



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:

Actuario

PRESENTA:

Guillermo Alfonso Martínez Contreras

TUTORA

Mat. Margarita Elvira Chávez Cano



CIUDAD UNIVERSITARIA, CD.MX.  
2018



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.

1.Datos del alumno.

Martínez

Contreras

Guillermo Alfonso

044-55-38-03-25-59

Universidad Nacional Autónoma de

México

Facultad de Ciencias

Actuaría

310125963

2.Datos del Tutor

Mat

Margarita Elvira

Chávez

Cano

3.Datos del sinodal 1

Dra

Lizbeth

Naranjo

Albarrán

4.Datos del sinodal 2

Dra

María del Pilar

Alonso

Reyes

5.Datos del sinodal 3

Dr.

Arrigo

Coen

Coria

6.Datos del sinodal 4

Act.

Edna Gabriela

López

Estrada

7.Datos del trabajo escrito.

Modelos de Regresión para Datos Circulares

92p

2018



*DEDICATORIA*

Siempre me guiaste por el mejor camino, sin tu apoyo, fuerzas y dedicación  
llegar aquí no hubiera sido posible, siempre serás un gran ejemplo a seguir,  
gracias mi querida madre.

# Agradecimientos

A mi asesora la Mat. Margarita Elvira Chávez Cano a quien agradezco el apoyo, los conocimientos y el tiempo dedicado para la realización de esta tesis, así como la influencia que representó en mi persona por el gusto en la estadística; le estaré siempre agradecido.

A mis sinodales por su tiempo y conocimientos compartidos para la elaboración de este trabajo.

A las personas (amigos, familia y profesores) que han estado a mi lado siempre apoyándome y guiándome en este camino pues sin su empuje y confianza esto no hubiera sido posible.

A mi Alma Máter, la Universidad Nacional Autónoma de México, quién es su Facultad de Ciencias me permitió obtener la mejor formación, siempre te pondré en alto.

# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Estadística Descriptiva</b>	<b>3</b>
1.1. Representación Gráfica . . . . .	3
1.1.1. Gráfica de Datos Originales . . . . .	3
1.1.2. Histogramas Lineales . . . . .	4
1.1.3. Diagrama de la Rosa . . . . .	5
1.1.4. Histograma Circular . . . . .	5
1.2. Valores Descriptivos . . . . .	7
1.2.1. Medidas de Tendencia Central . . . . .	8
1.2.2. Medidas de Dispersión . . . . .	10
<b>2. Distribuciones Circulares</b>	<b>11</b>
2.1. Introducción . . . . .	11
2.2. Distribuciones Circulares . . . . .	12
2.2.1. Distribución Uniforme . . . . .	13
2.2.2. Distribución Cardioide . . . . .	14
2.2.3. Distribución von Mises . . . . .	14
<b>3. Modelos de Regresión</b>	<b>17</b>
3.1. Regresión Circular-Lineal . . . . .	17
3.1.1. Ajuste de un Polinomio Trigonométrico . . . . .	20
3.1.2. Ejemplo . . . . .	24
3.1.3. Polinomio con 1 <sup>ero</sup> y 2 <sup>do</sup> Armónico . . . . .	26
3.2. Correlación Circular-Lineal . . . . .	31
3.3. Regresión Circular-Circular . . . . .	32
3.3.1. Estimación de los Coeficientes de Regresión . . . . .	33
3.3.2. Casos Especiales . . . . .	36
3.3.3. Estimación de la matriz $\mathbb{T}$ . . . . .	36

3.3.4. Determinación de $m$ . . . . .	37
3.4. <i>MAE</i> una Medida de Bondad de Ajuste . . . . .	40
<b>4. Aplicación de los Modelos</b> . . . . .	<b>41</b>
4.1. Regresión Circular-Lineal . . . . .	41
4.1.1. Primera Aplicación . . . . .	41
4.1.2. Segunda Aplicación . . . . .	52
4.2. Regresión Circular-Circular . . . . .	59
4.2.1. Primera Aplicación . . . . .	59
4.2.2. Segunda Aplicación . . . . .	65
<b>Conclusiones</b> . . . . .	<b>69</b>
<b>A. Códigos en R</b> . . . . .	<b>71</b>
A.1. Gráficas . . . . .	71
A.1.1. Figura 2.1, 2.2 y 2.3 . . . . .	71
A.1.2. Figura 3.2 . . . . .	72
A.2. Aplicaciones . . . . .	72
A.2.1. 1 <sup>ra</sup> Aplicación Regresión C-L . . . . .	72
A.2.2. 2 <sup>da</sup> Aplicación Regresión C-L (Temperatura Máxima). . . . .	76
A.2.3. 2 <sup>da</sup> Aplicación Regresión C-L (Temperatura Mínima). . . . .	80
A.2.4. 1 <sup>ra</sup> Aplicación Regresión C-C . . . . .	83
A.2.5. 2 <sup>da</sup> Aplicación Regresión C-C . . . . .	83
<b>Bibliografía</b> . . . . .	<b>84</b>



# Introducción

La estadística circular es una rama de la estadística que trata con datos que se pueden representar como puntos en la circunferencia del círculo unitario, los datos de este tipo son llamados “circulares” para ser distinguidos de los datos lineales utilizados con mayor frecuencia, teniendo así como soporte el círculo unitario y no la recta de los números reales. Estos datos necesitan ser analizados de forma adecuada pues tienen características especiales que deben ser analizadas de diferente manera a la estadística tradicional.

