



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**MODELOS DE REGRESIÓN PARA DATOS
CIRCULARES**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

**A C T U A R I A
P R E S E N T A:**

NANCY MINERVA TORRES ROJAS



**DIRECTORA DE TESIS:
MAT. MARGARITA ELVIRA CHÁVEZ CANO
2009**



Universidad Nacional
Autónoma de México



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.

HOJA DE DATOS DEL JURADO

1. Datos del alumno

Torres
Rojas
Nancy Minerva
53 42 05 64
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
301032469

2. Datos del tutor

Mat.
Margarita Elvira
Chávez
Cano

3. Datos del sinodal 1

Dra.
Ruth Selene
Fuentes
García

4. Datos del sinodal 2

Act.
Jaime
Vázquez
Alamilla

5. Datos del sinodal 3

M. en C.
Emily
Sánchez
García

6. Datos del sinodal 4

Act.
Francisco
Sánchez
Villarreal

7. Datos del trabajo escrito

Modelos de regresión para datos circulares
50 p
2009

AGRADECIMIENTOS

A mi asesora:

Por su gran capacidad intelectual y su disposición de compartir su conocimiento conmigo.

Por su apoyo incondicional y la confianza que depositó en mí.

Por todas aquellas cosas, que sin ser obligación, ha realizado en mi favor con dedicación y cariño.

Por su excelente calidad humana y la amistad brindada, le agradezco a usted, con cariño y admiración.

A mi madre:

Te agradezco los hábitos de estudio que con esfuerzo y dedicación inculcaste en mí y en mis hermanos desde que éramos niños.

Te agradezco la calidez y sensibilidad para tratar a los demás que nos has enseñado con tu ejemplo, y que sin duda alguna es una herramienta indispensable para las buenas relaciones humanas.

Te agradezco por ser una madre llena de excelentes virtudes, te agradezco con todo mi cariño.

A mi tía Rebe:

Gracias por el cariño que siempre me has demostrado y por el apoyo que me has brindado. Por todo lo que aún sin corresponderte has hecho por mí. Te agradezco con todo mi cariño.

A mis hermanos:

A ustedes que me han demostrado su amor justo en los momentos que más los he necesitado, que me han brindado su apoyo sin importar inconvenientes y adversidades. Los amo, y valoro todos sus actos y muestras de cariño.

A René:

Gracias por tu apoyo, por tu cariño y por ser un hombre diferente y valiente. Te agradezco las vivencias que hemos tenido juntos y que sin duda alguna me han ayudado a crecer como persona. Te amo.

Índice

INTRODUCCIÓN	2
1. REGRESIÓN CIRCULAR-LINEAL (REGRESIÓN PERIÓDICA)	4
1.1. INTRODUCCIÓN	4
1.2. MODELO DE REGRESIÓN PERIÓDICA	8
1.3. AJUSTE DE FUNCIONES PERIÓDICAS NO LINEALES	19
1.4. CORRELACIÓN CIRCULAR-LINEAL	22
2. OTROS MODELOS DE REGRESIÓN CIRCULAR	24
2.1. REGRESIÓN LINEAL-CIRCULAR	24
2.1.1. INTRODUCCIÓN	24
2.1.2. ALGUNOS MODELOS	24
2.2. REGRESIÓN CIRCULAR-CIRCULAR	27
2.2.1. INTRODUCCIÓN	27
2.2.2. ALGUNOS MODELOS	27
3. UNA APLICACIÓN DEL MODELO DE REGRESIÓN PERIÓDICA	29
CONCLUSIONES	47
GLOSARIO	48
ANEXO	49
IDENTIDADES TRIGONOMÉTRICAS	49
FUNCIÓN BESSEL	49
NÚMEROS COMPLEJOS	49

INTRODUCCIÓN

Se le llama Estadística Circular al análisis y estudio de datos que se distribuyen sobre una circunferencia, a este tipo de datos se les conoce como datos circulares. Tales datos se presentan en diversas áreas de estudio como biología, medicina, geología, geografía, meteorología y física. Las observaciones de direcciones sobre un plano y de fenómenos cíclicos pueden ser manejados como datos circulares. Se les llama métodos de regresión circular a todos aquellos modelos de regresión en los que al menos una variable (ya sea la variable de respuesta o alguna de las variables explicativas) es de tipo circular. A las variables circulares también se les conoce como variables angulares, por ello, en este texto ambas se mencionan de manera indistinta.

En este trabajo se desarrollan algunos modelos de regresión circular, incluyendo sólo los modelos en los que se tiene una única variable explicativa. El nombre que se le da a cada tipo de modelo presenta en primer lugar el tipo de variable explicativa y en segundo lugar el tipo de variable de respuesta. Es decir, modelo de regresión circular-lineal (que es el que se desarrolla con mayor detalle a lo largo de este trabajo) se refiere al modelo de regresión donde la variable independiente es circular y la variable dependiente es lineal.

Los métodos de regresión circular han sido usados en diversas aplicaciones, tales como, cristalografía por MacKenzie (1957), vectorcardiografía por Downs *et al.* (1970), predicción de la dirección de un gran movimiento durante un terremoto por Rivest (1997), entre otras.

Específicamente, para el modelo de regresión circular-lineal, las aplicaciones más importantes que se han realizado se encuentran en el área de la Biología conocida como *ritmos biológicos*¹, donde se pretende encontrar una relación entre el tiempo (el cual hace el papel de la variable circular) y algún fenómeno presente en el cuerpo de un ser vivo. Por ejemplo, se puede intentar encontrar una relación entre la hora del día y la concentración de cierta hormona en un organismo. Estos temas ya han sido abordados por distintos autores, por ejemplo: estudios de correlaciones entre *ritmos biológicos circadianos*², donde un periodo de 24hrs es considerado como los 360° de una circunferencia (Binkley, 1990; Downs, 1974; More-Ede *et al.*, 1982); algunas aplicaciones médicas en los periodos circadianos del tratamiento de quimioterapia para el cáncer, con el fin de reducir el número y severidad de sus efectos tóxicos colaterales (Hrushesky, 1985); etc. Es importante hacer la diferenciación entre el concepto de *ciclo* y el concepto de *ritmo*, para ello, se recomienda al lector buscar estos términos en el Glosario.

Uno de los más antiguos modelos de regresión para una variable de respuesta circular y una o varias variables lineales fue publicado por Gould (1969). Mardia (1975) desarrolló un coeficiente de correlación de tipo no paramétrico para datos circulares. Johnson y Wehrly (1978) mejoraron el modelo de Gould. Stephens (1979) adaptó un esquema de regresión direccional

¹Ver Glosario

²Ver Glosario

para datos espaciales de rocas magnéticas. Fisher y Lee (1992) generalizaron los modelos de Johnson y Wehrly por medio de una función liga que proyecta la recta real sobre el círculo unitario. Recientemente Lund (1999) desarrolló un modelo de regresión donde las variables independientes consisten en una variable de tipo circular y un conjunto de variables lineales, y Follman y Proschan (1999) desarrollaron una prueba de uniformidad circular circadiana de los tiempos en una crisis epiléptica usando tiempos de crisis sucesivos correlacionados sobre los mismos individuos.

A pesar de que la teoría existente es de gran utilidad, trabajos recientes sobre la genética y aspectos moleculares de ritmos circadianos en mamíferos (Lowrey *et al.*, 2000; Shearman *et al.*, 2000), sugieren la necesidad de modelos de regresión circular más poderosos. Es por esto, que es de suma importancia la investigación y desarrollo de estos temas.

1. REGRESIÓN CIRCULAR-LINEAL (REGRESIÓN PERIÓDICA)

1.1. INTRODUCCIÓN

Como se sabe, en el modelo de Regresión Lineal se requiere que tanto la variable dependiente como la independiente sean lineales. En la Regresión Periódica la variable independiente debe ser de tipo circular, como puede ser un ángulo o el tiempo en el caso de fenómenos periódicos, con lo que respecta a la variable dependiente, en esta regresión debe ser de tipo lineal. A este tipo de regresión también se le conoce como Regresión Circular-Lineal.

Dado que la variable independiente es de tipo circular, tiene implícito un periodo. Para desarrollar el modelo de regresión, se supone que dicho periodo es conocido (por ejemplo 360° o 24 hrs.), por lo tanto, no se tiene el problema de estimar este periodo.

Se ajusta un modelo de regresión cuando se sospecha que existe dependencia de una variable con respecto a otra u otras, con el objetivo de encontrar una expresión (es decir, una función) que permita predecir el valor de la variable dependiente, teniendo el valor de la(s) variable(s) independiente(s). Para lo anterior, lo primero que se hace es suponer el modelo, después se utilizan datos observados (tanto de la variable dependiente como de la(s) variable(s) independiente(s)) para estimar los parámetros desconocidos del modelo supuesto, una vez que se encuentran estos parámetros se analiza si el modelo ajusta bien a los datos o no, es decir, si sirve o no para predecir.

El modelo más simple de un fenómeno periódico es una curva cosenoidal dada por la siguiente ecuación:

$$y = M + A \cos(w(t - t_0)) \quad (1.1.1)$$

donde t es la variable independiente, generalmente el tiempo, sujeta a un periodo conocido T . La ecuación contiene cuatro parámetros que son:

M : nivel medio o mesor

A : amplitud ($A \geq 0$)

w : frecuencia angular

t_0 : fase pico o acrofase

Para entender el significado de cada uno de los parámetros se recomienda buscar cada término en el Glosario. Además, se puede analizar la ecuación (1.1.1) parte por parte:

$\cos(w(t - t_0))$: Se sabe que la función $\cos(\theta)$ toma valores que oscilan entre -1 y 1, por lo tanto, sea cual sea el valor de w , t y t_0 , la función $\cos(w(t - t_0))$ va a tomar valores entre -1 y 1 (esto se debe a que se puede ver a θ como $w(t - t_0)$). También se sabe que el valor

promedio de los valores que toma la función $\cos(\theta)$ es cero, por ello, el nivel medio (es decir, el mesor) de los valores que toma la función $\cos(w(t - t_0))$ es también cero. La amplitud de $\cos(\theta)$ es uno, por ser esta la distancia del mesor (que es cero) al valor máximo de la función (que es uno), por tanto, al tener el mismo mesor y valor máximo, la amplitud de la función $\cos(w(t - t_0))$ es también uno. Se sabe también que la función $\cos(\theta)$ se maximiza en $\theta = 0$, así que si se toma a θ como $w(t - t_0)$ se debe encontrar el valor de t que satisfaga la ecuación $w(t - t_0) = 0$, dado que $w \neq 0$ se debe cumplir $t - t_0 = 0$, por lo tanto, el valor que maximiza la función es $t = t_0$.

$A \cos(w(t - t_0))$: Al multiplicar la función $\cos(w(t - t_0))$ por un número real A , los valores de esta nueva función oscilan entre $-A$ y A , nuevamente el promedio de los valores que toma la función es claramente cero, por lo que, la amplitud es ahora A . Dado que el valor máximo que puede tomar la función es A , se debe encontrar en qué valor de la variable t se alcanza este máximo, es decir, en qué valor de t la función $\cos(w(t - t_0))$ vale uno; como ya se sabe la función $\cos(\theta)$ alcanza el valor de uno cuando $\theta = 0$, por lo que $w(t - t_0)$ debe ser igual a cero, por lo tanto, el máximo de $A \cos(w(t - t_0))$ se alcanza en $t = t_0$.

$M + A \cos(w(t - t_0))$: Esta función que es ya la del modelo difiere de la anterior sólo en la adición del valor M , lo que hace que la gráfica de la función se desplace M unidades. Dado que a todos los valores se les aumentó la misma cantidad, el promedio de los valores que toma la función es el mesor de la función anterior más M , es decir, el mesor resulta ser ahora M . Evidentemente el valor máximo de la función es ahora $A + M$, pero este valor máximo se sigue alcanzando en $t = t_0$ ya que los valores de la variable independiente no tuvieron ninguna alteración. Ahora, fijándose en la amplitud se puede observar que ésta sigue siendo A , eso se debe a que la amplitud es la distancia entre dos valores de la función (el mesor y el valor máximo) y como estos valores también se desplazaron M unidades, la distancia entre ellos sigue siendo la misma.

La frecuencia angular w está relacionada al periodo T por la siguiente ecuación:

$$w = 2\pi/T \quad \text{o} \quad 360^\circ/T \quad (1.1.2)$$

dependiendo si se están manejando grados o radianes.

Dado que el periodo T es conocido, la frecuencia angular w también es conocida. Se puede interpretar a w como el número de veces que se encuentra contenido el periodo en una rotación completa.

La acrofase t_0 es medida en las mismas unidades que la variable independiente t . Dado que la variable t es periódica módulo T , también $t_0 \pm T, t_0 \pm 2T, t_0 \pm 3T$, etc, pueden servir como acrofase, sin embargo, por practicidad se limita la acrofase al intervalo de 0 a T . Es decir, el valor que se toma en cuenta para el modelo es t_0 .

En la Figura 1.1.1 se encuentra una gráfica de la función (1.1.1), en la que se pueden apreciar cada uno de sus parámetros.

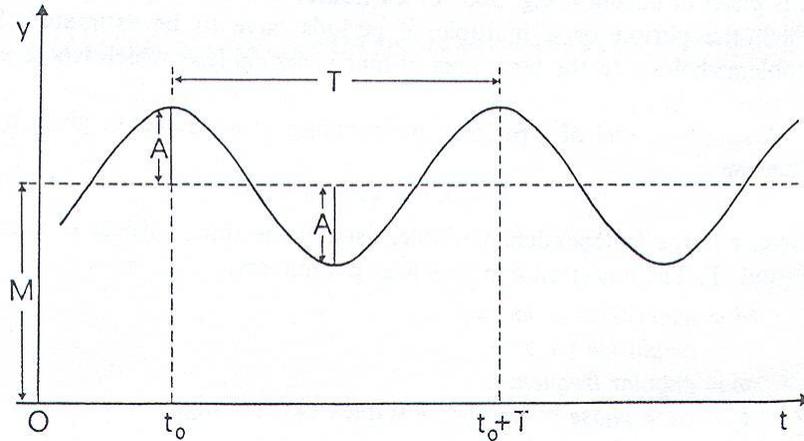


Figura 1.1.1

Debido a que el modelo más simple para este tipo de regresión, ecuación (1.1.1), contiene la función $\cos(\theta)$, se puede intuir que su gráfica no resultará una línea recta, lo cual se confirma en la Figura 1.1.1. Entonces surge la pregunta de por qué se le llama variable lineal a la variable dependiente y , la respuesta es muy simple, en Estadística Circular a las variables que no son circulares se les llama lineales sólo para distinguir unas de otras. Es decir, lo único que se toma en cuenta para su clasificación es la presencia o ausencia de un ciclo.

Se puede reescribir la ecuación (1.1.1) como sigue:

$$g(t) = M + A \cos(\omega t - \phi) \quad (1.1.3)$$

donde y está representada por $g(t)$ y ωt_0 por ϕ .

Obsérvese que si a la expresión $\omega = 2\pi/T$ se le multiplica por t_0 , el resultado (ωt_0) queda expresado en radianes, ya que las unidades de T y t_0 son las mismas y se eliminan, lo mismo pasa si en lugar de 2π se tiene 360° quedando el resultado expresado en grados. Por esta razón a ϕ se le llama ángulo acrofase, es decir, es el equivalente a t_0 pero expresado en ángulos o radianes. Por lo tanto, ϕ es la fase angular donde $g(t)$ alcanza su valor máximo.

Cuando el conjunto de datos $y_1, y_2, y_3, \dots, y_n$ en los instantes de tiempo $t_1, t_2, t_3, \dots, t_n$ están ya dados (es decir, fueron observados), a la ecuación (1.1.1) debe agregársele un término error. Para entender el por qué de este término, basta hacer la siguiente observación: a pesar de introducir el valor t_i observado en dicha ecuación, lo más probable es que el valor arrojado no sea el valor y_i observado sino algo muy similar. Esto se debe a que hay cuestiones en la realidad que no se pueden controlar tales como los errores de medición de los datos. Por lo tanto, el modelo con los valores observados queda de la siguiente manera:

$$y_i = M + A \cos(\omega(t_i - t_0)) + \epsilon_i \quad (1.1.4)$$

Los errores ϵ_i son variables aleatorias y se deben por lo general, como ya se dijo, a inexactitudes de medición. En algunos casos puede suponerse que los errores son independientes y normalmente distribuidos.

Se ha hablado ya de tres modelos diferentes: el *modelo teórico*, ecuación (1.1.1), el cual representa la relación que se supone entre la variable independiente y la dependiente; el *modelo muestral* que es el que se encuentra representado por la ecuación (1.1.4), en el cual se incluyen los datos observados de cada una de las variables; y por último el *modelo ajustado*, que es en el que se incluyen los estimadores de los parámetros desconocidos del modelo teórico y el cual es representado con la siguiente ecuación:

$$\hat{y}_i = \hat{M} + \hat{A} \cos(w(t_i - \hat{t}_0))$$

Los estimadores en el modelo ajustado \hat{M} , \hat{A} y \hat{t}_0 , serán calculados con los datos observados t_i y y_i . Nótese que \hat{y}_i resulta un valor estimado ya que está siendo calculado con los tres estimadores primero mencionados, la razón de ponerle gorrito es sólo para distinguir que \hat{y}_i es el valor que arroja el modelo ajustado y no el valor observado y_i .

