tome  $C_0$  y  $S_0$  estará sobre el rango de los datos usados para ajustar el modelo; por lo tanto no tiene sentido el hablar de observaciones futuras.

Finalmente, haciendo inferencia simultánea sobre los parámetros del modelo, se obtiene una *Región Simultánea del 95% de confianza para  $M$ ,  $\beta$  y  $\gamma$*  :

$$2M^2 + \beta^2 + \gamma^2 - 4053.72M + 137.44\beta + 221.14\gamma + 8282853.30 \leq 0 \quad (3.4.12)$$

y una *Región Simultánea del 95% de confianza para  $\beta$  y  $\gamma$*  :

$$\beta^2 + \gamma^2 + 137.44\beta + 221.14\gamma + 66860.44 \leq 0 \quad (3.4.13)$$

### **3.5 Consumo de Energía Eléctrica a Nivel Nacional ( Sector Industrial )**

Los datos que a continuación se presentan corresponden al consumo mensual promedio de energía eléctrica, a nivel nacional, para el sector industrial durante el periodo de enero de 1990 a diciembre de 1995.

<b>MES</b>	<b>CONSUMO DE ENERGÍA ELÉCTRICA ( Miles de millones de watts/hora )</b>
ENE	3219.83
FEB	3062.00
MAR	3311.83
ABR	3300.33
MAY	3432.67
JUN	3461.00
JUL	3489.83
AGO	3552.83
SEP	3522.67
OCT	3582.33
NOV	3553.50
DIC	3486.60

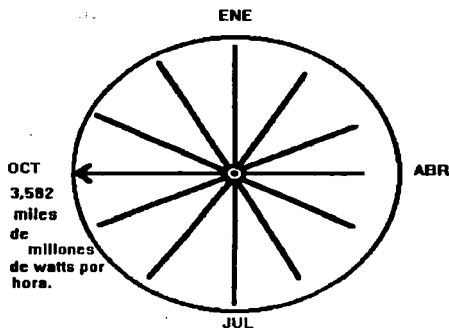
**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

---

si a cada mes se le asocia un ángulo que fluctúa entre  $0^\circ$  y  $360^\circ$ , se tiene lo siguiente :

<b>DIRECCIÓN</b>	<b>GRADOS</b>
ENE	0
FEB	30
MAR	60
ABR	90
MAY	120
JUN	150
JUL	180
AGO	210
SEP	240
OCT	270
NOV	300
DIC	330

de esta forma se puede construir un histograma circular (Ver figura (3.5.1)).



*Figura 3.5.1* Histograma circular para el consumo mensual promedio de energía eléctrica (Sector Industrial) .

La longitud del vector medio,  $r$ , para los meses registrados en este fenómeno es igual a  $3.33 \times 10^{-17}$ , esto confirma que los ángulos asociados a cada mes están uniformemente distribuidos en una rotación completa de  $360^\circ$ ; por lo tanto se puede ajustar una función periódica para modelar este fenómeno (Ver Figura (1.3.7)).

Si se representa este fenómeno cíclico en un plano cartesiano, se observa que efectivamente, una función periódica podría ser un buen modelo (Ver figura (3.5.2)).

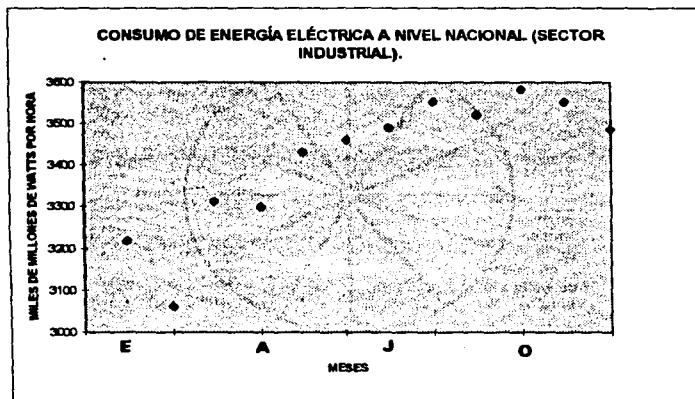


Figura 3.5.2 Consumo mensual promedio de energía eléctrica a nivel nacional (Sector Industrial)

para este caso se tiene que el periodo,  $T$ , es igual a  $360^\circ$ , entonces  $\omega = 1$ , es decir, en el periodo solo habrá un valle y solo habrá una cresta. Con toda la información anterior se obtiene lo siguiente :