Los datos circulares pueden ser encontrados en diversos campos, algunos de estos son: la biología, la física, la geografía, la medicina, la ecología, la geología etcétera; algunos ejemplos de datos circulares y aplicaciones en las áreas mencionadas son:

- Estudio de trayectorias de animales.
- Dirección del viento, variable que es importante en el estudio de la contaminación del medio ambiente.
- Estudio de fenómenos con ocurrencia periódica.
- La incidencia de la aparición de una enfermedad en particular en varios momentos del año.
- Estudios de la ocurrencia de terremotos en alguna zona.
- Estudios de percepción bajo condiciones específicas, por ejemplo la simulación de la gravedad cero.

En la estadística circular se tienen tres tipos de modelos cuando se tiene una única variable dependiente, los cuales son:

- Modelo de regresión circular-circular.
- Modelo de regresión circular-lineal.
- Modelo de regresión lineal-circular.

El nombre que recibe cada uno de los modelos está dado por el tipo de variable dependiente y por el tipo de variable independiente, es decir el modelo de regresión circular-lineal hace referencia al modelo de regresión en el cual la variable dependiente es de tipo circular y la variable independiente es de tipo lineal.

La presente tesis está dividida en cuatro capítulos: En el primer capítulo se presentan las herramientas principales de la estadística descriptiva para datos circulares, las cuales dan una descripción general de los datos. En el segundo capítulo se presentan los conceptos básicos sobre distribuciones de probabilidad circulares y se mencionan algunas de ellas, así como sus características. En el tercer capítulo se encuentran dos modelos de regresión para datos circulares, el primero es el modelo de regresión circular-lineal, se muestra el desarrollo del modelo el cual es ilustrado con un ejemplo sobre el ritmo circadiano en roedores, posteriormente se muestra una medida de correlación entre una variable lineal y una circular; el segundo es el modelo de regresión circular-circular, y se considera su desarrollo del modelo, así como algunos casos especiales que éste puede tener. En el cuarto capítulo se encuentran dos aplicaciones para cada uno de los modelos, se muestra de manera detallada cómo se llevó a cabo su aplicación con ayuda del software *R*.

Por último se presentan las conclusiones de este trabajo, las referencias utilizadas y el código de las aplicaciones implementadas en el software *R* para el desarrollo de los ejemplos.

# Capítulo 1

## Estadística Descriptiva

El uso de la estadística descriptiva aplicada a datos circulares brinda una descripción de éstos para poder realizar un análisis del conjunto de datos.

Dentro del análisis es importante tener una representación gráfica de la información los datos, así como resumir, de forma numérica, algunos de sus aspectos más relevantes, ya que al hacer esto se tiene una idea inicial de las características que los datos pueden tener.

### 1.1. Representación Gráfica

La representación gráfica de los datos es importante pues permite tener una idea de las características que la muestra aporta, tales como posibles concentraciones de los datos, y con ello ayudar a identificar una posible distribución que ajuste a los datos o poder observar si la muestra es unimodal o multimodal.

Las representaciones gráficas más usadas en la estadística circular son las siguientes:

#### 1.1.1. Gráfica de Datos Originales

Es la representación más simple de los datos circulares, ésta representación consiste en graficar un diagrama de dispersión de las observaciones, cada una de estas se representa como un punto sobre el círculo unitario.

En la Tabla 1.1 se tiene la dirección medida en grados que tomaron 76 tortugas después de haber desovado. La Figura 1.1 representa gráficamente estos datos.

---

<sup>1</sup>Fisher N.I.(1993). Statistical Analysis of Circular Data. Cambridge University Press.

---



---

8	9	13	13	14	18	22	27	30	34
38	38	40	44	45	47	48	48	48	48
50	53	56	57	58	58	61	63	64	64
64	65	65	68	70	73	78	78	78	83
83	88	88	88	90	92	92	93	95	96
98	100	103	106	113	116	113	118	138	153
153	155	204	215	223	226	237	238	243	244
250	251	257	268	285	319	343	350		

---



---

Tabla 1.1: Orientación de 76 tortugas después de haber desovado

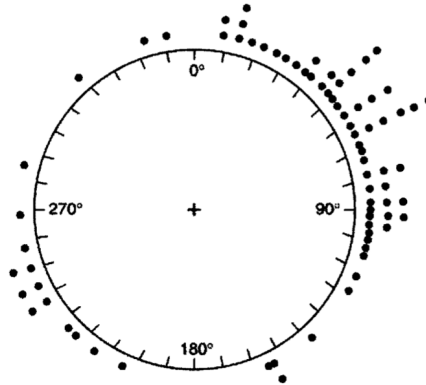


Figura 1.1: Representación de la orientación que tomaron 76 tortugas después de haber desovado.<sup>1</sup>

### 1.1.2. Histogramas Lineales

Debido a la gran experiencia que se ha adquirido en la interpretación de de un histograma en la estadística lineal resulta de gran utilidad el poder representar los datos de manera análoga.