En resumen, se tiene lo siguiente:

$$\begin{array}{ll} y = M + A \cos(w(t - t_0)) + \epsilon & \text{modelo teórico} \\ y_i = M + A \cos(w(t_i - t_0)) + \epsilon_i & \text{modelo muestral} \\ \hat{y}_i = \hat{M} + \hat{A} \cos(w(t_i - \hat{t}_0)) & \text{modelo ajustado} \end{array}$$

Aunque el modelo (1.1.1) es el más importante en la regresión periódica, más adelante se discutirán otros modelos que ofrecen un mejor ajuste de los datos. La generalización del modelo (1.1.3) está dada por la siguiente ecuación:

$$g(t) = M + A_1 \cos(wt - \phi_1) + A_2 \cos(2wt - \phi_2) + \dots + A_k \cos(kwt - \phi_k) \quad (1.1.5)$$

La función anterior es llamada polinomio trigonométrico o suma parcial de una serie de Fourier. Ésta contiene las frecuencias angulares $w, 2w, \dots, kw$, las cuales son múltiplos de w . Además del periodo original T también se encuentran implícitos periodos más pequeños $T/2, T/3, \dots, T/k$ los cuales se encuentran exactamente dentro del periodo original. A los términos $A_1 \cos(wt - \phi_1)$, $A_2 \cos(2wt - \phi_2)$, etc. se les llama primero, segundo, etc. armónicos.

Cuando se tiene dado un conjunto de datos $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$ se puede encontrar siempre un polinomio trigonométrico que ajuste a los datos, siempre y cuando, los valores t_i de la variable independiente sean diferentes. Se debe encontrar este polinomio siguiendo los

siguientes criterios:

a) El modelo debe ajustar a los datos con suficiente precisión. Sin embargo, el precio que se tiene que pagar por ello es un gran número de términos, lo cual implica también un gran número de parámetros que posteriormente se tendrán que estimar.

b) El modelo debe ser lo más suave posible, es decir, que la gráfica no debe cambiar frecuentemente o abruptamente.

c) El número de parámetros debe ser lo más pequeño posible, ya que, algunos de ellos son difíciles de estimar.

Es raro que los tres puntos anteriores sean cumplidos al mismo tiempo. Además, mientras que para el primer punto es favorable tener un gran número de parámetros, para el segundo y el tercero esto resulta de poca ayuda. El recurso más frecuentemente usado consiste en restringir la ecuación (1.1.5) al primer armónico ($k = 1$) lo cual actúa en favor del segundo y tercer punto, pero desfavorablemente para el primero. En este caso el polinomio trigonométrico de la ecuación (1.1.5) se reduce a la ecuación (1.1.3).

1.2. MODELO DE REGRESIÓN PERIÓDICA

Comenzando con el caso más simple, es decir, el de la ecuación (1.1.3) y utilizando la identidad trigonométrica de la expresión (3.0.6) del Anexo se obtiene la siguiente función:

$$g(t) = M + A \cos(wt) \cos(\phi) + A \sin(wt) \sin(\phi)$$

La cual se puede reescribir como:

$$g(t) = M + X \cos(wt) + Y \sin(wt) \quad (1.2.1)$$

Haciendo:

$$X = A \cos(\phi), \quad Y = A \sin(\phi) \quad (1.2.2)$$

Puede observarse que, dado que la amplitud A y el ángulo acrofase ϕ son parámetros, X y Y también son parámetros y no variables.

Se debe hacer la observación de que el modelo representado por la ecuación (1.2.1) es un modelo teórico, ya que es el mismo modelo de la ecuación (1.1.1) pero con algunas expresiones renombradas, debido a esto, se debe trabajar también con una expresión nueva para los modelos muestral y ajustado.

Por lo anterior, de nuevo se tienen tres modelos, que vienen siendo equivalentes a los tres modelos anteriormente mencionados:

$$\begin{aligned}
g(t) &= M + X \cos(wt) + Y \operatorname{sen}(wt) + \epsilon && \text{modelo teórico} \\
g(t_i) &= M + X \cos(wt_i) + Y \operatorname{sen}(wt_i) + \epsilon_i && \text{modelo muestral} \\
\hat{g}(t_i) &= \hat{M} + \hat{X} \cos(wt_i) + \hat{Y} \operatorname{sen}(wt_i) && \text{modelo ajustado}
\end{aligned}$$

Siguiendo con el análisis del modelo (1.2.1), usando la identidad trigonométrica de la expresión (3.0.7) del Anexo se obtiene la siguiente expresión:

$$X^2 + Y^2 = A^2 \cos^2(\phi) + A^2 \operatorname{sen}^2(\phi) = A^2(\cos^2(\phi) + \operatorname{sen}^2(\phi)) = A^2$$

$$\text{Por lo tanto, } A = (X^2 + Y^2)^{1/2} \quad (1.2.3)$$

Utilizando la ecuación (3.0.8) del Anexo se puede escribir lo siguiente:

$$Y/X = A \operatorname{sen}(\phi)/A \cos(\phi) = \operatorname{sen}(\phi)/\cos(\phi) = \tan(\phi)$$

De lo anterior se deduce que:

$$\phi = \arctan(Y/X)$$

Pero hay algo que se debe de considerar para esta última expresión, el codominio de la función $\arctan(x)$, que es el intervalo $[-\pi/2, \pi/2]$. Por lo tanto, al calcular ϕ con esta expresión siempre va a resultar en este intervalo. Sin embargo, el parámetro ϕ de nuestro modelo no está restringido. La manera de saber fácilmente si ϕ está o no en dicho intervalo es con el signo de X . Si $X = A \cos(\phi) > 0$ significa que $\cos(\phi) > 0$ (ya que $A \geq 0$) lo que implica que ϕ está entre $-\pi/2$ y $\pi/2$ sin tomar a los extremos, por otro lado, si $X < 0$ entonces $\cos(\phi) < 0$ lo que significa que ϕ está entre $\pi/2$ y $3\pi/2$ también sin tomar a los extremos.

Por lo anterior, si $X < 0$, el ángulo ϕ no está en el codominio de la función $\arctan(x)$, así que, se puede concluir:

$$\begin{aligned}
\phi &= \{\arctan(Y/X) && \text{si } X > 0 \\
&180^\circ + \arctan(Y/X) && \text{si } X < 0
\end{aligned} \quad (1.2.4)$$

Se puede observar que si $X = 0$ queda indeterminada la ecuación (1.2.4) y se tendrían dos opciones: $\phi = \pi/2$ o $\phi = 3\pi/2$. En este caso, se puede saber el valor de ϕ fijándose en el valor de Y ; si $Y > 0$, el ángulo ϕ debe ser igual a $\pi/2$, en cambio, si $Y < 0$, el ángulo ϕ debe ser $3\pi/2$.

La ecuación (1.2.1) cuenta con cuatro parámetros M, X, Y, w . Sin embargo, sólo se deben estimar a tres de ellos: M, X y Y ; dado que la frecuencia angular w se supone conocida. Dicha ecuación es lineal en estos tres parámetros, por lo que el procedimiento de estimación

se facilita en gran medida. Lo anterior, también es la razón de que el modelo (1.2.1) sea un modelo de regresión lineal.

El método más usado para estimar a los parámetros, es el método de los mínimos cuadrados, que es el que se utilizará en este desarrollo.

Sean $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$ los puntos dados y $g(t_i)$ los valores arrojados por el modelo ajustado, el valor ajustado del modelo teórico que está representado por la ecuación (1.2.1), en los valores t_1, t_2, \dots, t_n de la variable independiente. Obsérvese que al ser $g(t_i)$ el valor del modelo ajustado antes mencionado, debería aparecer como $(\hat{g}(t_i))$, pero en este caso no es necesario hacer esta distinción ya que al valor del modelo ajustado en el valor t_i se le está llamando $g(t_i)$ que no se confunde con el valor observado y_i . Por lo tanto, aunque se haga referencia a $(\hat{g}(t_i))$ en adelante se utilizará sólo $g(t_i)$ por practicidad. De esta manera se definen los residuales:

$$\delta_i = g(t_i) - y_i \quad (1.2.5)$$

Por como está definida δ_i , ésta es la desviación que hay entre el valor del modelo ajustado en la abscisa t_i y el valor observado.

Por lógica se puede deducir que el mejor modelo $g(t)$, debe ser aquel que minimice la suma de estas desviaciones (es decir, que minimice $\sum_{i=1}^n \delta_i$). Para ello, se debe tomar en cuenta que δ_i puede tomar valores negativos y positivos, dependiendo si $g(t_i)$ es mayor a y_i o viceversa. Esto puede conducir a tomar un modelo cuya $\sum_{i=1}^n \delta_i$ sea pequeña y cada una de las δ_i sea muy grande en su valor absoluto, es claro que un modelo así no estaría acercándose a los datos. Por lo tanto, se debe minimizar una función que tenga que ver con la suma de las desviaciones y que a su vez cada una de ellas no tome valores negativos.

Por lo anterior, la función que se toma en cuenta para minimizar es:

$$S = \sum_{i=1}^n \delta_i^2$$

Por otro lado, de la ecuación (1.2.1) se obtiene:

$$g(t_i) = M + X \cos(wt_i) + Y \operatorname{sen}(wt_i)$$

Si se recuerda que la expresión anterior se refiere al modelo ajustado, M, X y Y deberían estar expresados como \hat{M}, \hat{X} y \hat{Y} , pero para fines prácticos en las siguientes líneas se omite el gorrito en todos los estimadores.

Como la frecuencia angular w es conocida, se puede conocer también el valor de los coeficientes c_i y s_i definidos como sigue:

$$c_i = \cos(wt_i), \quad s_i = \text{sen}(wt_i) \quad (1.2.6)$$

Con lo anterior se puede reescribir a la ecuación (1.2.5) de la siguiente forma:

$$\delta_i = M + Xc_i + Ys_i - y_i \quad (1.2.7)$$

Obsérvese que en la ecuación anterior M , X y Y son ya los estimadores que se desean encontrar. Con (1.2.7) se puede reescribir a la función S (que es la que se debe minimizar) como:

$$S = \sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n (M + Xc_i + Ys_i - y_i)^2 \quad (1.2.8)$$

Ahora, con la ayuda de las herramientas del cálculo diferencial se puede encontrar la expresión para los estimadores M , X y Y que minimizan a S .

Primero se deriva S con respecto a X :

$$\begin{aligned} \frac{\partial S}{\partial X} &= \frac{\partial \sum (M + Xc_i + Ys_i - y_i)^2}{\partial X} \\ &= \sum \frac{\partial (M + Xc_i + Ys_i - y_i)^2}{\partial X} \\ &= \sum 2(M + Xc_i + Ys_i - y_i)c_i \\ &= 2 \sum (Mc_i + Xc_i^2 + Ys_i c_i - y_i c_i) \\ &= 2(M \sum c_i + X \sum c_i^2 + Y \sum s_i c_i - \sum y_i c_i) \end{aligned}$$

Ahora se deriva S con respecto a Y :

$$\begin{aligned} \frac{\partial S}{\partial Y} &= \frac{\partial \sum (M + Xc_i + Ys_i - y_i)^2}{\partial Y} \\ &= \sum \frac{\partial (M + Xc_i + Ys_i - y_i)^2}{\partial Y} \\ &= \sum 2(M + Xc_i + Ys_i - y_i)s_i \\ &= 2 \sum (Ms_i + Xc_i s_i + Ys_i^2 - y_i s_i) \\ &= 2(M \sum s_i + X \sum c_i s_i + Y \sum s_i^2 - \sum y_i s_i) \end{aligned}$$

Por último, se deriva S con respecto a M :

$$\begin{aligned}
\frac{\partial S}{\partial M} &= \frac{\partial \sum (M + Xc_i + Ys_i - y_i)^2}{\partial M} \\
&= \sum \frac{\partial (M + Xc_i + Ys_i - y_i)^2}{\partial M} \\
&= \sum 2(M + Xc_i + Ys_i - y_i) \\
&= 2 \sum (M + Xc_i + Ys_i - y_i) \\
&= 2(nM + X \sum c_i + Y \sum s_i - \sum y_i)
\end{aligned}$$

Los valores de M , X y Y que minimizan a la función S son aquellos que satisfacen el siguiente sistema de ecuaciones:

$$\begin{aligned}
2(M \sum c_i + X \sum c_i^2 + Y \sum s_i c_i - \sum y_i c_i) &= 0 \\
2(M \sum s_i + X \sum c_i s_i + Y \sum s_i^2 - \sum y_i s_i) &= 0 \\
2(nM + X \sum c_i + Y \sum s_i - \sum y_i) &= 0
\end{aligned} \tag{1.2.9}$$

Se puede omitir a la constante 2 de las ecuaciones anteriores por dos razones. La primera, porque al dividir entre 2 a las tres ecuaciones, en el lado izquierdo de las ecuaciones el 2 desaparece y en el lado derecho se vuelve a tener cero. Y la segunda, porque al tener dos factores igualados a cero se tienen tres opciones: ambos son cero, el primero es cero y el segundo distinto de cero o el segundo es cero y el primero distinto de cero; claramente 2 es distinto de cero por lo que en las tres ecuaciones quedan los otros factores igualados a cero. Por lo anterior, se obtiene el sistema:

$$\begin{aligned}
M \sum c_i + X \sum c_i^2 + Y \sum s_i c_i &= \sum y_i c_i \\
M \sum s_i + X \sum c_i s_i + Y \sum s_i^2 &= \sum y_i s_i \\
nM + X \sum c_i + Y \sum s_i &= \sum y_i
\end{aligned} \tag{1.2.10}$$

Las ecuaciones anteriores son conocidas como *ecuaciones normales*. Se puede observar que estas ecuaciones son lineales en los estimadores M , X y Y por lo que la solución de éstas es relativamente sencilla.

Para simplificar el sistema (1.2.10) se definen las siguientes constantes:

$$\begin{aligned}
a &= \sum c_i & b &= \sum c_i^2 & c &= \sum s_i c_i \\
d &= \sum y_i c_i & e &= \sum s_i & f &= \sum s_i^2 \\
g &= \sum y_i s_i & k &= n & p &= \sum y_i
\end{aligned}$$

Obteniendo así el siguiente sistema:

$$\begin{aligned}
aM + bX + cY &= d \\
eM + cX + fY &= g \\
kM + aX + eY &= p
\end{aligned} \tag{1.2.11}$$

Puede encontrarse la solución del sistema anterior con diversos métodos matemáticos, sin embargo, con cualquiera de ellos hay que seguir primero una gran cantidad de pasos antes de llegar a la solución. Por esta razón, se omiten dichos pasos y se escriben a continuación las expresiones de cada estimador, obtenidas al resolver el sistema con el método de eliminación sucesiva.

$$\begin{aligned}
M &= \frac{-afd + ecd - bge + bfp - pc^2 + cag}{-a^2f + 2aec + fkb - e^2b - kc^2} \\
X &= \frac{age - cgk - afp + fkd + ecp - e^2d}{-a^2f + 2aec + fkb - e^2b - kc^2} \\
Y &= \frac{apc - peb + gkb - ckd - a^2g + aed}{-a^2f + 2aec + fkb - e^2b - kc^2}
\end{aligned}$$

Regresando a la notación de los tres modelos la cual se tenía en un principio, empleada para distinguir a los parámetros de sus estimadores, se obtiene:

$$\begin{aligned}
\hat{M} &= \frac{-afd + ecd - bge + bfp - pc^2 + cag}{-a^2f + 2aec + fkb - e^2b - kc^2} \\
\hat{X} &= \frac{age - cgk - afp + fkd + ecp - e^2d}{-a^2f + 2aec + fkb - e^2b - kc^2} \\
\hat{Y} &= \frac{apc - peb + gkb - ckd - a^2g + aed}{-a^2f + 2aec + fkb - e^2b - kc^2}
\end{aligned} \tag{1.2.12}$$

Al sustituir el valor de cada una de las constantes anteriores se obtiene al fin la solución del sistema (1.2.10), con esta solución se puede regresar a la ecuación (1.2.1) obteniendo así el modelo buscado. En ocasiones quedarse con el modelo (1.2.1) es suficiente, pero en otros casos, cuando el valor de A y ϕ es requerido se debe encontrar el modelo (1.1.3) y para ello se pueden obtener fácilmente los valores de sus parámetros con las ecuaciones (1.2.3) y (1.2.4).