$$\bar{Y} = 3414.6183$$

$$S_{ss} = 6$$

$$\bar{C} = -2.7756 \times 10^{-17}$$

$$S_{yy} = 277417.67$$

$$\bar{S} = -1.8504 \times 10^{-17}$$

$$S_{sy} = -827.4496$$

$$\sum_{i=1}^n C_i = -3.3307 \times 10^{-16}$$

$$S_{cs} = 2.7756 \times 10^{-16}$$

**Aplicación a Fenómenos Cíclicos y a Fenómenos de Direcciones.**

---

$$\sum_{i=1}^n S_i = -2.2204 \times 10^{-16}$$

$$S_{cc} = 6$$

$$S_{cy} = -717.9096$$

$$\sum_{i=1}^n e_i^2 = 115807.35$$

A continuación se calculan los estimadores que son útiles para proponer una función que modele a este fenómeno cíclico :

$$\begin{aligned}\hat{\gamma} &= -137.9083 \\ \hat{\beta} &= -119.6510 \\ \hat{M} &= 3414.6183\end{aligned}\tag{3.5.1}$$

por lo tanto la función periódica que se propone para modelar el consumo mensual promedio de energía eléctrica, a nivel nacional, para el sector industrial es :

$$\hat{Y} = 3414.6183 - 119.6510 \cos\phi - 137.9083 \sin\phi\tag{3.5.2}$$

En la figura (3.5.3) se muestra la función periódica ajustada.

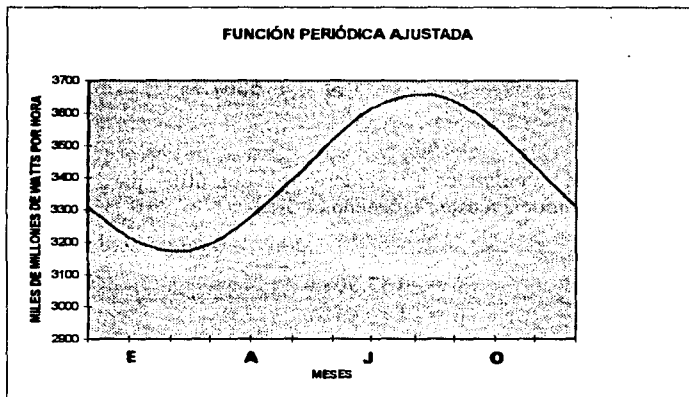


Figura 3.5.3 *Función Periódica Ajustada*

A continuación se presentan los estimadores puntuales para el ángulo de la acrofase y para la amplitud :

$$\hat{\phi}_A = 229.05^\circ \quad (3.5.3)$$

$$\hat{A} = 182.5789$$

Los intervalos de confianza del 95% para los estimadores de los parámetros de la función ajustada, son los siguientes :



$$\begin{aligned}
 3394.6460 < M < 3434.5906 \\
 -217.5225 < \beta < -21.7795 \\
 -235.7798 < \gamma < -40.03676
 \end{aligned}
 \tag{3.5.4}$$

de lo anterior se puede inferir que  $M \neq 0$ ,  $\beta \neq 0$  y  $\gamma \neq 0$ ; esto se confirmará al realizar la prueba de significancia global.

El intervalo de confianza del 95% para  $\sigma^2$  es el siguiente

$$8014.6823 < \sigma^2 < 93593.269 \tag{3.5.5}$$

**Prueba de significancia global**

**H<sub>0</sub>:**  $\beta = \gamma = 0$  vs

**H<sub>a</sub>:** Al menos uno de estos dos parámetros, es diferente de cero.

A continuación se presenta la tabla de análisis de varianza :

	Grados de libertad	Suma de cuadrados	Cuadrados medios	F
<b>Ho</b>	2	23310460.9	11655230.50	1355.13
<b>Error (Modelo completo)</b>	9	77407.35	8600.81	
<b>Total</b>	11	23387868.25		

Tomando  $\alpha = 0.05$ , se tiene

$$1355.13 > 4.26 = F_{(2,9)}^{0.95} \quad (3.5.6)$$

por lo tanto se rechaza  $H_0$ , es decir, el modelo que se ha propuesto tiene sentido. Se puede notar que la prueba es altamente significativa como se esperaba.

A continuación se obtiene el coeficiente de correlación lineal circular muestral

$$R_{y,\phi} = 0.9983 \quad (3.5.7)$$

lo cual indica que en la muestra existe correlación entre un determinado mes del año y el consumo promedio de energía eléctrica que hay en él. Por lo tanto

$$R_{y,\phi}^2 = 0.9966 \quad (3.5.8)$$

es la proporción de la variación total de  $Y$  (consumo mensual promedio de energía eléctrica) que queda explicada por la regresión periódica; en otras palabras, el 72.10% de la variación en el consumo mensual promedio de energía eléctrica es explicada por la variación en el mes de consumo. Esto indica que se tiene un ajuste bastante bueno.