Los histogramas lineales se construyen considerando los datos como si fueran “lineales” y no angulares, por lo que no se grafican en círculo sino en dos ejes, para su construcción se selecciona el intervalo que va de  $0^\circ$  a  $360^\circ$  y se transforma en intervalos de tiempo, los cuales también se agrupan en intervalos. En el eje de las ordenadas se graficarán las frecuencias y en el eje de las abscisas los datos de

tiempo o de orientación. La Figura 1.2 muestra el histograma lineal con los datos de la Tabla 1.2.

### 1.1.3. Diagrama de la Rosa

Una forma muy usual de representar un histograma circular es por medio de un diagrama de rosa en el cual las barras usadas en un histograma son reemplazadas por sectores, el área en los sectores será determinada por la frecuencia de los datos que presenten en su clase. La Figura 1.3 muestra el diagrama de la rosa con los datos de la Tabla 1.2

### 1.1.4. Histograma Circular

Los datos circulares agrupados pueden ser representados por este histograma, donde cada barra está centrada en el punto medio de su correspondiente grupo de ángulos y el área de esta barra es proporcional a la frecuencia que se tiene de los datos en su grupo y tendrán forma rectangular. Este histograma es muy similar al histograma lineal, ya que pareciera que no es más que un círculo con un histograma lineal envolviéndolo. La Figura 1.4 muestra el histograma circular con los datos de la Tabla 1.2

La visualización de las tres gráficas anteriores se llevará acabo con los datos presentados en la Tabla 1.2.

En la Tabla 1.2 se muestran datos de la dirección en que desaparecen en el horizonte 714 aves del tipo Ánades Reales Británicos<sup>2</sup>.

---

<sup>2</sup>Mardia K.V (1975). Statistics of Directional Data. London: Academic Press.

Dirección (en grados)	No. de aves	Dirección (en grados)	No. de aves
0-19	40	180-199	3
20-39	22	200-219	11
40-59	20	220-239	22
60-79	9	240-259	24
80-99	6	260-279	58
100-119	3	280-299	136
120-139	3	300-319	138
140-159	1	320-339	143
160-179	6	340-359	69
		Total	714

Tabla 1.2: Dirección en que desaparecen en el horizonte 714 Ánades Reales Británicos

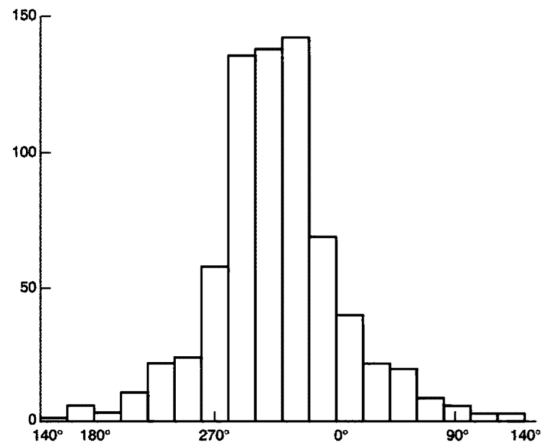


Figura 1.2: Histograma lineal de la dirección en que desaparecen en el horizonte 714 Ánades Reales Británicos.

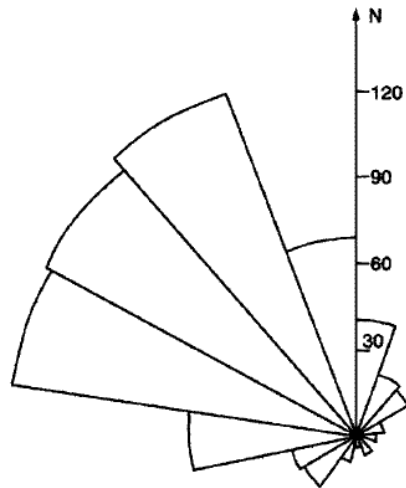


Figura 1.3: Diagrama de rosa de la dirección en que desaparecen en el horizonte 714 Ánades Reales Británicos.

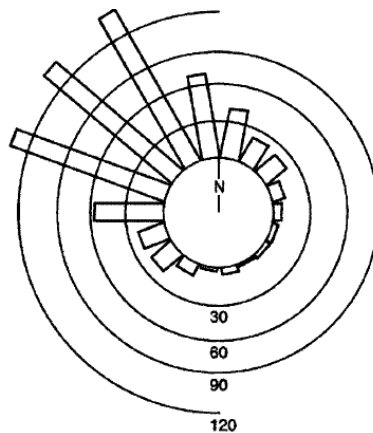


Figura 1.4: Histograma circular de la dirección en que desaparecen en el horizonte 714 Ánades Reales Británicos.

## 1.2. Valores Descriptivos

Los valores descriptivos o medidas descriptivas son valores numéricos que se obtienen a partir de los datos, estas medidas proporcionan características impor-

tantes así como dirección y dispersión, entre otros, la finalidad de calcular estos valores, es poder reducir en cifras las características importantes que ayudan a entender de mejor manera el comportamiento de los datos.