Desde un punto de vista meramente matemático los valores $t_1, t_2, t_3, \dots, t_n$ pueden ser escogidos arbitrariamente. Estos instantes de tiempo no están tampoco restringidos a un mismo intervalo (por ejemplo de 0 a T) incluso algunos de ellos pueden coincidir, sin embargo, para obtener resultados significativos se debe escoger a estos instantes de tiempo igualmente espaciados sobre un periodo T completo. Si lo anterior no resulta práctico se pueden escoger los valores $t_1, t_2, t_3, \dots, t_n$ lo suficientemente densos sobre el eje de tiempo t . En el primer caso, un mínimo de datos $n = 6$ resulta suficiente para un modelo razonable, en cambio, cuando los instantes de tiempo no pueden ser escogidos espaciados igualmente se requiere un mínimo de datos $n = 8$ (sin olvidar que en este caso deben ser razonablemente densos).

La Figura 1.2.1 ilustra algunos casos en los que se puede decir si el modelo a obtener (según el número y distribución de sus datos) resultará significativo o no.

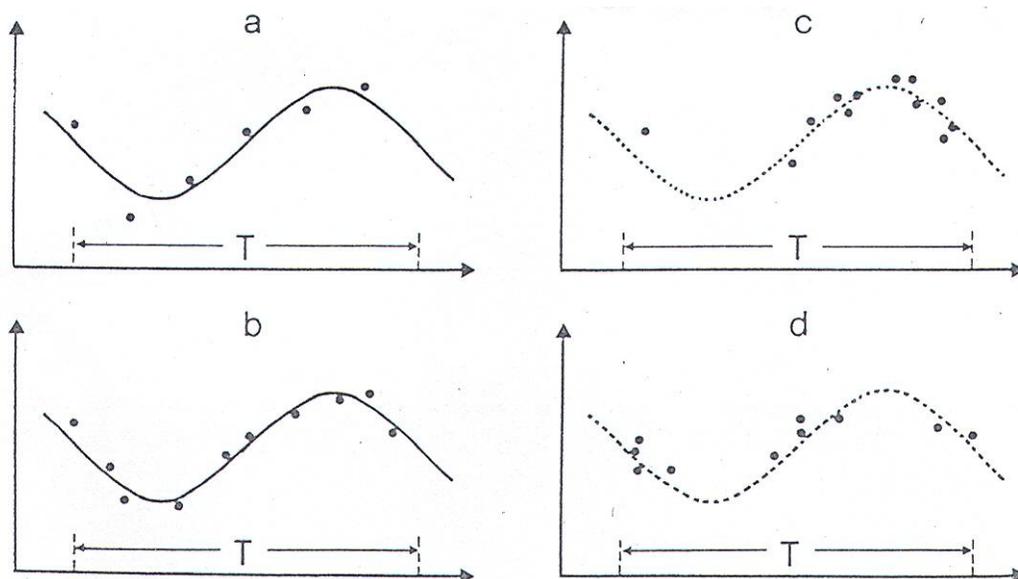


Figura 1.2.1

En la gráfica a) los instantes de tiempo se encuentran igualmente espaciados ocupando un periodo T completo, cumpliendo también con un mínimo de 6 datos, por lo que se puede concluir que el modelo a obtener sí será útil. Para la gráfica b) se tienen 10 datos razonablemente densos sobre todo el periodo T por lo que también se espera un modelo significativo. En c) no se tienen instantes de tiempo igualmente espaciados por lo que deben estar lo suficientemente densos sobre el eje de tiempo, sin embargo, se encuentran saturados en una zona del periodo T y dispersos en la otra, por lo que a pesar de cumplir con el mínimo de 8 datos no se pueden obtener parámetros confiables para el modelo. Por último, en d) tampoco se puede esperar un modelo realmente predictivo, por las mismas razones que en c).

La calidad del ajuste (es decir, del modelo) se intensifica con una gran cantidad de datos, siempre y cuando éstos se encuentren dispersos sobre el periodo T .

Hay ocasiones en que un modelo con primero y segundo armónico ajusta mejor a los datos que el modelo que se ha estudiado, en estos casos el procedimiento para encontrar los estimadores es análogo. Con la ayuda de las ecuaciones (1.2.1) y (1.2.2) se obtiene:

$$g(t) = M + A_1 \cos(\omega t - \phi_1) + A_2 \cos(2\omega t - \phi_2) \quad (1.2.13)$$

Definiendo:

$$\begin{aligned} X_1 &= A_1 \cos(\phi_1) & Y_1 &= A_1 \sin(\phi_1) \\ X_2 &= A_2 \cos(\phi_2) & Y_2 &= A_2 \sin(\phi_2) \end{aligned}$$

Podemos reescribir la ecuación (1.2.13) como sigue:

$$g(t) = M + X_1 \cos(\omega t) + Y_1 \sin(\omega t) + X_2 \cos(2\omega t) + Y_2 \sin(2\omega t)$$

Recuérdese que la frecuencia angular w es conocida, por lo que los parámetros que habría que estimar con este modelo son: M, X_1, X_2, Y_1, Y_2 . El procedimiento para el cálculo de estos parámetros es el mismo que se siguió con el modelo anterior, sólo que en este caso el sistema de ecuaciones obtenido será de 5 ecuaciones con 5 incógnitas.

Ejemplo 1

En 1974, Descovich estudió el ritmo circanual de la secreción de norepinefrina en la orina de seres humanos sanos. La norepinefrina es una de las hormonas conocidas como catecolaminas (adrenalina, dopamina y norepinefrina), las sustancias de este grupo pueden ser producidas en las glándulas suprarrenales ejerciendo una función hormonal, o ser producidas en las terminaciones nerviosas por lo que se les considera también neurotransmisores. Las catecolaminas están asociadas al estrés y a la obesidad.

En la Figura 1.2.2 se describen promedios mensuales de esta secreción. Para obtener estimadores de los parámetros (mesor, amplitud y acrofase) se ajusta un modelo cosenoidal.

Se tienen $n = 12$ datos, uno por cada mes del año. Estos datos están tomados a la mitad de cada mes. En la mitad de enero, en la mitad de febrero, etc. Se pueden asociar estos instantes de tiempo con los ángulos $0^\circ, 30^\circ, 60^\circ, \dots, 330^\circ$, respectivamente. El periodo T es un año o 360° , debido a que se está manejando el tiempo con grados, la frecuencia angular está dada por:

$$w = 360^\circ / T = 1$$

y las abscisas son $t_1 = 0^\circ, t_2 = 30^\circ, t_3 = 60^\circ, \dots, t_{12} = 330^\circ$.

Se pueden calcular las cantidades auxiliares c_i y s_i por medio de la ecuación (1.2.6) para cada una de las abscisas, en la Tabla 1.2.1 se encuentran los instantes de tiempo t_i , su respectiva c_i y s_i y los datos y_i observados para cada instante de tiempo.

Debido a que los instantes de tiempo son igualmente espaciados en todo un periodo completo de 360° se puede intuir que $\sum c_i$ y la $\sum s_i$ serán cero.

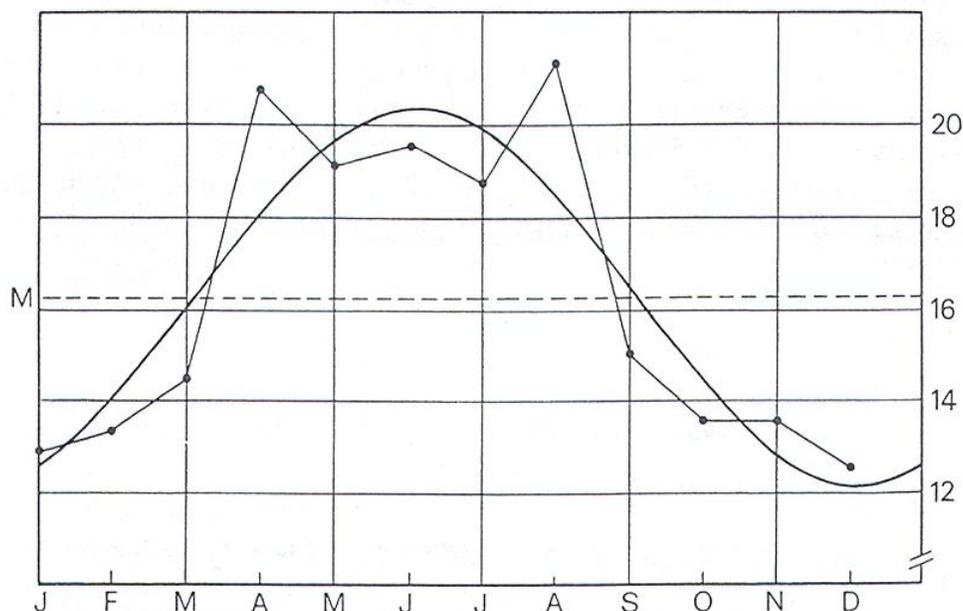


Figura 1.2.2

Month	t_i	c_i	s_i	y_i^\dagger
J	0°	1	0	12.9
F	30°	0.866	0.500	13.3
M	60°	0.500	0.866	14.4
A	90°	0	1	20.8
M	120°	-0.500	0.866	19.1
J	150°	-0.866	0.500	19.6
J	180°	-1	0	18.8
A	210°	-0.866	-0.500	21.3
S	240°	-0.500	-0.866	15.0
O	270°	0	-1	13.6
N	300°	0.500	-0.866	13.6
D	330°	0.866	-0.500	12.6
Total	—	0	0	195.0

Tabla 1.2.1

Con los datos anteriores se encuentran los siguientes valores:

$$\begin{array}{lll} \sum c_i = 0 & \sum s_i = 0 & \sum s_i c_i = 0 \\ \sum c_i^2 = 6 & \sum s_i^2 = 6 & \sum y_i = 195 \\ \sum y_i c_i = -21.94 & \sum y_i s_i = 10.94 & \end{array}$$

Con estos resultados las ecuaciones normales (1.2.10) se vuelven extremadamente simples:

$$6X = -21.94 \quad 6Y = 10.94 \quad 12M = 195$$

De este modo, se obtienen los siguientes parámetros:

$$X = -3.66 \quad Y = 1.82 \quad M = 16.3$$

Una vez teniendo los valores de X , Y y M de las ecuaciones (1.2.3) y (1.2.4) se obtienen los parámetros originales:

$$A = 4.1 \quad \phi = 153.5^\circ$$

Después de encontrar los parámetros desconocidos M , A y ϕ se obtiene la ecuación del modelo ajustado:

$$y = M + A \cos(\omega t - \phi) = 16.3 + (4.1) \cos(t - 153.5)$$

Donde t está medida en grados (como se dijo al principio), mientras que y , M y A en las unidades de las observaciones y_i . La gráfica del modelo obtenido se encuentra en la Figura 1.2.2.

Finalmente, se puede encontrar la acrofase en términos de los meses del año (para poder interpretar los resultados obtenidos de una forma más real). Para lograr lo anterior, basta observar que 150° son equivalentes a la mitad del mes de junio, por lo que se puede decir que 153° son aproximadamente equivalentes al 16 de junio.

Ejemplo 2

Ehrhardt estudió en 1977 la regulación del transporte de AIB en el hígado de ratas (AIB = ácido 2-aminoisobutírico). Él logró encontrar un ritmo circadiano, el cual se describe en la Figura 1.2.3.

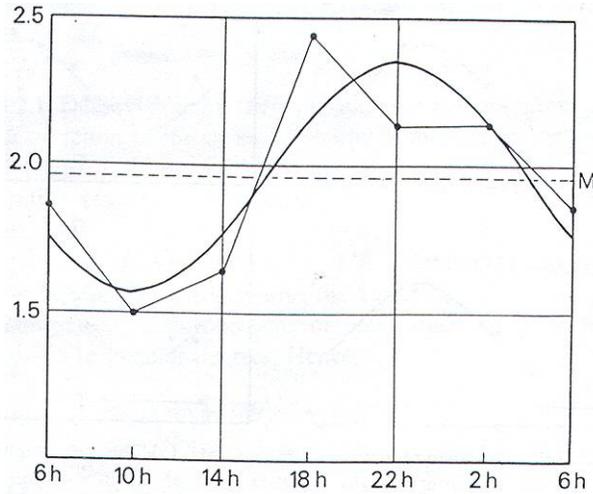


Figura 1.2.3

Se ajustará una función cosenoidal a estos datos. El periodo T y la frecuencia angular son:

$$T = 24\text{hrs.} \quad w = 360^\circ/T = 15^\circ\text{hr}^{-1}$$

Las abscisas t_i son 6, 10, 14, 18, 22, 2 horas, sus valores correspondientes en grados son 90° , 150° , 210° , 270° , 330° , 30° (tomando en cuenta que 24hrs. equivalen a 360°), a estos valores también se les puede ver como sus correspondientes fases angulares ωt_i . De la misma manera que en el ejemplo anterior, con la ecuación (1.2.6) se calculan las cantidades auxiliares c_i y s_i . Con lo anterior y los valores observados y_i se obtiene la Tabla 1.2.2

Time	ωt_i	c_i	s_i	y_i
6 h	90°	0	1.0	1.87
10 h	150°	-0.866	0.5	1.50
14 h	210°	-0.866	-0.5	1.64
18 h	270°	0	-1.0	2.45
22 h	330°	0.866	-0.5	2.14
2 h	30°	0.866	0.5	2.15
Total	—	0	0	11.75

Tabla 1.2.2

Ya que los instantes de tiempo están igualmente espaciados, se obtiene $\sum c_i = 0$ y $\sum s_i = 0$ (tal como resultó en el ejemplo anterior). Además se obtiene lo siguiente:

$$\begin{array}{lll} \sum c_i = 0 & \sum s_i = 0 & \sum s_i c_i = 0 \\ \sum c_i^2 = 3 & \sum s_i^2 = 3 & \sum y_i = 11.75 \\ \sum y_i c_i = 0.996 & \sum y_i s_i = -0.645 & \end{array}$$

Con los resultados anteriores las ecuaciones normales resultan muy sencillas:

$$3X = 0.996$$

$$3Y = -0.645$$

$$6M = 11.75$$

Se puede encontrar el valor de los parámetros X , Y y M de las ecuaciones anteriores con un simple despeje, quedando:

$$X = 0.332 \quad Y = -0.215 \quad M = 1.96$$

Para encontrar los parámetros que se tenían originalmente, es decir, A y ϕ , se utilizan las ecuaciones (1.2.3) y (1.2.4)

$$A = 0.396 \quad \phi = 327.1^\circ$$

Con los parámetros anteriores, se obtiene al fin la ecuación del modelo ajustado:

$$y = M + A \cos(\omega t - \phi) = 1.96 + (0.396) \cos(15t - 327.1)$$

El mesor, M , la amplitud, A , y el valor, y , son medidos en la unidad de los valores observados y_i , mientras que el tiempo, t , está medido en horas. La línea ajustada, los datos observados y el mesor, M , se encuentran ilustrados en la Figura 1.2.3.

La acrofase estimada t_0 , se puede calcular despejando simplemente la ecuación $\phi = \omega t_0$, entonces $t_0 = \phi/\omega = 21.8$ hrs, o bien, 21hrs 48min.

1.3. AJUSTE DE FUNCIONES PERIÓDICAS NO LINEALES

Regularmente, las oscilaciones periódicas presentan un patrón complicado y con más de un máximo y un mínimo en un sólo periodo. En general, ninguna función matemática es apropiada para ajustar tales datos. Sin embargo, los polinómios trigonométricos ofrecen un excelente ajuste, aunque se requieren demasiados términos para éste. Incluso ni siquiera cincuenta términos de un polinomio trigonométrico serían suficientes para ajustar el patrón, relativamebnte sencillo, de un electrocardiograma.

A pesar de esto, existen algunos modelos para el caso particular en que se encuentran únicamente un máximo y un mínimo en un intervalo de un ciclo. La Figura 1.3.1 ilustra un conjunto de oscilaciones que se encuentran más o menos distorsionadas de una oscilación senoidal (la cual se encuentra en el centro de la figura con $v = 0$).

Obsérvese que los máximos y los mínimos no varían según el ciclo en el que estén, es decir, no son más grandes o más pequeños. A este tipo de oscilaciones se les llama *oscilaciones sesgadas*, ya que, corresponden a las distribuciones de probabilidad sesgadas.

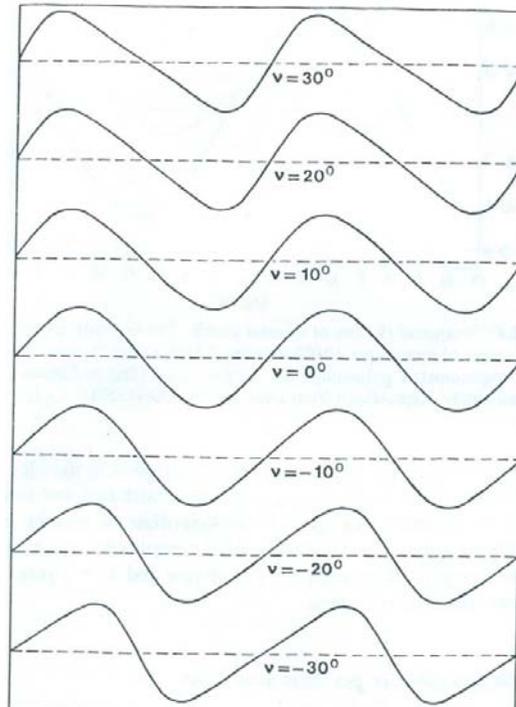


Figura 1.3.1

El parámetro v de sesgo indica qué tanto difiere la gráfica de la oscilación sesgada de la de una oscilación senosoidal.