Con lo anterior se puede realizar una prueba de hipótesis para el coeficiente de correlación lineal circular poblacional

$$H_0: \rho_{Y,\phi} = 0 \quad \text{vs} \quad H_a: \rho_{Y,\phi} \neq 0$$

**ESTADÍSTICA DE PRUEBA :**

$$U = \frac{\frac{n-3}{2} R_{y,\phi}^2}{1 - R_{y,\phi}^2}$$

**REGLA DE DECISIÓN :**

Rechazar  $H_0$  si  $u > F_{(2,n-3)}^{1-\alpha}$

Sea  $\alpha = 0.05$

$$U = 1355.13 > 4.26 = F_{(2,9)}^{0.95} \quad (3.5.9)$$

por lo tanto se rechaza  $H_0$ ; esta prueba es altamente significativa como se esperaba.

Después de haber ajustado una función periódica que modela el fenómeno cíclico en cuestión, el consumo mensual promedio de energía eléctrica para el sector industrial, se continuará con la parte de predicción.

Suponga que se desea obtener un intervalo de confianza del 95% para la respuesta media del consumo promedio de energía eléctrica entre el 15 de octubre y el 15 de noviembre, es decir, se desea obtener un intervalo del 95% de confianza para la respuesta media en un punto  $(C_0, S_0)$ . El ángulo asociado con dicho periodo es  $315^\circ$ , por lo tanto

$$(C_0, S_0) = (0.7071, -0.7071) \quad (3.5.10)$$

recordemos que  $C_0$  y  $S_0$  son los valores de sus respectivas variables regresoras los cuales están sobre el rango de los datos usados para ajustar el modelo, entonces el intervalo que buscado es el siguiente :

$$3110.3498 < E(Y|C_0, S_0) < 3744.6902 \quad (3.5.11)$$

Nuevamente es importante señalar, que para este fenómeno se pueden obtener intervalos de confianza para la respuesta media, sólo si se toman en cuenta periodos de 30 días.

También se debe señalar que debido a que los ángulos  $\phi_1$  están uniformemente distribuidos sobre una rotación completa de  $360^\circ$ , cualquier valor que tome  $C_0$  y  $S_0$  estará sobre el rango de los datos usados para ajustar el modelo; por lo tanto no tiene sentido el hablar de observaciones futuras.

Finalmente, haciendo inferencia simultánea sobre los parámetros del modelo, se obtiene una *Región Simultánea del 95% de confianza para  $M$ ,  $\beta$  y  $\gamma$*  :

$$2M^2 + \beta^2 + \gamma^2 - 6829.23M + 119.65\beta + 137.90\gamma + 23203566 \leq 0 \quad (3.5.12)$$

y una *Región Simultánea del 95% de confianza para  $\beta$  y  $\gamma$*  :

$$\beta^2 + \gamma^2 + 119.65\beta + 137.90\gamma + 15063.23 \leq 0 \quad (3.5.13)$$

---

## **CONCLUSIONES**

La elaboración del presente trabajo nos ha llevado a proponer algunos resultados que pueden contribuir al desarrollo y difusión del ANÁLISIS DE REGRESIÓN PERIÓDICA. Estos resultados no son incluidos en este apartado debido a que el desarrollo de los mismos, es presentado en el segundo capítulo.

Junto con algunas conjeturas tomadas de diversos artículos de investigación, hemos integrado un conjunto de resultados que proponemos como herramientas para modelar fenómenos cíclicos y de direcciones, así como para hacer inferencias de la población involucradas en ellos.

En nuestro afán por mostrar como se aplica la teoría del modelo más simple de regresión periódica, hemos encontrado un par de funciones periódicas que pueden ser útiles para modelar el consumo de energía eléctrica en México.

Hemos encontrado además que no todo fenómeno de direcciones puede ser modelado por una función periódica. Sin embargo, esto no implica que no existan fenómenos de direcciones, a los cuales se les pueda modelar y hacer buenas inferencias a través del uso de la teoría de la REGRESIÓN PERIÓDICA.

---

Finalmente diremos que el ANÁLISIS DE REGRESIÓN PERIÓDICA es todavía un área poco desarrollada y difundida, en donde aún faltan muchas cosas por hacer. El presente trabajo es apenas un pequeño intento por voltear los ojos a esta parte de la ESTADÍSTICA MATEMÁTICA que muestra cuestiones de gran interés, y que ha sido un poco olvidada por las personas que estudiamos estadística. En este sentido, se hace una atenta invitación para desarrollar y difundir otros modelos de REGRESIÓN PERIÓDICA.

De esta forma quedan los resultados presentados en este trabajo para su análisis y discusión.