### 1.2.1. Medidas de Tendencia Central

#### 1. Dirección Media

Consideremos la siguiente muestra de valores dados en ángulos:

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

Se busca definir el valor de la dirección media, veamos la media aritmética:

$$\frac{1}{3}(\phi_1 + \phi_2 + \phi_3) = 160^\circ,$$

es claro que este resultado no es algo aceptable, reemplazando a  $\phi_2$  por su ángulo equivalente  $-10^\circ$  al realizar esto el valor cambia a  $40^\circ$  el cual es un resultado de cierta manera aceptable, sin embargo de manera general es necesario definir un procedimiento apropiado para poder calcular la **dirección media**.

Por lo anterior, en la estadística circular el proceso para definir la dirección media, es muy diferente al de la estadística lineal, se define un proceso adecuado para calcularla de tal forma que la dirección media coincida con el centro de masa de los datos.

Se representa cada observación como un vector sobre el círculo unitario y se obtiene el vector resultante de la suma vectorial de las observaciones; de esta forma se calculan las coordenadas asociadas a la dirección media.

Sea  $\theta_1, \dots, \theta_n$  un conjunto de observaciones circulares medidas en ángulos, con vectores unitarios  $e_1, \dots, e_n$ , tales que, aplicando la transformación de coordenadas polares a rectangulares a cada  $e_i$  se tiene que sus coordenadas quedan expresadas de la siguiente manera:

$$(\cos(\theta_j), \text{sen}(\theta_j)), \quad j = 1, \dots, n$$

entonces

$$R = \left( \sum_{i=1}^n \cos(\theta_i), \sum_{i=1}^n \text{sen}(\theta_i) \right) = (C, S),$$

donde:

$$C = \sum_{i=1}^n \cos(\theta_i) \quad \text{y} \quad S = \sum_{i=1}^n \text{sen}(\theta_i),$$



y se obtiene la longitud del vector resultante:

$$R' = \| R \| = \sqrt{C^2 + S^2}.$$

Considerando que las componentes rectangulares de cada vector unitario están dadas por  $(\cos(\theta_i), \text{sen}(\theta_i))$ , las componentes para la dirección media serán:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \quad \text{y} \quad \bar{S} = \frac{1}{n} \sum_{i=1}^n \text{sen}(\theta_i),$$

así el ángulo del vector resultante,  $\bar{\theta}$  será el ángulo medio de la muestra y está dado por las ecuaciones:

$$\cos(\bar{\theta}) = \frac{\bar{C}}{\bar{R}} \quad \text{y} \quad \text{sen}(\bar{\theta}) = \frac{\bar{S}}{\bar{R}}$$

donde la longitud media  $\bar{R}$  es:

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}.$$

Una forma más explícita de  $\bar{\theta}$  está dada por:

$$\bar{\theta} = \begin{cases} \tan^{-1}(\bar{S}/\bar{C}) & \text{si } \bar{C} \geq 0 \\ \pi + \tan^{-1}(\bar{S}/\bar{C}) & \text{si } \bar{C} < 0, \end{cases}$$

donde la función inversa de la tangente “ $\tan^{-1}$ ” toma valores en  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ .

## 2. Longitud Media

La longitud media  $\bar{R}$  se definió como:

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2},$$

como los vectores están en el círculo unitario se cumple que  $0 \leq \bar{R} \leq 1$ . Si las direcciones  $\theta_1, \dots, \theta_n$  se encuentran estrechamente agrupadas, el valor de  $\bar{R}$  será muy cercano a 1, de la misma manera si  $\theta_1, \dots, \theta_n$  se encuentran muy dispersas  $\bar{R}$  tendrá valores muy cercanos a 0.

## 3. Moda Angular

El ángulo modal es el ángulo que más se repite y puede haber más de una moda en la muestra.

## 1.2.2. Medidas de Dispersión

### 1. Varianza Circular

Se considera a la “longitud media” como una medida de concentración. Entonces, se define a la varianza como una medida de dispersión. Como  $\bar{R}$  disminuye de 1 a 0 cuando la dispersión aumenta, entonces la varianza queda definida de la siguiente manera:

$$V = 1 - \bar{R},$$

con

$$0 \leq V \leq 1.$$

### 2. Desviación Estándar Circular

Esta medida se obtiene de la transformación de la *varianza circular*, que está dada como:

$$v = (-2 \log(1 - V))^{\frac{1}{2}} = (-2 \log \bar{R})^{\frac{1}{2}}.$$

Notemos que  $v$  toma valores en  $[0, \infty]$ , para una  $V$  (varianza circular) muy pequeña,  $v$  queda reducida de la siguiente manera:

$$v \simeq (2V)^{\frac{1}{2}}.$$

# Capítulo 2

## Distribuciones Circulares

### 2.1. Introducción

Una distribución circular es una distribución de probabilidad cuya probabilidad total se encuentra en la circunferencia del círculo unitario, los valores que pertenecen al rango están situados entre 0 y  $2\pi$ .