La siguiente función es apropiada para representar oscilaciones sesgadas ligeramente:

$$y = M + A \cos(\psi + v \cos \psi) \quad (1.3.1)$$

El ángulo ψ denota a la variable:

$$\psi = \omega t - \phi \quad (1.3.2)$$

la cual ya se conoce de la sección 1.2, al igual que el nivel medio M y la amplitud A .

La función (1.3.1), consiste de una función interior

$$\psi + v \cos \psi \quad (1.3.3)$$

y una función exterior $M + A \cos(\cdot)$. Ambas funciones incluyen cosenos.

La función interior contiene un nuevo parámetro, v . El ángulo, ψ es medido en grados o radianes, y v debe estar medido en la misma unidad. Para $v = 0$ se obtiene una simple función coseno, el primer armónico. El parámetro v desvía a la oscilación de una simple oscilación seno. Por lo tanto, v es llamado parámetro de sesgo. los casos $v = 30^\circ, 20^\circ, 10^\circ, \dots, -30^\circ$ se muestran en la Figura 1.3.1.

El parámetro de sesgo está limitado al intervalo:

$$-30^\circ \leq v \leq 30^\circ \quad (1.3.4)$$

o el equivalente en radianes. Para valores de v más allá de este intervalo, la gráfica no es tan suave como es requerida.

Al ángulo particular $\psi = \psi_0$, se le llama ángulo acrofase, para el cual la función (1.3.1) alcanza su máximo, $M + A$. Para determinar ψ_0 , se deriva la ecuación (1.3.1) con respecto a ψ y se iguala lo obtenido a cero:

$$dy/d\psi = -\text{sen}(\psi + v \cos \psi) * (1 - v \text{sen} \psi) = 0$$

Para que la expresión anterior sea efectivamente cero, alguno de los dos factores o los dos deben ser cero. Dado que el ángulo v está limitado por la ecuación (1.3.4), únicamente el primer factor puede ser cero. Por lo tanto, ψ_0 debe satisfacer la ecuación:

$$\psi_0 + v \cos \psi_0 = 0 \quad (1.3.5)$$

Es común que también sea de interés calcular la distancia de un máximo al mínimo consecutivo de una oscilación sesgada. Sea δ esta distancia, expresada en grados, y $\delta\%$ la misma distancia como un porcentaje del periodo (tal como se muestra en la Figura 1.3.2). Entonces:

$$\delta = 180^\circ - 2\psi_0, \quad \delta\% = (\delta/360^\circ) * 100\% \quad (1.3.6)$$

En la Tabla 1.3.1 se muestran diferentes valores de ψ_0 , δ , $\delta\%$ y v :

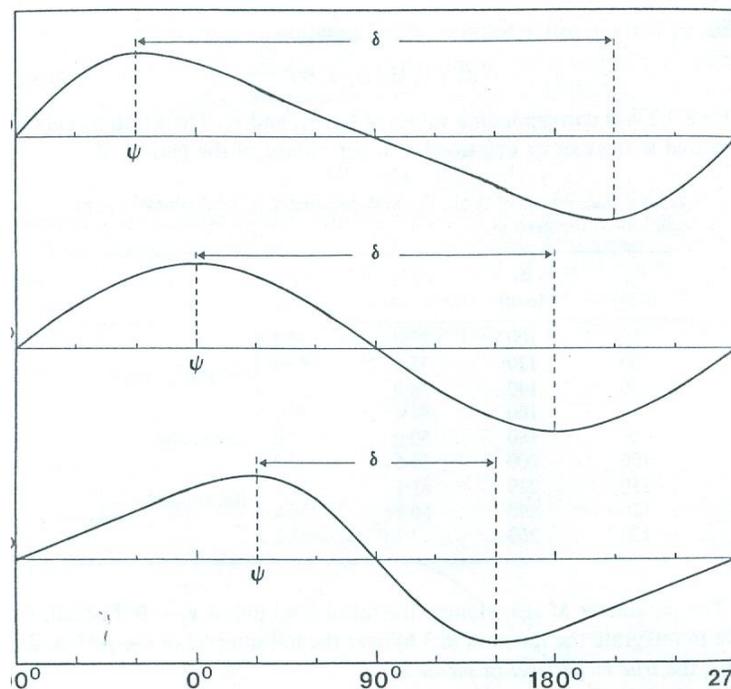


Figura 1.3.2

ψ_o (deg)	δ (deg)	$\delta\%$	v (deg)
-25	230	63.9	27.6
-20	220	61.1	21.3
-15	210	58.3	15.5
-10	200	55.6	10.2
-5	190	52.8	5.0
0	180	50.0	0.0
5	170	47.2	-5.0
10	160	44.4	-10.2
15	150	41.7	-15.5
20	140	38.9	-21.3
25	130	36.1	-27.6

Tabla 1.3.1

Teniendo un conjunto de datos (t_i, y_i) , se puede ajustar un modelo apropiado estimando los parámetros desconocidos de la ecuación (1.3.1). Además de las oscilaciones sesgadas existen otros tipos de oscilaciones para las cuales los modelos se vuelven bastante más complicados.

1.4. CORRELACIÓN CIRCULAR-LINEAL

Existen diversas situaciones en las que una variable circular, ϕ , está relacionada de alguna manera con una variable lineal, y . Frecuentemente, la dirección del viento medida angularmente está correlacionada con variables lineales tales como humedad, cantidad de lluvia, concentración de contaminantes, etcétera. En el área de ritmos biológicos, por ejemplo, la tasa de natalidad (variable lineal) está correlacionada con la hora del día (variable angular), también la acrofase de un ritmo biológico (variable circular) puede depender de la edad del sujeto (variable lineal).

Para calcular un coeficiente de correlación entre este tipo de variables existen métodos paramétricos y no paramétricos. En este texto se presentará únicamente un método paramétrico, el cual resultará más sencillo, ya que, los parámetros que éste necesita son los que ya se presentaron en el modelo de regresión periódica.

Para desarrollar dicho coeficiente, se supondrá que $0^\circ < \phi \leq 360^\circ$ y que y es una variable lineal que de alguna manera depende de ϕ . De esta manera, lo que se tendrán son parejas de datos (ϕ_i, y_i) .

En la relación entre estas variables sólo debe existir un punto máximo y un punto mínimo dentro del intervalo de un periodo. Se puede ver además que una línea de regresión entre estas variables sería aproximadamente senoidal, con ecuación:

$$y = M + A \cos(\phi - \phi_0) \quad (1.4.1)$$

donde M es el nivel medio, A la amplitud, y ϕ_0 el ángulo acrofase.

Nótese que lo que se desea no es estimar los parámetros mencionados, sino probar estadísticamente que y está correlacionada con ϕ , es decir, que y depende de ϕ . Para lo anterior, se introduce una nueva notación:

$$x = \cos(\phi - \phi_0), \quad x_i = \cos(\phi_i - \phi_0) \quad (1.4.2)$$

Con esta notación, la expresión (1.4.1) se puede ver como una ecuación lineal:

$$y = M + Ax \quad (1.4.3)$$

con pendiente A y ordenada al origen M .

Por lo tanto, a las parejas (x_i, y_i) se les puede aplicar el análisis de correlación lineal ordinario. Nótese que los valores y_i de dichas parejas son los que se tienen como datos y que los valores x_i , son los que se obtienen al aplicar la función (1.4.2) a los datos ϕ_i . El único problema que se presenta es que no se conoce el valor de ϕ_0 , por lo que, se debe encontrar un valor razonable para este parámetro.

Se puede obtener un estimador del parámetro desconocido ϕ_0 , siguiendo los pasos descritos en la sección 1.2 para el ajuste de un polinomio trigonométrico sobre la ecuación (1.4.1).

Una vez obtenido el valor de ϕ_0 , y después los valores x_i , se puede calcular el coeficiente de correlación lineal de Pearson o cualquier otro que también busque una relación lineal. Este coeficiente de Pearson está dado por la ecuación:

$$\rho = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2 \sum (y_i - \bar{Y})^2}} \quad (1.4.4)$$

2. OTROS MODELOS DE REGRESIÓN CIRCULAR

2.1. REGRESIÓN LINEAL-CIRCULAR

2.1.1. INTRODUCCIÓN

La teoría de los modelos de regresión cuando la variable de respuesta (es decir, la variable dependiente) es angular se encuentra algo abandonada en el área de estadística, aunque fenómenos de este tipo son comunes en un buen número de áreas de aplicación tales como, Biología, Geología, Paleontología, Física y Meteorología. Un ejemplo común puede ser la dependencia de la dirección del viento sobre la velocidad del mismo.

Dicha teoría se encuentra ampliamente relacionada con el avance en el estudio de las familias más interesantes de distribuciones conjuntas de una variable angular θ y una variable lineal continua X , Gould (1969), Laycock (1975) y Mardia (1972) encontraron algunos modelos de distribuciones conjuntas, sin embargo, modelos más sencillos y por lo tanto más aplicables fueron propuestos por Johnson y Wehrly (1978).

La Cardiología es un campo de la Medicina en el cual las variables angulares están sumamente involucradas, en especial se encuentran muy relacionadas con el vectorcardiograma, un aparato muy útil de diagnóstico que es el análogo en tercera dimensión del electrocardiograma unidimensional común.

En algún instante de tiempo durante una palpitación, la actividad eléctrica del corazón (la cual causa que el corazón se contraiga) puede ser resumida por un vector de campo eléctrico, el cual es medido en la práctica por un sistema de terminales colocadas sobre la superficie del cuerpo; el vectocardiograma es el registro del progreso de este vector a través del espacio. Una manera de analizar el vectorcardiograma es restringir nuestra atención a la orientación del vector, ignorando su longitud. Es por ello que las técnicas apropiadas para el análisis de variables angulares son requeridas en esta aplicación.

2.1.2. ALGUNOS MODELOS

Existen varias formas en las cuales se puede construir un modelo para θ en términos de varias variables explicativas (no se olvide que se ve a θ como variable dependiente y a las demás como independientes). Por ejemplo, se puede desear:

A: Modelar la dirección media de θ en términos de un vector de covariables $x = (x_1, \dots, x_k)'$.

B: Modelar la dispersión de θ en términos de x_1, \dots, x_k .

C: Modelar ambos, la dirección media y la dispersión simultáneamente, en términos de x_1, \dots, x_k .

La razón por la cual en la clasificación anterior de los modelos, en los de tipo A se menciona un vector de covariables y en B y C no, es porque en las expresiones que se tienen para modelar la dirección media es conveniente que dichas covariables aparezcan expresadas como un vector, mientras que en los modelos para modelar la dispersión y los que modelan a ambos es mejor que aparezcan de manera individual cada una de las covariables.

Una peculiaridad de los modelos de regresión con una variable de respuesta circular es que se tiene disponible un modelo especial nulo, por ejemplo en la distribución uniforme, para la cual no hay dirección media. Esta es una razón para considerar a los modelos de regresión basados sobre la dispersión tanto como a los basados en la dirección media; otra característica es el hecho de que muchos conjuntos encontrados en la práctica (por ejemplo, mediciones de la dirección del viento y la velocidad de éste) presentan rasgos tales como variabilidad incrementada de θ para valores pequeños de X .

Además de la razón ya mencionada para tomar en cuenta a los modelos basados en la dispersión tanto como a los basados en la dirección media, otra razón es que el ángulo medio de una muestra tomada de una distribución unimodal indica sólo una dirección predilecta, pero no contiene ninguna información acerca de cómo es que se extienden los valores de la muestra que se encuentran agrupados alrededor de la media. Así como en estadística lineal se tiene que: *Una medida de dispersión es tan importante como una medida de ubicación*, este hecho también se cumple en estadística circular. Algunas de las medidas de dispersión para variables circulares son: varianza angular, desviación angular y rango.

Modelos de dispersión para una desviación angular presentan ciertas dificultades, debido a que no existe una medida natural de escala para distribuciones circulares. Por esta razón, es conveniente trabajar con la familia de distribuciones von Mises, la cual tiene construida una medida de dispersión, y además comparte muchas de las propiedades principales, en inferencia estadística, con la distribución normal que se tiene para datos lineales. La función de densidad de una distribución von Mises $VM(\mu, k)$ está dada por:

$$f(\theta; \mu, k) = \frac{1}{2\pi I_0(k)} e^{k \cos(\theta - \mu)}, \quad 0 \leq \theta < 2\pi \quad (2.1.1)$$

donde $0 \leq \mu < 2\pi$ es la dirección media y $k \geq 0$ el parámetro de concentración. Aquí $I_0(k)$ es la Función Bessel³ de primer tipo y orden cero, y está dada por:

$$I_0(k) = \frac{1}{2\pi} \int_0^{2\pi} e^{k \cos(\theta)} d\theta = \sum_{r=0}^{\infty} \left(\frac{k}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2 \quad (2.1.2)$$

En la función de densidad (2.1.1) el valor $k = 0$ corresponde a la distribución circular uniforme; para $k > 0$, la densidad es unimodal y simétrica con respecto a μ , además el nivel de concentración se incrementa conforme lo hace k . Para $k \geq 2$ la densidad en la antípoda $\mu + \pi$ es efectivamente cero, y f está bien aproximada por una distribución normal con varianza

³Ver Anexo, Función Bessel

$1/k$. En lo siguiente, se supone que θ es una variable von Mises.

Se presenta un modelo de regresión de tipo A propuesto por Gould (1969):

$$\mu = \mu_0 + \sum \beta_j x_j \quad (2.1.3)$$

para la dirección media, resultando en una forma "barber's pole"⁴.

Gould dio un método iterativo para calcular los estimadores máximo verosímiles de los parámetros del modelo, y construyó algunos métodos aproximados de inferencia para el caso cuando k puede ser considerado muy grande. Laycock (1975) también discutió este modelo, y observó que los estimadores máximo verosímiles son equivalentes a los de mínimos cuadrados. Johnson y Wehrly (1978) encontraron que la función de verosimilitud en el modelo de Gould tiene una infinidad de picos de la misma altura, dejando ambiguamente definidos a los estimadores máximo verosímiles. Como una alternativa, para una única variable explicativa, ellos sugirieron una aproximación vía un modelo específico para la distribución conjunta de θ y una variable lineal X con una función de distribución marginal completamente especificada $F(x)$. La distribución condicional de θ dada $X = x$ propuesta por ellos es $VM(\mu + 2\pi F(x), k)$, un modelo que permite la estimación directa de μ y k . Se conoce a éste como el modelo tipo A de Johnson y Wehrly.

Para modelos tipo B con una única variable explicativa, sugirieron modelar la distribución condicional de θ dada $X = x$ como $VM(\mu, kx)$, el cual permite también la estimación directa de los parámetros, y también incluye el caso nulo de concentración cero.

⁴Ver Glosario

2.2. REGRESIÓN CIRCULAR-CIRCULAR

2.2.1. INTRODUCCIÓN

En los modelos de regresión circular-circular se busca predecir una dirección dada otra, con la cual la primera está relacionada. Por ejemplo, en un sismo, se puede ajustar un modelo para predecir la dirección del descenso más empinado con la dirección del movimiento de la tierra.

Otras aplicaciones se encuentran en el área de la Biología que estudia aspectos sobre la migración de especies, por ejemplo, la dirección en la cual viene una parvada puede estar relacionada con la dirección en la que regresa, o la dirección del vuelo de ésta con la dirección del viento.

La mayoría de las veces, la literatura acerca de métodos estadísticos para muestras bivariadas de ángulos (Fisher, 1993; Jupp & Mardia, 1989), ha sido relacionada con la inferencia de coeficientes de correlación que sirven para medir la dependencia entre muestras angulares.

Algunos de los modelos que existen para este tipo de regresión expresan a la dirección y como una rotación de la dirección x (tomando como variable dependiente a y y como independiente a x). Otros modelos, predicen el ángulo y usando una rotación del ángulo "descentrado" x , donde el término ángulo "descentrado" se puede definir como: la dirección de la suma del vector unitario correspondiente al ángulo x más un vector no centrado, es decir, $(u_1 + \cos x, u_2 + \sin x)$. En dichos modelos, los parámetros a estimar son u_1 y u_2 .