**“POR MI RAZA HABLARÁ EL ESPÍRITU”**

**Ciudad Universitaria, México, D.F., Mayo de 1997.**

---

## ***BIBLIOGRAFÍA***

- [1] Anderson, T.W.(1984). "An Introduction to Multivariate Statistical Analysis". John Wiley & Sons, Toronto, Canada.
- [2] Batschelet, E.(1975). "Introduction to Mathematics for Life Scientist". 2nd edn. Springer, Berlin, Heidelberg, New York.
- [3] Batschelet, E.(1981). "Circular Statistics in Biology". Academic Press Inc. (LONDON) LTD.
- [4] Chatterjee, S. & Price, B.(1977). "Regression Analysis by Example". John Wiley & Sons, New York.
- [5] Draper, N. & Smith, H.(1981). "Applied Regression Analysis". John Wiley & Sons, New York. Second Edition.
- [6] Fort, Ann & Millis, J.N.(1970). "Fitting Sine Curves to 24h Urinary Data". Nature 226, pp. 657-658.
- [7] Johnson, R.A. & Wehrly, T.(1977). "Measures and Models for Angular Correlation and Angular-Linear Correlation". J. Roy. Stat. Soc, Ser. B 39, 222-229.
- [8] Jupp, P. E. & Mardia K. V.(1980). "A General Correlation Coefficient for Directional Data and Related Regression Problems". Biometrika 67, 163-173.

- 
- [9] Liddell, I.G. & Ord, J.K.(1978). "Linear-Circular Correlation Coefficients : Some Further Results". *Biometrika* 65, 448-450.
- [10] Mardia, K.V.(1972). "Statistics of Directional Data". Academic Press London and New York.
- [11] Mardia, K.V.(1975). "Statistics of Directional Data". *Journal Royal Statistics Soc. Serie B*, V.37, pp. 349-393.
- [12] Mardia, K.V.(1976). "Linear-Circular Correlation Coefficients and Rhythmometry". *Biometrika* 63, pp. 403-405.
- [13] Mardia, K.V. et al(1982). "Multivariate Analysis". Academic Press, Inc. London .
- [14] Montgomery, D.C. & Peck, E.A.(1992). "Introduction to Linear Regression Analysis". John Wiley & Sons, New York.
- [15] Mood, Alexander & Graybill, Franklin(1974). "Introduction to the Theory of Statistics". McGraw-Hill International Editions, Singapore.
- [16] Moritz, Robert E.(1972). "Elements of Plane Trigonometry". John Wiley & Sons, New York.
- [17] Schmidt-Koenig, K.(1964). "Über die Orientierung del Vögel". *Naturwissenschaften* 51, 423-431.
- [18] Shashoua, V.E.(1973). "Seasonal Changes in the Learning and Activity Patterns of Goldfish". *Science* 181, 572-574.



- 
- [19] Srivastava, M.S. & Carter, E.M.(1983). "An Introduction to Applied Multivariate Statistics". North-Holland. New York.
- [20] Stephens, M.A.(1979). "Vector Correlation". *Biometrika* 66, 41-48.
- [21] Ware, J.H. & Bowden, R.E.(1977). "Circadian Rhythm Analysis when Output is Collected at Intervals". *Biometrics* 33, 566-571.
- [22] Wonnacott, Thomas & Wonnacott, Ronald.(1981). "Regression : A Second Course in Statistics". Krieger Publishing Company, Malabor, Florida.
- [23] Zar, J.H. (1974). "Biostatistical Analysis". Prentice-Hall, Englewood Cliffs , N.J.

---

## **APÉNDICE. Símbolos**

<b>Y</b>	Variable Dependiente
<b>t</b>	Variable Independiente
<b>M</b>	Nivel Medio de una Función Periódica
<b>A</b>	Amplitud de una Función Periódica
<b><math>\omega</math></b>	Frecuencia Angular de una Función Periódica
<b><math>t_A</math></b>	Acrofase de una Función Periódica
<b>T</b>	Periodo de la Variable Dependiente
<b><math>\phi_A</math></b>	Ángulo de la Acrofase de una Función Periódica
<b><math>\epsilon_i</math></b>	Error Aleatorio
<b><math>\bar{e}_i</math></b>	Vectores Unitarios
<b>m</b>	Vector Medio de la Muestra
<b>R</b>	Longitud del Vector Resultante
<b>r</b>	Longitud del Vector Medio
<b><math>\bar{\phi}</math></b>	Ángulo Medio de la Muestra
<b><math>\mathcal{N}</math></b>	Masa asignada a cada punto en un Diagrama de Dispersión
<b><math>\rho_{Y,\phi}^2</math></b>	Cuadrado del Coeficiente de Correlación Lineal-Circular Poblacional
<b><math>R_{Y,\phi}^2</math></b>	Cuadrado del Coeficiente de Correlación Lineal-Circular Muestral
<b>C</b>	Centro de Masa