Las distribuciones circulares son esencialmente de dos tipos: discretas ó continuas. Las distribuciones discretas asignan probabilidad a un número finito de direcciones; en las distribuciones continuas la función de densidad deberá cumplir las siguientes propiedades:

1.  $f(\theta) \geq 0$ .
2.  $\int_0^{2\pi} f(\theta)d\theta = 1$ .
3.  $f(\theta) = f(\theta + (k * 2\pi))$ , para cualquier entero  $k$  (i.e.,  $f$  es periódica).

Al igual que en las distribuciones lineales existe una relación entre la función de densidad  $f(\theta)$  y la función de distribución  $F(\theta)$  la cual queda determinada por la siguiente igualdad:

$$F(\theta) = \int_0^{\theta} f(\beta)d\beta.$$

La función de distribución  $F(\theta)$  cumplirá que:

$$F(0) = 0 \quad \text{y} \quad F(2\pi) = 1.$$

La función de distribución  $F(\theta)$  correspondiente a  $f(\theta)$  (función de densidad) se

define sobre un intervalo  $(\theta_1, \theta_2)$  como:

$$F(\theta_2) - F(\theta_1) = \int_{\theta_1}^{\theta_2} f(\theta) d\theta.$$

Una distribución circular en ocasiones se puede generar a partir de distribuciones conocidas sobre la recta o el plano, por diferentes métodos, algunos de ellos son:

1. Envolviendo al círculo unitario con una distribución lineal (*wrapping*).
2. Por medio de una transformación de una variable aleatoria lineal bivariada a sus componentes direccionales, las cuales son llamadas distribuciones de “desplazamiento” (*offset*).
3. Iniciando una distribución sobre la recta, se aplica una función inyectiva que sea una correspondencia a cada punto  $x \in \mathbb{R}$  en un punto  $\theta$  del círculo unitario, y se ajustan los valores  $-\infty$  e  $\infty$  de manera que correspondan con  $2\pi$ .

### Distribución Envuelta

Dada cualquier distribución en la recta, será posible envolverla alrededor de la circunferencia del círculo unitario; i.e. si  $X$  es una variable aleatoria sobre la recta con función de densidad  $h(x)$  y función de distribución  $H(x)$ , entonces la correspondiente variable aleatoria  $\theta$  estará dada por:

$$\theta = X(\text{mod } 2\pi).$$

La función de distribución estará dada por:

$$F(\theta) = \sum_{k=-\infty}^{\infty} [H(\theta + 2k\pi) - H(2k\pi)].$$

La función de densidad estará dada por:

$$f(\theta) = \sum_{k=-\infty}^{\infty} h(\theta + 2k\pi).$$

## 2.2. Distribuciones Circulares

A continuación se definen tres distribuciones importantes para datos circulares: la uniforme, la cardioide y la von Mises.

### 2.2.1. Distribución Uniforme

La distribución uniforme es la distribución más básica en el círculo, ésta es la única que es invariante bajo rotación y reflexión. Esta distribución será de utilidad si todas las direcciones entre 0 y  $2\pi$  pueden ser elegidas con la misma probabilidad, en consecuencia con esto la función de densidad será constante sobre toda la circunferencia. La función de densidad para una variable aleatoria  $\theta$  se define como:

$$f(\theta) = \frac{1}{2\pi} \quad \text{para} \quad 0 \leq \theta < 2\pi.$$

Su función de distribución está dada por:

$$F(\theta) = \frac{\theta}{2\pi} \quad \text{para} \quad 0 \leq \theta < 2\pi.$$

La Figura 2.1 muestra la gráfica de la función de densidad Uniforme.

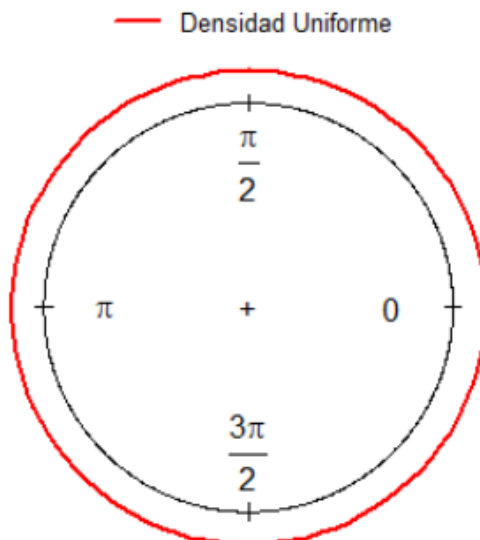


Figura 2.1: Gráfica de la función de densidad Uniforme

## 2.2.2. Distribución Cardioide

La distribución cardioide, recibe su nombre por el gran parecido con la curva cardioide, dependerá de dos parámetros  $\mu$  y  $\rho$ , cuya función de densidad es:

$$f(\theta; \mu, \rho) = \frac{1}{2\pi}(1 + 2\rho \cos(\theta - \mu)),$$

sus parámetros toman valores en:

$$0 \leq \mu < 2\pi, \quad -\frac{1}{2} < \rho < \frac{1}{2}.$$

Esta función alcanza su máximo cuando  $\cos(\theta) = 1$ , i.e. cuando  $\theta = \mu$ .