2.2.2. ALGUNOS MODELOS

Se consideran dos variables aleatorias α y β , ambas circulares, con una función de densidad conjunta $f(\alpha, \beta)$. Donde $0 < \alpha, \beta \leq 2\pi$. Para predecir el valor de β para un valor dado de α , considérese la regresión o esperanza condicional del vector $e^{i\beta}$ dado α , es decir,

$$E[e^{i\beta}|\alpha] = \rho(\alpha)e^{i\mu(\alpha)} = g_1(\alpha) + ig_2(\alpha) \quad (2.2.1)$$

donde $\mu(\alpha)$ representa la dirección media condicional de β dado α y $0 \leq \rho(\alpha) \leq 1$ la concentración condicional hacia esta dirección. Y donde:

$$\begin{aligned} g_1(\alpha) &= E[\cos \beta|\alpha], \\ g_2(\alpha) &= E[\sin \beta|\alpha] \end{aligned} \quad (2.2.2)$$

Es importante aclarar que en la ecuación (2.2.1) se está considerando $E(e^{i\beta}|\alpha)$ porque el número imaginario (o complejo) $e^{i\beta}$ servirá para predecir el valor de β dado un valor de α , pero las igualdades de dicha expresión resultan de las diversas expresiones que se tienen para los números complejos. Para comprender dicha ecuación se recomienda al lector revisar en el Anexo la parte de números complejos y después hacer el siguiente desarrollo:

$$\begin{aligned}
E[e^{i\beta}|\alpha] &= E[\cos \beta + i \operatorname{sen} \beta|\alpha] \\
&= E[\cos \beta|\alpha] + E[i \operatorname{sen} \beta|\alpha] \\
&= E[\cos \beta|\alpha] + iE[\operatorname{sen} \beta|\alpha] \\
&= g_1(\alpha) + ig_2(\alpha)
\end{aligned}$$

De esta manera queda claro por qué $E[e^{i\beta}|\alpha] = g_1(\alpha) + ig_2(\alpha)$ en la ecuación (2.2.1), lo único que falta aclarar es la primera igualdad de esta ecuación, y sólo se trata de una forma general de expresar $E[e^{i\beta}|\alpha]$.

Por todo lo anterior β queda determinada como:

$$\mu(\alpha) = \beta = \arctan \frac{g_2(\alpha)}{g_1(\alpha)} \quad (2.2.3)$$

La predicción de β de esta manera es óptima en el sentido de que ésta minimiza la expresión:

$$E\{|e^{i\beta} - g(\alpha)|^2\}$$

lo que resulta muy similar a la idea de los mínimos cuadrados.

Al no tener una expresión específica para $g_1(\alpha)$ y $g_2(\alpha)$ resulta en general difícil estimarlas a partir de los datos. Así que $g_1(\alpha)$ y $g_2(\alpha)$ serán aproximadas por funciones adecuadas. Debido a que $g_1(\alpha)$ y $g_2(\alpha)$ son funciones periódicas, con periodo 2π , éstas pueden ser expresadas en términos de sus expansiones en series de Fourier. Esto conduce a aproximar a $g_1(\alpha)$ y $g_2(\alpha)$ por polinomios trigonométricos de un grado adecuado m , es decir,

$$\begin{aligned}
g_1(\alpha) &\approx \sum_{k=0}^m (A_k \cos k\alpha + B_k \operatorname{sen} k\alpha), \\
g_2(\alpha) &\approx \sum_{k=0}^m (C_k \cos k\alpha + D_k \operatorname{sen} k\alpha).
\end{aligned} \quad (2.2.4)$$

Entonces se obtiene el modelo siguiente, el cual es de hecho un modelo lineal general:

$$\begin{aligned}
\cos \beta &\approx \sum_{k=0}^m (A_k \cos k\alpha + B_k \operatorname{sen} k\alpha) + \epsilon_1, \\
\operatorname{sen} \beta &\approx \sum_{k=0}^m (C_k \cos k\alpha + D_k \operatorname{sen} k\alpha) + \epsilon_2.
\end{aligned} \quad (2.2.5)$$

donde $(\epsilon_1, \epsilon_2)^t$ es el vector error con vector medio 0 y matriz de dispersión Σ desconocida.

3. UNA APLICACIÓN DEL MODELO DE REGRESIÓN PERIÓDICA

El viento es un fenómeno meteorológico definido como “aire en movimiento”, el cual se caracteriza por dos elementos: velocidad y dirección.

Generalmente, dichas características son representadas por medio de vectores en forma polar, donde la magnitud está dada por la velocidad del viento y la orientación por la dirección de éste.

Los meteorólogos crearon una gráfica llamada Rosa de los vientos que permite representar simultáneamente la relación que existe entre las características del viento.

Los vientos globales se generan como consecuencia del desplazamiento del aire desde zonas de alta presión a zonas de baja presión, determinando los vientos dominantes de un área o región. Aún así hay que tener en cuenta numerosos factores locales que influyen o determinan la intensidad y periodicidad de los movimientos del aire. Estos factores, difíciles de simplificar por su multiplicidad, son los que permiten hablar de vientos locales, los cuales son en muchos lugares más importantes que los de carácter general.

Para esta aplicación se toma en cuenta sólo una de las características del viento, su velocidad, con la finalidad de determinar si tiene un comportamiento periódico.

Los datos fueron obtenidos a 48m. sobre el nivel del mar en la Central Eólica ubicada en Guerrero Negro, Baja California Sur. Son datos promedio de la velocidad del viento registrados en esta central en el año 2007 y se encuentran medidos en m/s.

Como ya se mencionó, en cada región existen factores locales que influyen en el movimiento del viento. Es por esto que los modelos obtenidos podrían ser utilizados sólo para describir el comportamiento de la velocidad del viento en el lugar de origen de los datos.

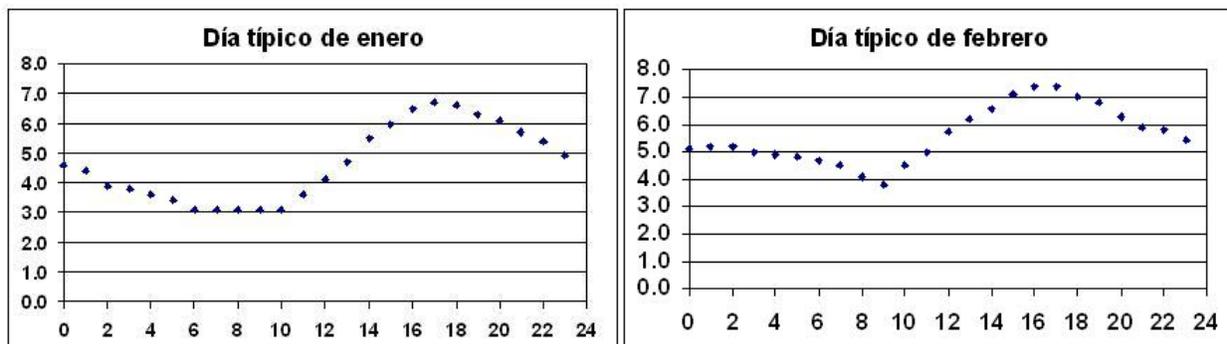
En la siguiente tabla se muestran los datos obtenidos:

VELOCIDAD PROMEDIO DEL VIENTO A 48m EN LA CENTRAL EÓLICA, GUERRERO NEGRO, BCS.																									
Mes/Hora	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Promedio
Enero	4.6	4.4	3.9	3.8	3.6	3.4	3.1	3.1	3.1	3.1	3.1	3.6	4.1	4.7	5.5	6.0	6.5	6.7	6.6	6.3	6.1	5.7	5.4	4.9	4.6
Febrero	5.1	5.2	5.2	5.0	4.9	4.8	4.7	4.5	4.1	3.8	4.5	5.0	5.7	6.2	6.6	7.1	7.4	7.4	7.0	6.8	6.3	5.9	5.8	5.4	5.6
Marzo	6.0	5.8	5.4	5.0	4.6	4.3	3.9	3.9	4.3	4.3	4.6	5.1	5.8	6.1	6.8	7.4	7.6	7.8	7.7	7.6	7.7	7.4	6.8	6.3	5.9
Abril	6.6	6.3	6.2	5.9	5.5	5.6	5.4	5.5	5.5	5.6	5.9	6.4	7.3	8.0	8.7	9.3	9.5	9.6	9.2	8.5	8.3	7.9	7.6	7.0	7.1
Mayo	6.4	6.1	5.8	5.6	5.4	5.2	5.0	4.8	4.6	4.9	5.4	6.1	7.0	7.6	8.2	8.7	9.0	9.3	9.2	8.8	8.3	7.9	7.3	6.8	6.8
Junio	6.0	5.6	5.4	5.2	5.1	5.0	4.8	4.8	4.8	4.8	5.1	5.8	6.5	7.3	8.0	8.5	9.0	9.1	9.0	8.7	8.3	7.6	7.0	6.5	6.6
Julio	5.8	5.7	5.6	5.4	5.3	5.2	5.0	5.1	5.1	5.2	5.6	6.2	6.8	7.5	8.0	8.5	9.0	9.1	8.9	8.6	7.9	7.3	6.7	6.2	6.7
Agosto	5.1	4.8	4.9	5.0	5.0	4.9	4.7	4.5	4.3	4.4	4.9	5.7	6.2	6.9	7.7	8.0	8.2	8.4	8.3	7.9	7.4	6.7	5.9	5.5	6.1
Septiembre	5.4	5.4	5.3	5.5	5.3	5.1	4.9	4.9	5.0	5.1	5.4	5.9	6.5	7.2	7.6	8.0	8.3	8.5	8.4	7.9	7.3	6.6	5.8	5.6	6.3
Octubre	3.5	3.4	3.6	3.5	3.6	3.3	3.1	3.0	2.8	2.8	3.1	3.9	4.9	5.7	6.1	6.5	6.7	7.0	6.9	6.3	5.7	5.4	4.9	4.0	4.6
Noviembre	4.5	4.0	3.6	3.3	3.1	2.8	2.8	2.7	2.3	2.3	2.9	3.4	4.3	4.9	5.5	6.0	6.2	6.3	6.2	6.3	6.0	5.7	5.2	4.9	4.4
Diciembre	3.9	3.6	3.2	2.9	3.0	2.9	2.8	2.8	2.8	2.5	2.1	2.8	3.6	4.1	4.7	5.6	6.1	6.2	6.0	5.6	5.2	4.6	4.4	4.3	4.0
Promedio	4.6	4.4	3.9	3.8	3.6	3.4	3.1	3.1	3.1	3.1	3.1	3.6	4.1	4.7	5.5	6.0	6.5	6.7	6.6	6.3	6.1	5.7	5.4	4.9	5.7

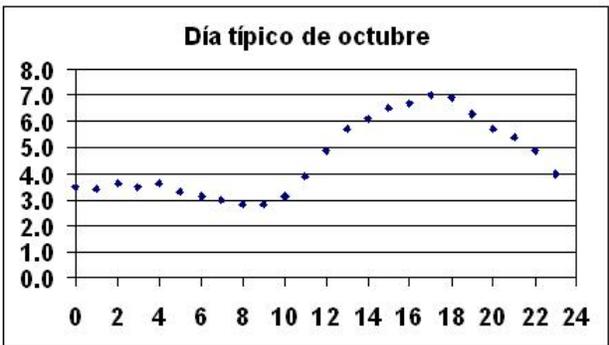
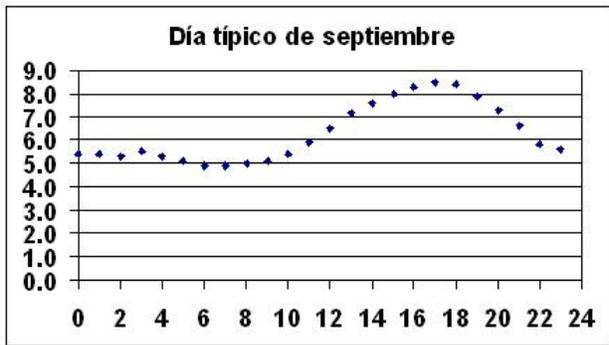
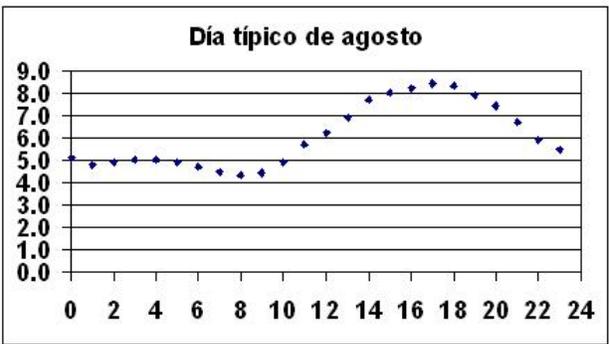
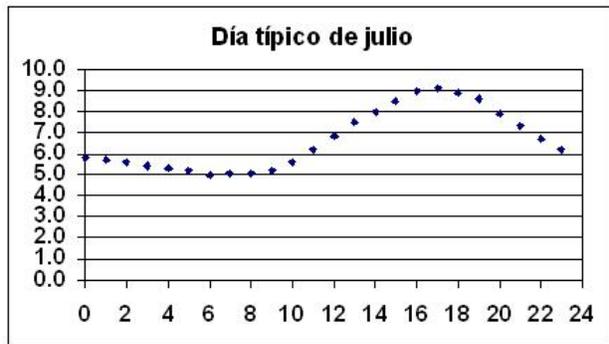
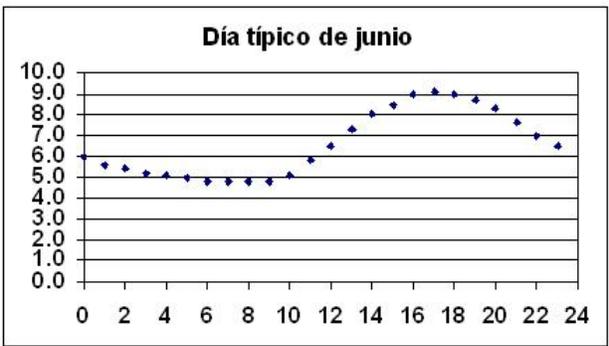
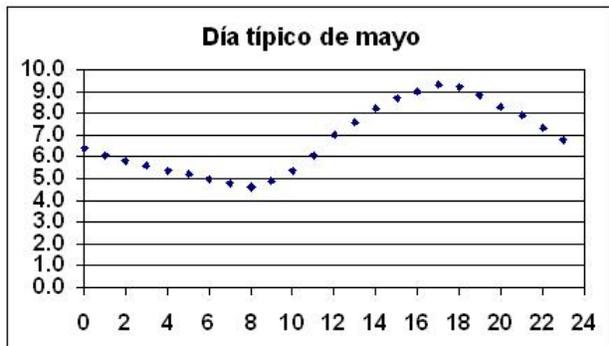
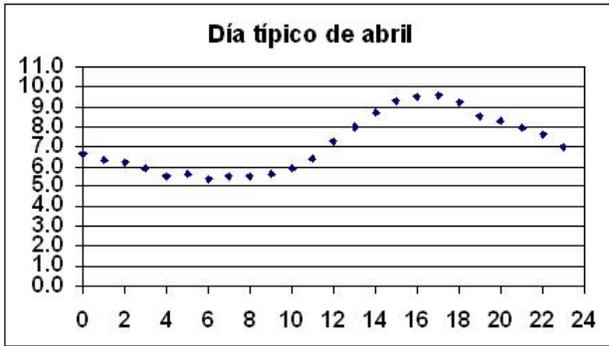
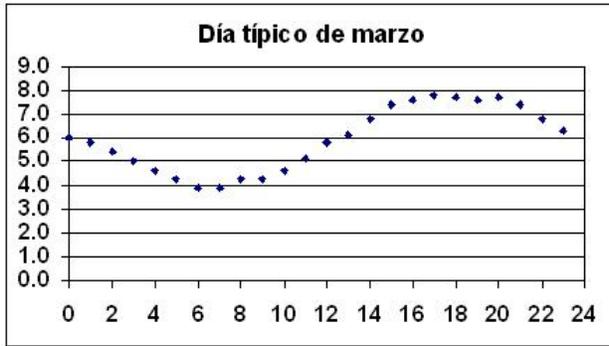
Tabla 3.1

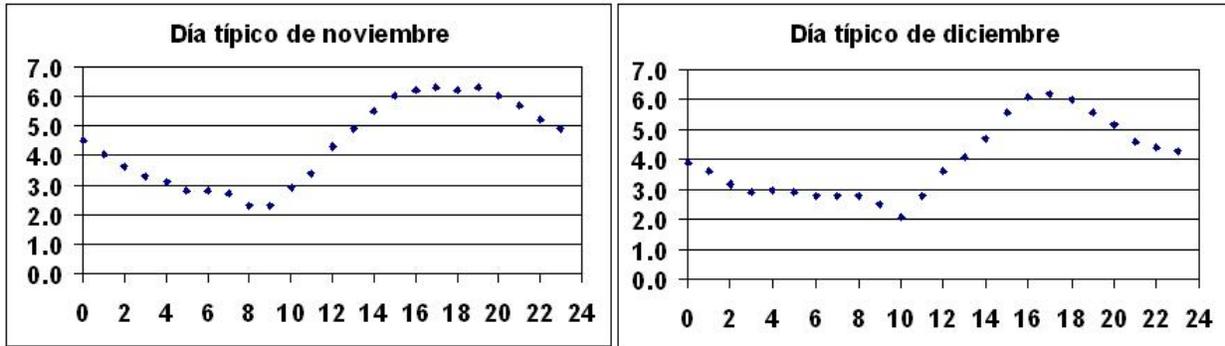
En la tabla anterior, los valores que se presentan en cada mes del año son las velocidades promedio de todos los días de ese mes a una hora determinada. Por ejemplo, el valor que se encuentra en el mes de enero a la hora 0 es un promedio de la velocidad del viento de todos los días de enero a esta hora, y así sucesivamente. De tal manera que los renglones de la tabla se pueden interpretar como un día típico⁵ de cada uno de los meses del año. Así mismo, el último renglón representa un día típico del año ya que sus valores son el promedio de los de todos los meses del año.