Su función de distribución está dada por:

$$F(\theta) = \frac{\theta}{2\pi} + \frac{\rho}{\pi} \text{sen}(\theta - \mu) \quad \text{para } 0 \leq \theta < 2\pi.$$

Donde

- La longitud media es  $\rho$ .
- La dirección media es  $\mu$ .
- La distribución es simétrica y unimodal en  $\mu$  (si  $\rho > 0$ ).
- Si  $\rho = 0$  la distribución se reduce a la distribución Uniforme. La Figura 2.2 muestra la gráfica de la función de densidad Cardioide con parámetro  $\rho=.4$  y  $\mu=5$ .

## 2.2.3. Distribución von Mises

La distribución von Mises una distribución simétrica y unimodal; esta distribución tiene un papel similar a la distribución normal en la Estadística lineal porque también es conocida como “**Distribución Normal Circular**”, tiene como función de densidad:

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}$$

con  $0 \leq \theta < 2\pi$  y  $0 \leq \kappa < \infty$  donde

- $I_0$  denota la función Bessel modificada tipo uno y orden 0,

$$I_0(\kappa) = \int_0^{2\pi} e^{[\kappa \cos(\theta)]} d\theta.$$

- El parámetro  $\mu$  es la dirección media.
- $\kappa$  es el parámetro de concentración.

La función de distribución es:

$$F(\theta) = \frac{1}{2\pi I_0(\kappa)} \int_0^\theta e^{[\kappa \cos(\beta - \mu)]} d\beta.$$

Se observa que la distribución uniforme es un caso particular de la distribución von Mises y esto ocurre cuando  $\kappa = 0$ ,

$$f(\theta) = \frac{1}{2\pi I_0(0)} e^{0 \cos(\theta - \mu)} = \frac{1}{2\pi}.$$

La Figura 2.3 muestra la gráfica de la función de densidad von Mises.

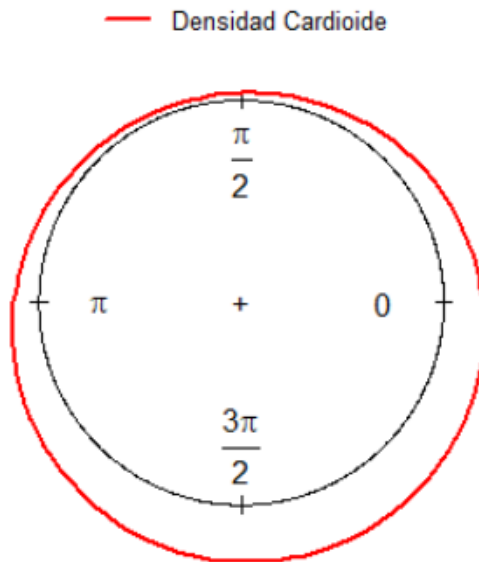


Figura 2.2: Gráfica de la función de densidad Uniforme

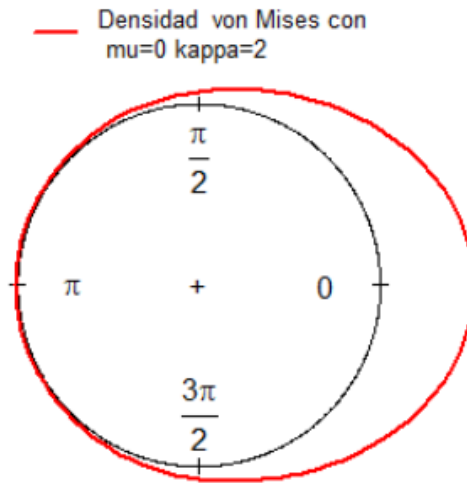


Figura 2.3: Función de densidad von Mises con parámetros  $\mu=0$  y  $\kappa = 2$ .

- Los códigos en  $R$  de las Figuras 2.1, 2.2, 2.3 pueden ser consultados en el A.1.1.



# Capítulo 3

## Modelos de Regresión

Un modelo de regresión es empleado cuando se desea estudiar alguna posible relación (dependencia) de una variable con respecto a otras, ya que este modelo llevará a una función que permitirá observar futuros valores, dicha función será estimada a partir de datos observados. En este capítulo se presentará el desarrollo de los modelos de regresión para datos circulares, los cuales son:

1. Regresión Circular-Lineal (regresión periódica).
2. Regresión Circular-Circular.

Además se presentarán sus respectivas medidas de correlación.

### 3.1. Regresión Circular-Lineal

En un modelo de regresión ordinario (lineal) la variable dependiente y las variables independientes tienen que ser lineales, lo cual es distinto al modelo que trabajaremos, ya que la *variable independiente* deberá ser de tipo circular ya sea un ángulo o el tiempo de ocurrencia de algún fenómeno de tipo periódico, y la *variable dependiente* deberá ser de tipo lineal.

Denotaremos a las variables de la siguiente manera:

- $y$  es la variable dependiente (lineal).
- $\alpha$  es la variable independiente (circular).

Se supone que la variable independiente, es circular y tiene un periodo conocido como  $360^\circ$ , un año, 24hrs etcétera; de esta manera no se tendrá que estimar este valor.

Un modelo básico que se ajusta bien a fenómenos de tipo periódico está dado por la siguiente ecuación:

$$y = A_0 + A_1 \cos(w(\alpha - \alpha_0)). \quad (3.1)$$

Esta ecuación está compuesta por los siguientes elementos:

- $\alpha$  = variable independiente con periodo  $T$ .
- $\alpha_0$  = acrofase.
- $A_1$  = amplitud.
- $A_0$  = nivel medio o mesor.
- $w$  = frecuencia angular.

En la Figura 3.1 se pueden identificar cada uno de los elementos mencionados anteriormente

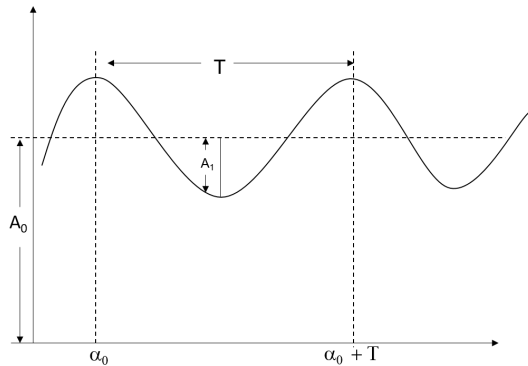


Figura 3.1: Gráfica de la función coseno con nivel medio  $A_0$ , amplitud  $A_1$

La frecuencia angular  $w$  se encuentra relacionada con el periodo  $T$  de la siguiente manera:

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

esto dependerá de si las mediciones están en grados o radianes.

Podemos reescribir la ecuación (3.1) de la siguiente manera:

$$g(\alpha) = A_0 + A_1 \cos(w\alpha - \phi) \quad (3.2)$$

con

$$\phi = w\alpha_0.$$

Cuando el conjunto de datos  $(\alpha_1, y_1), (\alpha_2, y_2), \dots, (\alpha_n, y_n)$  están dados es decir, fueron observados, es necesario agregar a la ecuación (3.1) un término de error que se denotará como  $\epsilon_i$ . Este término es agregado ya que en el momento en que estos datos fueron observados puede existir un error en su medición, debido a esto los  $\epsilon_i$  se consideran variables aleatorias; en algunos casos se puede suponer que son independientes y siguen una distribución normal; añadiendo este término el modelo quedará de la siguiente manera:

$$y_i = A_0 + A_1 \cos(w(\alpha_i - \alpha_0)) + \epsilon_i.$$

La generalización del modelo (3.1) está dada por la siguiente ecuación:

$$\begin{aligned} g(\alpha) = A_0 &+ A_1 \cos(w\alpha - \phi_1) \\ &+ A_2 \cos(2w\alpha - \phi_2) + \dots + A_k \cos(kw\alpha - \phi_k), \end{aligned} \quad (3.3)$$

conocido como **“Polinomio Trigonométrico”** o suma parcial de una serie de Fourier, notemos que esta ecuación contiene frecuencias angulares  $w, 2w, \dots, kw$ , esto es, múltiplos de “ $w$ ”.

A los términos  $A_1 \cos(wt - \phi)$ ,  $A_2 \cos(2wt - \phi_2)$ , etcétera. Se les llama primer armónico, segundo armónico, así sucesivamente.

Para los datos  $(\alpha_1, y_1), (\alpha_2, y_2), \dots, (\alpha_n, y_n)$  se podrá encontrar un polinomio trigonométrico que ajuste a esto con la condición de que los valores de la variable independiente ( $\alpha_i$ ) sean diferentes; por lo tanto se deberá encontrar un polinomio que cumpla con las siguientes tres condiciones:

1. El modelo deberá ajustar a los datos con suficiente exactitud; no obstante el precio a pagar por este ajuste es un gran número de parámetros, los cuales se tendrán que estimar y esto puede llegar a ser una tarea difícil.
2. El modelo debe ser tan suave como sea posible, es decir la gráfica no debe mostrar cambios frecuentes o abruptos a menos que los datos sugieran dicho comportamiento.
3. El número de parámetros deberá de ser lo más pequeño posible, esto debido a que la estimación de alguno de ellos puede resultar difícil.

Raramente las tres condiciones se cumplen de manera simultánea; ya que mientras para el primer criterio se menciona un buen ajuste sin considerar la limitación del número de términos, el tercer criterio sugiere lo contrario así como el segundo.

Así que de manera frecuente se restringe la ecuación (3.3) al primer armónico  $k = 1$  de esta manera se buscará satisfacer tanto el segundo como el tercer criterio buscando que el modelo sea parsimonioso, de esta manera el polinomio quedará reducido a la siguiente ecuación:

$$g(\alpha) = A_0 + A_1 \cos(w\alpha - \phi).$$

### 3.1.1. Ajuste de un Polinomio Trigonométrico

El caso más simple nos lleva a la ecuación (3.1) y utilizando la siguiente identidad trigonométrica:

$$\cos(\alpha - \phi) = \cos(\alpha) \cos(\phi) + \text{sen}(\alpha) \text{sen}(\phi),$$

la ecuación (3.1) queda de la siguiente forma:

$$g(\alpha) = A_0 + A_1 \cos(w\alpha) \cos(\phi) + A_1 \text{sen}(w\alpha) \text{sen}(\phi),$$

definiendo:

$$X = A_1 \cos(\phi) \quad \text{y} \quad Z = A_1 \text{sen}(\phi)$$

la ecuación queda de la siguiente forma:

$$g(\alpha) = A_0 + X \cos(w\alpha) + Z \text{sen}(w\alpha), \quad (3.4)$$

se observa que tanto  $X$  como  $Z$  son parámetros ya que dependen de la amplitud ( $A_1$ ) y de el ángulo acrofase ( $\phi$ ).