Para observar el comportamiento de la velocidad del viento a través del tiempo, se trazan las gráficas de los días típicos de cada mes, tomando como variable independiente la hora del día y como dependiente la velocidad; de esta manera se obtiene lo siguiente:



⁵En este trabajo se maneja un día típico como un día promedio





En las gráficas anteriores se observa que cada uno de los días siguen un mismo patrón, lo que supone que la velocidad del viento presenta un fenómeno periódico de 24hrs. A pesar de esto, los valores de cada gráfica son muy distintos. Por ejemplo, en un día típico de enero las velocidades oscilan entre 3.1 y 6.7 m/s mientras que en el del mes de abril oscilan entre 5.4 y 9.6 m/s. Esto resulta lógico ya que la velocidad del viento depende, entre otras cosas, de la estación del año.

Se pudo haber ajustado un sólo modelo de frecuencia diaria con los datos del último renglón de la Tabla 5.1 y se habría obtenido un modelo para predecir la velocidad del viento a una hora determinada de un día típico del año. Pero este modelo tendría poca precisión, ya que, la velocidad del viento depende en gran medida del mes en curso. Es por esto, que se ajustará un modelo con un periodo de 24hrs. para cada uno de los meses del año.

Se tienen los mismos valores $t_1 = 0, t_2 = 1, t_3 = 2, \dots, t_{24} = 23$ de la variable independiente t medida en horas para cada uno de dichos modelos. También se tiene el mismo periodo $T = 24$ hrs. por lo que la frecuencia angular:

$$w = 360^\circ/T = 360^\circ/24hrs. = 15^\circ/hr$$

es también la misma para todos los modelos.

Para obtener los valores c_i y s_i de la ecuación (1.2.6), a cada instante de tiempo t_i se le asocia un ángulo, es decir, su correspondiente fase angular wt_i resultando: $wt_1 = 0^\circ, wt_2 = 15^\circ, wt_3 = 30^\circ, \dots, wt_{24} = 345^\circ$. Los valores wt_i también son los mismos para cada uno de los modelos.

A continuación se presentan los modelos que resultan para los días típicos de cada mes:

Enero

En la siguiente tabla se encuentran los instantes de tiempo t_i , sus respectivas c_i y s_i , los datos observados y_i , y otras cantidades como c_i^2 y s_i^2 que son necesarias para estimar los parámetros del modelo.

Hora (t _i)	Enero (y _i)	wt _i	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	4.60	0	1.0000	0.0000	1.0000	0.0000	0.0000	4.6000	0.0000
1	4.40	15	0.9659	0.2588	0.9330	0.0670	0.2500	4.2501	1.1388
2	3.90	30	0.8660	0.5000	0.7500	0.2500	0.4330	3.3775	1.9500
3	3.80	45	0.7071	0.7071	0.5000	0.5000	0.5000	2.6870	2.6870
4	3.60	60	0.5000	0.8660	0.2500	0.7500	0.4330	1.8000	3.1177
5	3.40	75	0.2588	0.9659	0.0670	0.9330	0.2500	0.8800	3.2841
6	3.10	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	3.1000
7	3.10	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-0.8023	2.9944
8	3.10	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-1.5500	2.6847
9	3.10	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-2.1920	2.1920
10	3.10	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-2.6847	1.5500
11	3.60	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-3.4773	0.9317
12	4.10	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-4.1000	0.0000
13	4.70	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-4.5399	-1.2164
14	5.50	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-4.7631	-2.7500
15	6.00	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-4.2426	-4.2426
16	6.50	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-3.2500	-5.6292
17	6.70	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-1.7341	-6.4717
18	6.60	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.6000
19	6.30	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	1.6306	-6.0853
20	6.10	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	3.0500	-5.2828
21	5.70	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	4.0305	-4.0305
22	5.40	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	4.6765	-2.7000
23	4.90	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	4.7330	-1.2682
Suma	111.30		0.0000	0.0000	12.0000	12.0000	0.0000	2.3791	-20.6463

Tabla 3.2

Con los valores de la Tabla 3.2, se obtienen las constantes que fueron definidas en la sección 1.2

$$\begin{aligned}
 a &= \sum c_i = 0 & b &= \sum c_i^2 = 12 & c &= \sum s_i c_i = 0 \\
 d &= \sum y_i c_i = 2.3791 & e &= \sum s_i = 0 & f &= \sum s_i^2 = 12 \\
 g &= \sum y_i s_i = -20.6463 & k &= n = 24 & p &= \sum y_i = 111.30
 \end{aligned}$$

De esta manera, se pueden sustituir las constantes anteriores en la solución (1.2.12) para así obtener los estimadores: M , X y Y :

$$M = 4.6375 \quad X = 0.1982 \quad Y = -1.7205$$

Con los estimadores anteriores se tiene ya el modelo de la ecuación (1.2.1). Sin embargo, siempre que sea posible se debe llegar al modelo (1.1.3). Para ello, se utilizan las ecuaciones (1.2.3) y (1.2.4):

$$\begin{aligned}
 A &= (X^2 + Y^2)^{1/2} = ((0.1982)^2 + (-1.7205)^2)^{1/2} = 1.7319 \\
 \phi &= \arctan(Y/X) = \arctan(-1.7205/0.1982) = -83.4267
 \end{aligned}$$

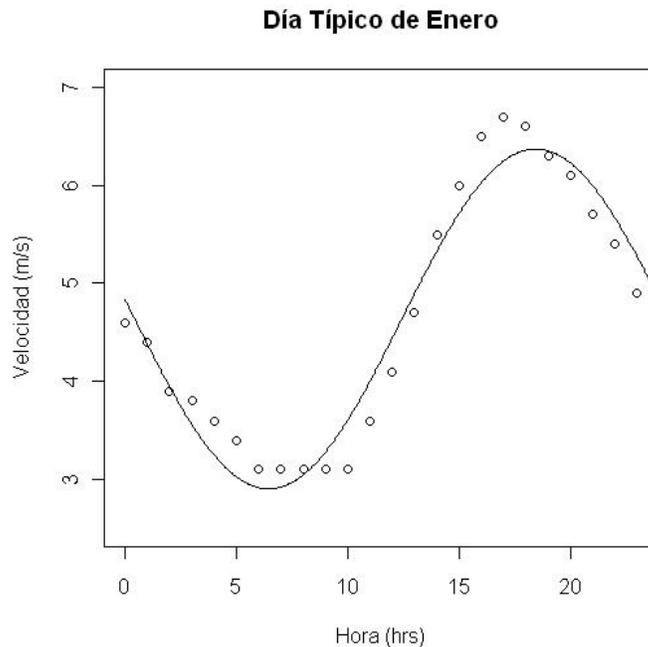
dado que $X > 0$, según la ecuación (1.2.4), el valor de $\phi = -83.4267$ es correcto. Sin embargo, es importante (para fines de interpretación) que el ángulo ϕ esté medido en dirección contraria a las manecillas del reloj a partir del eje x. Por lo tanto, al valor que ya se tiene se le suman 360° . De esta manera, se obtienen los estimadores:

$$M = 4.6375 \quad A = 1.7319 \quad \phi = 276.5732$$

Por lo tanto, el modelo (1.1.3) para las velocidades del viento de un día típico del mes de enero es:

$$g(t) = 4.63 + 1.73 \cos(15t - 276.57)$$

Para ver que tan preciso es el modelo ajustado, se representan en una misma gráfica tanto el modelo como los valores observados:



En la gráfica anterior se puede observar que el ajuste es bastante bueno, lo que también indica que no hay errores en los cálculos efectuados. Sin embargo, se deben hacer pruebas de bondad de ajuste siempre que se encuentre un modelo de regresión. En los textos consultados para elaborar este trabajo sólo se hace referencia a la existencia de dichas pruebas para este tipo de regresión, pero no se describe ninguna de ellas. Por esto, se utilizará el coeficiente de correlación circular-lineal ρ , de la Sección 1.4 como un argumento estadístico para probar la correlación que existe entre la hora del día y la velocidad del viento.

Para obtener dicho coeficiente se construyen las siguientes variables:

$$x_i = \cos(\omega t_i - \phi)$$

donde los valores t_i son los datos que se tienen en la Tabla 3.2.

Con las variables construidas y los datos y_i de la misma tabla, se aplica la fórmula de la expresión (1.4.4), obteniendo el coeficiente de correlación circular-lineal:

$$\rho = 0.9738$$

el cual resulta bastante alto, por lo que podemos concluir que las dos variables: hora del día y velocidad del viento están fuertemente correlacionadas en el mes de enero.

Siguiendo el mismo procedimiento, se ajustan los modelos para los días típicos de cada uno de los meses restantes. Por lo tanto, para cada mes, sólo se presenta la tabla de datos, el modelo ajustado y la gráfica correspondiente.

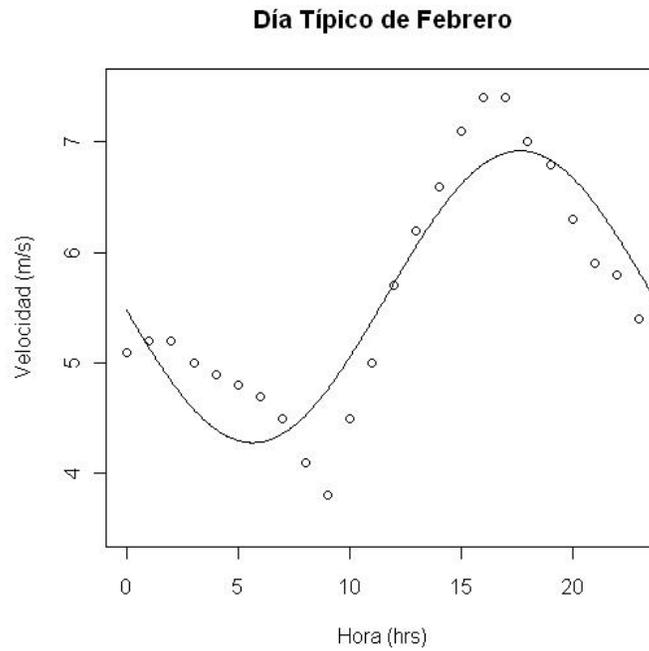
Febrero

Hora	Febrero								
(t_i)	(y_i)	wt	c_i	s_i	c_i²	s_i²	s_i c_i	y_i c_i	y_i s_i
0	5.10	0	1.0000	0.0000	1.0000	0.0000	0.0000	5.1000	0.0000
1	5.20	15	0.9659	0.2588	0.9330	0.0670	0.2500	5.0228	1.3459
2	5.20	30	0.8660	0.5000	0.7500	0.2500	0.4330	4.5033	2.6000
3	5.00	45	0.7071	0.7071	0.5000	0.5000	0.5000	3.5355	3.5355
4	4.90	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.4500	4.2435
5	4.80	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.2423	4.6364
6	4.70	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	4.7000
7	4.50	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.1647	4.3467
8	4.10	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.0500	3.5507
9	3.80	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-2.6870	2.6870
10	4.50	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-3.8971	2.2500
11	5.00	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-4.8296	1.2941
12	5.70	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-5.7000	0.0000
13	6.20	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-5.9887	-1.6047
14	6.60	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-5.7158	-3.3000
15	7.10	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-5.0205	-5.0205
16	7.40	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-3.7000	-6.4086
17	7.40	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-1.9153	-7.1479
18	7.00	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-7.0000
19	6.80	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	1.7600	-6.5683
20	6.30	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	3.1500	-5.4560
21	5.90	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	4.1719	-4.1719
22	5.80	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	5.0229	-2.9000
23	5.40	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	5.2160	-1.3976
Suma	134.40		0.0000	0.0000	12.0000	12.0000	0.0000	-1.4938	-15.7856

Tabla 3.3

$$g(t) = 5.60 + 1.32 \cos(15t - 264.59)$$

$$\rho = 0.9100$$



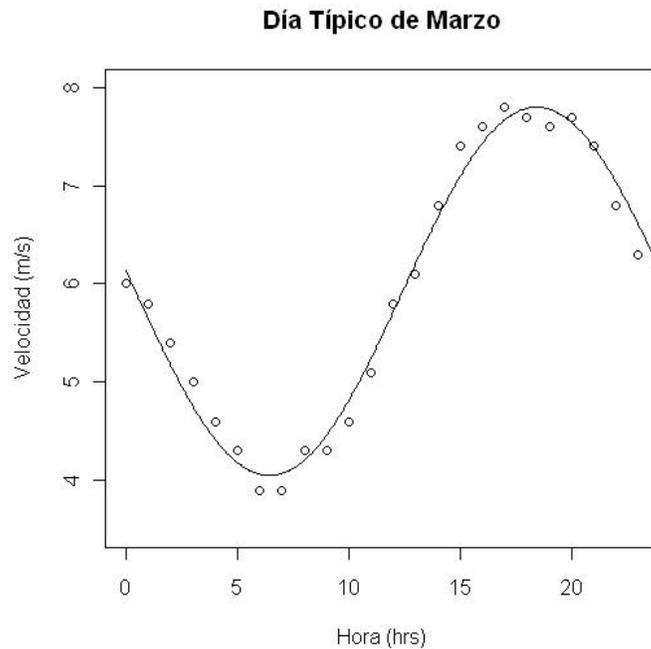
Marzo

Hora	Marzo								
(t)	(y _i)	wt	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	6.00	0	1.0000	0.0000	1.0000	0.0000	0.0000	6.0000	0.0000
1	5.80	15	0.9659	0.2588	0.9330	0.0670	0.2500	5.6024	1.5012
2	5.40	30	0.8660	0.5000	0.7500	0.2500	0.4330	4.6765	2.7000
3	5.00	45	0.7071	0.7071	0.5000	0.5000	0.5000	3.5355	3.5355
4	4.60	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.3000	3.9837
5	4.30	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.1129	4.1535
6	3.90	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	3.9000
7	3.90	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.0094	3.7671
8	4.30	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.1500	3.7239
9	4.30	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-3.0406	3.0406
10	4.60	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-3.9837	2.3000
11	5.10	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-4.9262	1.3200
12	5.80	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-5.8000	0.0000
13	6.10	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-5.8921	-1.5788
14	6.80	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-5.8890	-3.4000
15	7.40	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-5.2326	-5.2326
16	7.60	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-3.8000	-6.5818
17	7.80	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-2.0188	-7.5342
18	7.70	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-7.7000
19	7.60	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	1.9670	-7.3410
20	7.70	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	3.8500	-6.6684
21	7.40	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	5.2326	-5.2326
22	6.80	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	5.8890	-3.4000
23	6.30	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	6.0853	-1.6306
Suma	142.20		0.0000	0.0000	12.0000	12.0000	0.0000	2.5089	-22.3745

Tabla 3.4

$$g(t) = 5.92 + 1.87 \cos(15t - 276.39)$$

$$\rho = 0.9915$$



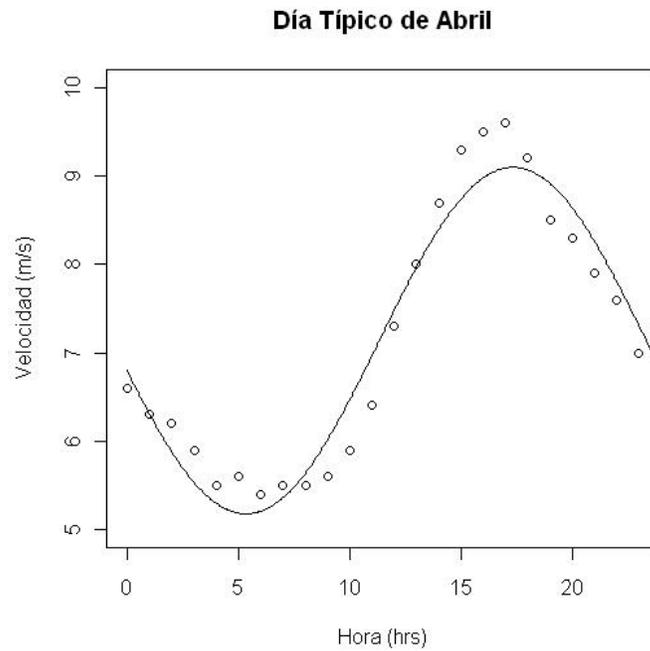
Abril

Hora	Abril								
(t _i)	(y _i)	wt _i	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	6.60	0	1.0000	0.0000	1.0000	0.0000	0.0000	6.6000	0.0000
1	6.30	15	0.9659	0.2588	0.9330	0.0670	0.2500	6.0853	1.6306
2	6.20	30	0.8660	0.5000	0.7500	0.2500	0.4330	5.3694	3.1000
3	5.90	45	0.7071	0.7071	0.5000	0.5000	0.5000	4.1719	4.1719
4	5.50	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.7500	4.7631
5	5.60	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.4494	5.4092
6	5.40	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	5.4000
7	5.50	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.4235	5.3126
8	5.50	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.7500	4.7631
9	5.60	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-3.9598	3.9598
10	5.90	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-5.1095	2.9500
11	6.40	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-6.1819	1.6564
12	7.30	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-7.3000	0.0000
13	8.00	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-7.7274	-2.0706
14	8.70	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-7.5344	-4.3500
15	9.30	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-6.5761	-6.5761
16	9.50	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-4.7500	-8.2272
17	9.60	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-2.4847	-9.2729
18	9.20	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-9.2000
19	8.50	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	2.2000	-8.2104
20	8.30	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	4.1500	-7.1880
21	7.90	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	5.5861	-5.5861
22	7.60	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	6.5818	-3.8000
23	7.00	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	6.7615	-1.8117
Suma	171.30		0.0000	0.0000	12.0000	12.0000	0.0000	-4.0920	-23.1762

Tabla 3.5

$$g(t) = 7.13 + 1.96 \cos(15t - 259.98)$$

$$\rho = 0.9698$$



Mayo

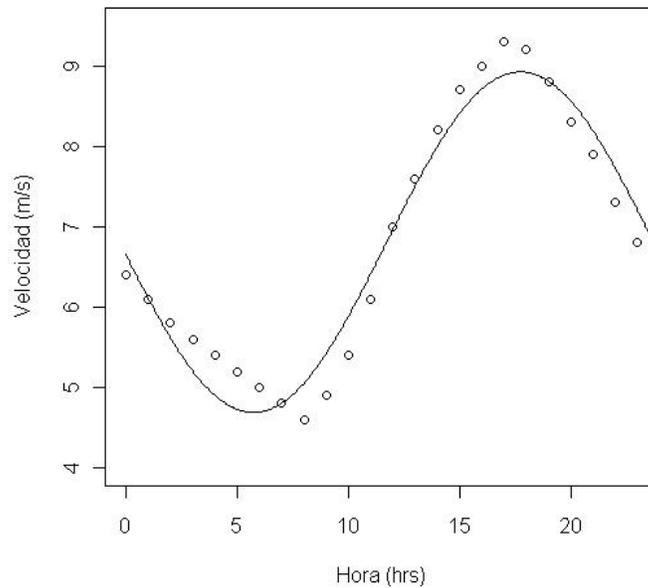
Hora (t)	Mayo (y _i)	wt _i	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	6.40	0	1.0000	0.0000	1.0000	0.0000	0.0000	6.4000	0.0000
1	6.10	15	0.9659	0.2588	0.9330	0.0670	0.2500	5.8921	1.5788
2	5.80	30	0.8660	0.5000	0.7500	0.2500	0.4330	5.0229	2.9000
3	5.60	45	0.7071	0.7071	0.5000	0.5000	0.5000	3.9598	3.9598
4	5.40	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.7000	4.6765
5	5.20	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.3459	5.0228
6	5.00	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	5.0000
7	4.80	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.2423	4.6364
8	4.60	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.3000	3.9837
9	4.90	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-3.4648	3.4648
10	5.40	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-4.6765	2.7000
11	6.10	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-5.8921	1.5788
12	7.00	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-7.0000	0.0000
13	7.60	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-7.3410	-1.9670
14	8.20	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-7.1014	-4.1000
15	8.70	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-6.1518	-6.1518
16	9.00	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-4.5000	-7.7942
17	9.30	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-2.4070	-8.9831
18	9.20	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-9.2000
19	8.80	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	2.2776	-8.5001
20	8.30	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	4.1500	-7.1880
21	7.90	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	5.5861	-5.5861
22	7.30	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	6.3220	-3.6500
23	6.80	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	6.5683	-1.7600
Suma	163.40		0.0000	0.0000	12.0000	12.0000	0.0000	-1.8523	-25.3787

Tabla 3.6

$$g(t) = 6.80 + 2.12 \cos(15t - 265.82)$$

$$\rho = 0.9767$$

Día Típico de Mayo



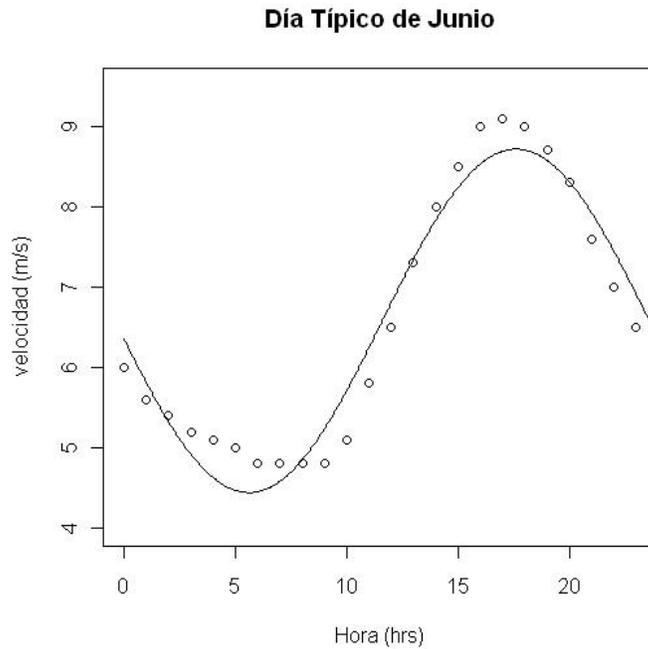
Junio

Hora (t)	Junio (y _i)	wt _i	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	6.00	0	1.0000	0.0000	1.0000	0.0000	0.0000	6.0000	0.0000
1	5.60	15	0.9659	0.2588	0.9330	0.0670	0.2500	5.4092	1.4494
2	5.40	30	0.8660	0.5000	0.7500	0.2500	0.4330	4.6765	2.7000
3	5.20	45	0.7071	0.7071	0.5000	0.5000	0.5000	3.6770	3.6770
4	5.10	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.5500	4.4167
5	5.00	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.2941	4.8296
6	4.80	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	4.8000
7	4.80	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.2423	4.6364
8	4.80	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.4000	4.1569
9	4.80	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-3.3941	3.3941
10	5.10	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-4.4167	2.5500
11	5.80	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-5.6024	1.5012
12	6.50	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.5000	0.0000
13	7.30	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-7.0513	-1.8894
14	8.00	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-6.9282	-4.0000
15	8.50	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-6.0104	-6.0104
16	9.00	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-4.5000	-7.7942
17	9.10	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-2.3553	-8.7899
18	9.00	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-9.0000
19	8.70	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	2.2517	-8.4036
20	8.30	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	4.1500	-7.1880
21	7.60	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	5.3740	-5.3740
22	7.00	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	6.0622	-3.5000
23	6.50	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	6.2785	-1.6823
Suma	157.90		0.0000	0.0000	12.0000	12.0000	0.0000	-2.6775	-25.5205

Tabla 3.7

$$g(t) = 6.57 + 2.13 \cos(15t - 264.01)$$

$$\rho = 0.9750$$



Julio

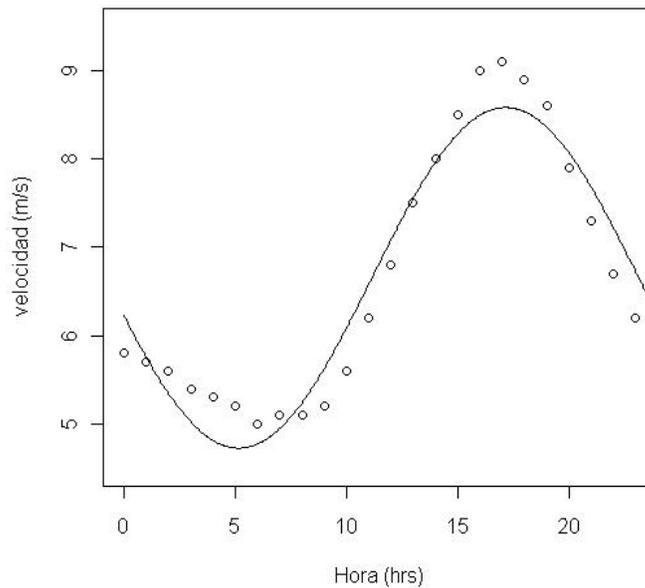
Hora (t)	Julio (y _i)	wt _i	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	5.80	0	1.0000	0.0000	1.0000	0.0000	0.0000	5.8000	0.0000
1	5.70	15	0.9659	0.2588	0.9330	0.0670	0.2500	5.5058	1.4753
2	5.60	30	0.8660	0.5000	0.7500	0.2500	0.4330	4.8497	2.8000
3	5.40	45	0.7071	0.7071	0.5000	0.5000	0.5000	3.8184	3.8184
4	5.30	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.6500	4.5899
5	5.20	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.3459	5.0228
6	5.00	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	5.0000
7	5.10	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.3200	4.9262
8	5.10	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.5500	4.4167
9	5.20	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-3.6770	3.6770
10	5.60	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-4.8497	2.8000
11	6.20	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-5.9887	1.6047
12	6.80	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.8000	0.0000
13	7.50	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-7.2444	-1.9411
14	8.00	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-6.9282	-4.0000
15	8.50	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-6.0104	-6.0104
16	9.00	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-4.5000	-7.7942
17	9.10	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-2.3553	-8.7899
18	8.90	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-8.9000
19	8.60	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	2.2258	-8.3070
20	7.90	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	3.9500	-6.8416
21	7.30	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	5.1619	-5.1619
22	6.70	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	5.8024	-3.3500
23	6.20	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	5.9887	-1.6047
Suma	159.70		0.0000	0.0000	12.0000	12.0000	0.0000	-5.1251	-22.5698

Tabla 3.8

$$g(t) = 6.65 + 1.92 \cos(15t - 257.20)$$

$$\rho = 0.9671$$

Día Típico de Julio



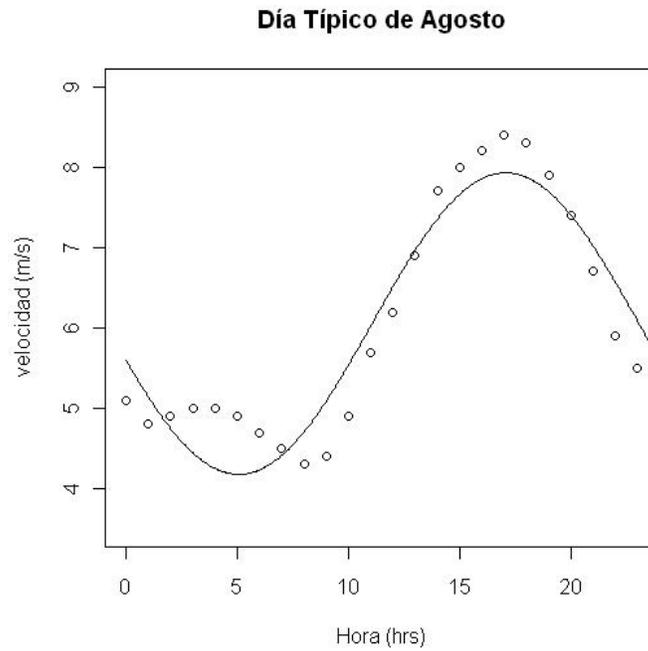
Agosto

Hora Agosto (t_i)	(y_i)	wt	c_i	s_i	c_i^2	s_i^2	$s_i c_i$	$y_i c_i$	$y_i s_i$
0	5.10	0	1.0000	0.0000	1.0000	0.0000	0.0000	5.1000	0.0000
1	4.80	15	0.9659	0.2588	0.9330	0.0670	0.2500	4.6364	1.2423
2	4.90	30	0.8660	0.5000	0.7500	0.2500	0.4330	4.2435	2.4500
3	5.00	45	0.7071	0.7071	0.5000	0.5000	0.5000	3.5355	3.5355
4	5.00	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.5000	4.3301
5	4.90	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.2682	4.7330
6	4.70	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	4.7000
7	4.50	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.1647	4.3467
8	4.30	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.1500	3.7239
9	4.40	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-3.1113	3.1113
10	4.90	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-4.2435	2.4500
11	5.70	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-5.5058	1.4753
12	6.20	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.2000	0.0000
13	6.90	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-6.6649	-1.7859
14	7.70	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-6.6684	-3.8500
15	8.00	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-5.6569	-5.6569
16	8.20	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-4.1000	-7.1014
17	8.40	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-2.1741	-8.1138
18	8.30	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-8.3000
19	7.90	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	2.0447	-7.6308
20	7.40	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	3.7000	-6.4086
21	6.70	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	4.7376	-4.7376
22	5.90	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	5.1095	-2.9500
23	5.50	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	5.3126	-1.4235
Suma	145.30		0.0000	0.0000	12.0000	12.0000	0.0000	-5.4513	-21.8603

Tabla 3.9

$$g(t) = 6.05 + 1.87 \cos(15t - 255.99)$$

$$\rho = 0.9462$$



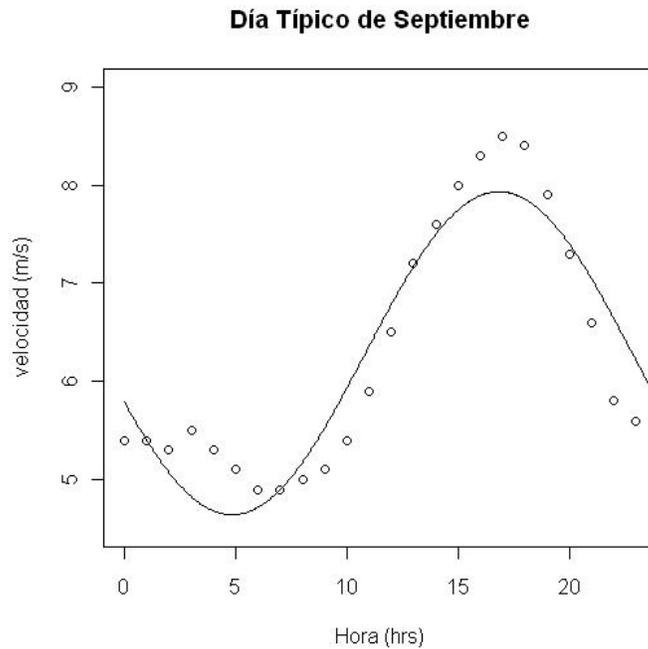
septiembre

Hora Septiembre (t)	(y _i)	wt	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	5.40	0	1.0000	0.0000	1.0000	0.0000	0.0000	5.4000	0.0000
1	5.40	15	0.9659	0.2588	0.9330	0.0670	0.2500	5.2160	1.3976
2	5.30	30	0.8660	0.5000	0.7500	0.2500	0.4330	4.5899	2.6500
3	5.50	45	0.7071	0.7071	0.5000	0.5000	0.5000	3.8891	3.8891
4	5.30	60	0.5000	0.8660	0.2500	0.7500	0.4330	2.6500	4.5899
5	5.10	75	0.2588	0.9659	0.0670	0.9330	0.2500	1.3200	4.9262
6	4.90	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	4.9000
7	4.90	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-1.2682	4.7330
8	5.00	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-2.5000	4.3301
9	5.10	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-3.6062	3.6062
10	5.40	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-4.6765	2.7000
11	5.90	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-5.6990	1.5270
12	6.50	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.5000	0.0000
13	7.20	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-6.9547	-1.8635
14	7.60	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-6.5818	-3.8000
15	8.00	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-5.6569	-5.6569
16	8.30	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-4.1500	-7.1880
17	8.50	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-2.2000	-8.2104
18	8.40	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-8.4000
19	7.90	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	2.0447	-7.6308
20	7.30	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	3.6500	-6.3220
21	6.60	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	4.6669	-4.6669
22	5.80	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	5.0229	-2.9000
23	5.60	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	5.4092	-1.4494
Suma	150.90		0.0000	0.0000	12.0000	12.0000	0.0000	-5.9345	18.8385

Tabla 3.10

$$g(t) = 6.28 + 1.64 \cos(15t - 252.51)$$

$$\rho = 0.9400$$



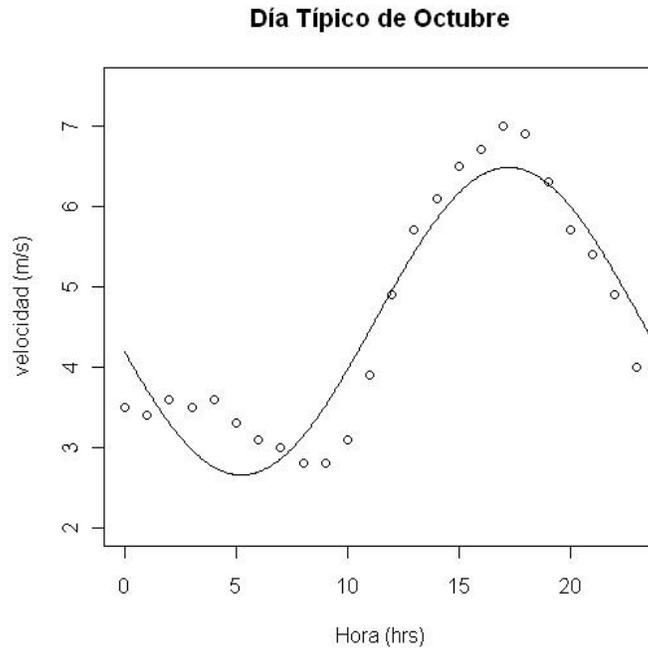
Octubre

Hora (t)	Octubre (y _i)	w _t	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	3.50	0	1.0000	0.0000	1.0000	0.0000	0.0000	3.5000	0.0000
1	3.40	15	0.9659	0.2588	0.9330	0.0670	0.2500	3.2841	0.8800
2	3.60	30	0.8660	0.5000	0.7500	0.2500	0.4330	3.1177	1.8000
3	3.50	45	0.7071	0.7071	0.5000	0.5000	0.5000	2.4749	2.4749
4	3.60	60	0.5000	0.8660	0.2500	0.7500	0.4330	1.8000	3.1177
5	3.30	75	0.2588	0.9659	0.0670	0.9330	0.2500	0.8541	3.1876
6	3.10	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	3.1000
7	3.00	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-0.7765	2.8978
8	2.80	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-1.4000	2.4249
9	2.80	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-1.9799	1.9799
10	3.10	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-2.6847	1.5500
11	3.90	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-3.7671	1.0094
12	4.90	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-4.9000	0.0000
13	5.70	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-5.5058	-1.4753
14	6.10	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-5.2828	-3.0500
15	6.50	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-4.5962	-4.5962
16	6.70	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-3.3500	-5.8024
17	7.00	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-1.8117	-6.7615
18	6.90	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.9000
19	6.30	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	1.6306	-6.0853
20	5.70	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	2.8500	-4.9363
21	5.40	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	3.8184	-3.8184
22	4.90	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	4.2435	-2.4500
23	4.00	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	3.8637	-1.0353
Suma	109.70		0.0000	0.0000	12.0000	12.0000	0.0000	-4.6176	-22.4886

Tabla 3.11

$$g(t) = 4.57 + 1.91 \cos(15t - 258.39)$$

$$\rho = 0.9432$$



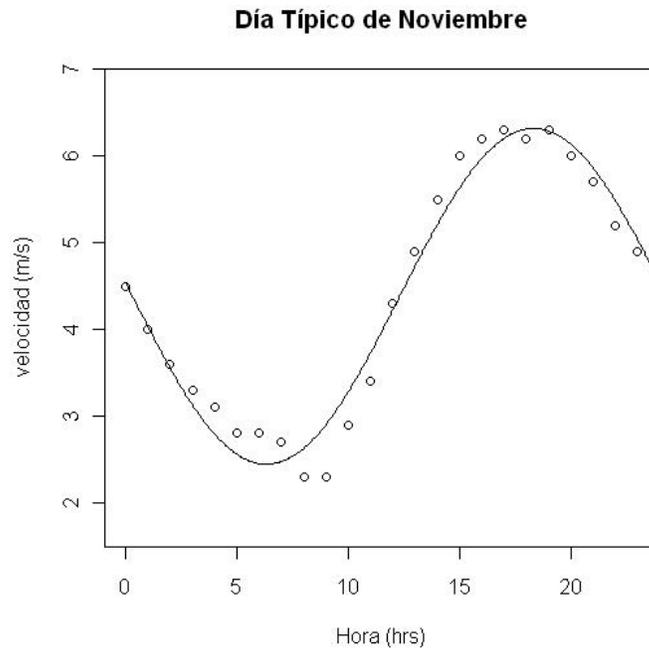
Noviembre

Hora (t)	Noviembre (y _i)	wt	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	4.50	0	1.0000	0.0000	1.0000	0.0000	0.0000	4.5000	0.0000
1	4.00	15	0.9659	0.2588	0.9330	0.0670	0.2500	3.8637	1.0353
2	3.60	30	0.8660	0.5000	0.7500	0.2500	0.4330	3.1177	1.8000
3	3.30	45	0.7071	0.7071	0.5000	0.5000	0.5000	2.3335	2.3335
4	3.10	60	0.5000	0.8660	0.2500	0.7500	0.4330	1.5500	2.6847
5	2.80	75	0.2588	0.9659	0.0670	0.9330	0.2500	0.7247	2.7046
6	2.80	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	2.8000
7	2.70	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-0.6988	2.6080
8	2.30	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-1.1500	1.9919
9	2.30	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-1.6263	1.6263
10	2.90	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-2.5115	1.4500
11	3.40	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-3.2841	0.8800
12	4.30	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-4.3000	0.0000
13	4.90	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-4.7330	-1.2682
14	5.50	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-4.7631	-2.7500
15	6.00	225	-0.7071	-0.7071	0.5000	0.5000	-0.5000	-4.2426	-4.2426
16	6.20	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-3.1000	-5.3694
17	6.30	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-1.6306	-6.0853
18	6.20	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.2000
19	6.30	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	1.6306	-6.0853
20	6.00	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	3.0000	-5.1962
21	5.70	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	4.0305	-4.0305
22	5.20	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	4.5033	-2.6000
23	4.90	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	4.7330	-1.2682
Suma	105.20		0.0000	0.0000	12.0000	12.0000	0.0000	1.9468	-23.1816

Tabla 3.12

$$g(t) = 4.38 + 1.93 \cos(15t - 274.80)$$

$$\rho = 0.9831$$



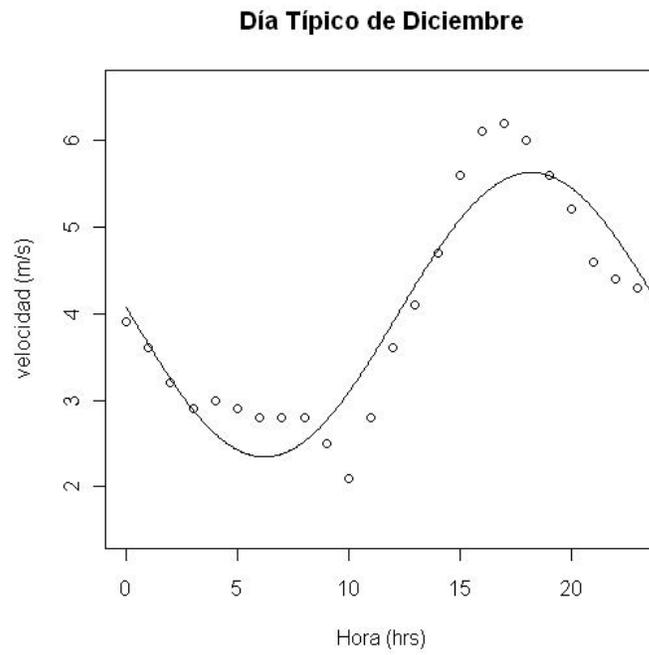
Diciembre

Hora (t)	Diciembre (y _i)	wt _i	c _i	s _i	c _i ²	s _i ²	s _i c _i	y _i c _i	y _i s _i
0	3.90	0	1.0000	0.0000	1.0000	0.0000	0.0000	3.9000	0.0000
1	3.60	15	0.9659	0.2588	0.9330	0.0670	0.2500	3.4773	0.9317
2	3.20	30	0.8660	0.5000	0.7500	0.2500	0.4330	2.7713	1.6000
3	2.90	45	0.7071	0.7071	0.5000	0.5000	0.5000	2.0506	2.0506
4	3.00	60	0.5000	0.8660	0.2500	0.7500	0.4330	1.5000	2.5981
5	2.90	75	0.2588	0.9659	0.0670	0.9330	0.2500	0.7506	2.8012
6	2.80	90	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	2.8000
7	2.80	105	-0.2588	0.9659	0.0670	0.9330	-0.2500	-0.7247	2.7046
8	2.80	120	-0.5000	0.8660	0.2500	0.7500	-0.4330	-1.4000	2.4249
9	2.50	135	-0.7071	0.7071	0.5000	0.5000	-0.5000	-1.7678	1.7678
10	2.10	150	-0.8660	0.5000	0.7500	0.2500	-0.4330	-1.8187	1.0500
11	2.80	165	-0.9659	0.2588	0.9330	0.0670	-0.2500	-2.7046	0.7247
12	3.60	180	-1.0000	0.0000	1.0000	0.0000	0.0000	-3.6000	0.0000
13	4.10	195	-0.9659	-0.2588	0.9330	0.0670	0.2500	-3.9603	-1.0612
14	4.70	210	-0.8660	-0.5000	0.7500	0.2500	0.4330	-4.0703	-2.3500
15	5.60	225	-0.7071	-0.7071	0.5000	0.5000	0.5000	-3.9598	-3.9598
16	6.10	240	-0.5000	-0.8660	0.2500	0.7500	0.4330	-3.0500	-5.2828
17	6.20	255	-0.2588	-0.9659	0.0670	0.9330	0.2500	-1.6047	-5.9887
18	6.00	270	0.0000	-1.0000	0.0000	1.0000	0.0000	0.0000	-6.0000
19	5.60	285	0.2588	-0.9659	0.0670	0.9330	-0.2500	1.4494	-5.4092
20	5.20	300	0.5000	-0.8660	0.2500	0.7500	-0.4330	2.6000	-4.5033
21	4.60	315	0.7071	-0.7071	0.5000	0.5000	-0.5000	3.2527	-3.2527
22	4.40	330	0.8660	-0.5000	0.7500	0.2500	-0.4330	3.8105	-2.2000
23	4.30	345	0.9659	-0.2588	0.9330	0.0670	-0.2500	4.1535	-1.1129
Suma		95.70	0.0000	0.0000	12.0000	12.0000	0.0000	1.0551	-19.6670

Tabla 3.13

$$g(t) = 3.98 + 1.64 \cos(15t - 273.07)$$

$$\rho = 0.9356$$



CONCLUSIONES

En el presente trabajo se eligieron los modelos de regresión circular-lineal (regresión periódica) para ser desarrollados con más detalle, debido a que existen más referencias para éstos y son de gran utilidad (tanto en la práctica como en la investigación) en diversas áreas de estudio.

Por lo anterior, este texto puede utilizarse como una buena referencia para introducir al lector al tema del análisis de regresión para variables circulares, y además como apoyo en el desarrollo de modelos de regresión periódica para cualquier área de conocimiento.

En el trabajo se ha incluido una aplicación con datos reales del modelo de regresión circular-lineal, en la que se ajustaron 12 modelos, aplicables a cada uno de los meses del año, para la relación que existe entre la hora del día y la velocidad del viento, en la región de Guerrero Negro en Baja California Sur. Los modelos obtenidos en dicha aplicación fueron construídos para ilustrar el proceso a seguir en el ajuste de los mismos, así como para ejemplificar la utilidad de este tipo de regresión.

Se obtuvieron muy buenos ajustes en la aplicación, lo que sugiere una relación circular entre la hora del día y la velocidad del viento. Por lo que, estos modelos dan una idea bastante aproximada del comportamiento de las variables involucradas, en esa zona del país.

Los ejemplos mostrados en este texto, abarcan diversas áreas como la medicina, biología y meteorología. Con estos, se ha pretendido transmitir al lector la importancia del estudio y desarrollo de estos temas, no sólo en las áreas ya mencionadas, sino también en áreas como la geología, geografía e incluso física.

GLOSARIO

Acrofase Parámetro que define el momento de máxima elevación de una variable. En otras palabras, la acrofase es el tiempo en el que la función alcanza su máximo.

Amplitud Se le denomina así a la mitad de la diferencia entre el punto más alto y el punto más bajo del modelo matemático. Este modelo matemático es en nuestro caso el modelo teórico.

Barber´s pole Es un tipo de símbolo usado por los barberos o peluqueros en muchas partes del mundo, es un poste con franjas blancas y rojas que se encuentran rodeandolo en forma diagonal (como los clásicos bastones de dulce). Se utiliza este término para varias cosas que tienen o resultan de esta forma.

Ciclo Es una sucesión de acontecimientos que tienen lugar de forma repetitiva y siempre en el mismo orden, sin tener en cuenta el tiempo en que dicha sucesión se repite.

Estadística Circular Es el análisis y el estudio de datos que se distribuyen sobre una circunferencia, es decir, datos cuya mejor representación es como puntos en una circunferencia.

Frecuencia angular Es el número de ciclos que se encuentran contenidos en una rotación completa del círculo (ya sea ésta 360° o 2π).

Mesor Se define como el valor intermedio entre el valor más alto y el más bajo del fenómeno ajustado a una función matemática. En nuestro estudio, dicha función matemática se refiere a lo que le hemos llamado el modelo teórico.

Periodo Intervalo de tiempo en el cual se repite un fenómeno cíclico o intervalo angular en el cual vuelve a ocurrir un fenómeno de direcciones.

Ritmo Es el nombre que se le da a un ciclo cuando éste ocurre en un intervalo de tiempo constante y previsible.

Ritmo biológico Se refieren a los cambios cíclicos que experimentan los distintos procesos fisiológicos y que ocurren en un intervalo de tiempo determinado. Según la periodicidad de estos cambios dividimos a los ritmos biológicos en:

Ritmo circadiano Es aquel que tiene una frecuencia próxima a la diaria, es decir, que presenta un periodo de entre 20 y 28 hrs. Por ejemplo, proceso sueño-vigilia, frecuencia cardiaca, etc.

Ritmo infradiano Es aquel cuya frecuencia es inferior a la diaria, es decir, el que presenta un periodo superior a las 28 hrs. El ejemplo más común es el ciclo menstrual.

Ritmo ultradiano Es el que presenta una frecuencia superior a la diaria, es decir, aquel cuyo periodo es inferior a las 20 hrs. Por ejemplo, secreción de neurotransmisores, ritmo de la respiración, etc.

Variable lineal Le llamamos variables lineales a las que pueden ser representadas en la recta real, le damos este término a las variables que no son circulares para así poder diferenciarlas.

ANEXO

IDENTIDADES TRIGONOMÉTRICAS

$$\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \operatorname{sen}(\alpha)\operatorname{sen}(\beta) \quad (3.0.6)$$

$$\operatorname{sen}^2(\alpha) + \cos^2(\alpha) = 1 \quad (3.0.7)$$

$$\tan(\phi) = \operatorname{sen}(\phi)/\cos(\phi) \quad (3.0.8)$$

FUNCIÓN BESSEL

La función Bessel de v -ésimo orden en su presentación integral es:

$$I_v(z) = \frac{(z/2)^v}{\Gamma(1/2)\Gamma(v+1/2)} \int_0^\pi e^{iz \cos \theta} \operatorname{sen}^{2v} \theta \, d\theta \quad \left(v > -\frac{1}{2}\right) \quad (3.0.9)$$

Y la expansión en serie de la misma está representada por:

$$I_v(z) = \sum_{r=0}^{\infty} (-1)^r \frac{(z/2)^{v+2r}}{r! \Gamma(v+r+1)} \quad (3.0.10)$$

NÚMEROS COMPLEJOS

Existen diversas formas de representar a los números complejos. La forma más común es la llamada forma *binomial* o *rectangular* cuya expresión es:

$$a + bi \quad \text{con } a, b \in \mathfrak{R}$$

Otra manera de representarlos es la conocida como *r cis θ* o *polar* cuya forma es:

$$r(\cos \theta + i \operatorname{sen} \theta)$$

donde θ es el ángulo formado por el eje real y el eje imaginario y r el modulo del número complejo. Por último se encuentra la expresión:

$$re^{i\theta}$$

a la que se le conoce como *fórmula de Euler*.

Podemos resumir lo anterior con la siguiente ecuación:

$$a + bi = r(\cos \theta + i \operatorname{sen} \theta) = re^{i\theta} \quad (3.0.11)$$

Es importante mencionar que la manera más usual de expresar al ángulo θ de las expresiones anteriores es en radianes.

Referencias

- [1] BATSCHELET, E. (1981). *Circular Statistics in Biology*. Academic Press, London.
- [2] JAMMALAMADAKA, S. R. & SENGUPTA, A. (2001). *Topics in Circular Statistics*. World Scientific Publishing Co. Pte. Ltd, Singapore.
- [3] SOLOMON, A. D. (1989). *The Essentials of Fourier Analysis*. Research and Education Association, New Jersey.
- [4] RIVEST, L. P. (1997). *A Decentred Predictor for Circular-Circular Regression*. *Biometrika*. **84**, 717-726.
- [5] FISHER, N. I. & LEE, A. J. (1992). *Regression Models for an Angular Response*. *Biometrics*. **48**, 665-677.
- [6] GOULD, A. L. (1969). *A Regression Technique for Angular Variates*. *Biometrics*. 683-700.
- [7] DOWNS, T. D. & MARDIA, K. V. (2002). *Circular Regression*. *Biometrika*. **89**, 683-697.
- [8] JOHNSON, R. A. & WEHRLY T. E. (1978). *Some Angular-Linear Distributions and Related Regression Models*. *Journal of the American Statistical Association*. **73**, 602-606.