La ecuación (3.4) cuenta entonces con 4 parámetros ( $A_0, X, Z, w$ ); pero  $w$  se supone conocido por lo que solo resulta de interés estimar  $A_0, X$  y  $Z$ ; se observa que la ecuación (3.4) es lineal, y el método que se utilizará para estimar los parámetros de interés es el de mínimos cuadrados.

Sean  $(\alpha_1, y_1), (\alpha_2, y_2), \dots, (\alpha_n, y_n)$  los puntos dados y  $g(\alpha_i)$  los valores ajustados resultantes de la ecuación (3.4), los residuales del modelo están definidos de la siguiente manera:

$$\delta_i = g(\alpha_i) - y_i, \quad (3.5)$$

los parámetros  $A_0, X, Z$  podrán ser determinados minimizando la suma de cuadrados de residuales (desviaciones):

$$S = \sum_{i=1}^n (\delta_i)^2 = \sum_{i=1}^n [g(\alpha_i) - y_i]^2,$$

se sabe que:

$$g(\alpha_i) = A_0 + X \cos(w\alpha_i) + Z \text{sen}(w\alpha_i)$$

y como  $w$  es conocido se pueden calcular los siguientes valores:

$$c_i = \cos(w\alpha_i) \quad s_i = \text{sen}(w\alpha_i), \quad (3.6)$$

con esto, la ecuación queda de la siguiente manera:

$$g(\alpha_i) = A_0 + Xc_i + Zs_i,$$

sustituyendo  $g(\alpha_i)$  en la ecuación (3.5) se tiene

$$\delta_i = A_0 + Xc_i + Zs_i - y_i,$$

de donde se plantea el problema de minimizar la suma de cuadrados de residuales siguiente:

$$S = \sum_{i=1}^n (\delta_i)^2 = \sum_{i=1}^n [A_0 + Xc_i + Zs_i - y_i]^2. \quad (3.7)$$

Ahora aplicando algunas herramientas de cálculo diferencial encontraremos la estimación de los parámetros ( $A_0$ ,  $X$ ,  $Z$ ) que minimizan a  $S$  en la ecuación (3.7).

Primero se calcularán las derivadas parciales de la ecuación (3.7) con respecto a los tres parámetros:

- Derivada parcial con respecto a  $X$ :

$$\begin{aligned} \frac{\partial S}{\partial X} &= \frac{\partial \sum [A_0 + Xc_i + Zs_i - y_i]^2}{\partial X} \\ &= 2 \sum [A_0 + Xc_i + Zs_i - y_i]c_i \\ &= 2 \left[ \sum A_0c_i + \sum Xc_i^2 + \sum Zs_i c_i - \sum y_i c_i \right] \\ &= 2 \left[ A_0 \sum c_i + X \sum c_i^2 + Z \sum s_i c_i - \sum y_i c_i \right]. \end{aligned}$$

- Derivada parcial con respecto a  $Z$ :

$$\begin{aligned} \frac{\partial S}{\partial Z} &= \frac{\partial \sum [A_0 + Xc_i + Zs_i - y_i]^2}{\partial Z} \\ &= 2 \sum [A_0 + Xc_i + Zs_i - y_i]s_i \\ &= 2 \left[ \sum A_0s_i + \sum Xc_i s_i + \sum Zs_i^2 - \sum y_i s_i \right] \\ &= 2 \left[ A_0 \sum s_i + X \sum c_i s_i + Z \sum s_i^2 - \sum y_i s_i \right]. \end{aligned}$$

- Derivada parcial con respecto a  $A_0$ .

$$\begin{aligned} \frac{\partial S}{\partial A_0} &= \frac{\partial \sum [A_0 + Xc_i + Zs_i - y_i]^2}{\partial A_0} \\ &= 2 \sum [A_0 + Xc_i + Zs_i - y_i] \\ &= 2 \left[ nA_0 + \sum Xc_i + \sum Zs_i - \sum y_i \right] \\ &= 2 \left[ nA_0 + X \sum c_i + Z \sum s_i - \sum y_i \right]. \end{aligned}$$

Donde cada suma va de 1 a  $n$ . Igualando a cero cada una de las derivadas

$$\frac{\partial S}{\partial X} = 0, \quad \frac{\partial S}{\partial Z} = 0, \quad \frac{\partial S}{\partial A_0} = 0,$$

se tiene el siguiente sistema de ecuaciones:

$$\left. \begin{aligned} 2[A_0 \sum c_i + X \sum c_i^2 + Z \sum s_i c_i - \sum y_i s_i] &= 0 \\ 2[A_0 \sum s_i + X \sum c_i s_i + Z \sum s_i^2 - \sum y_i s_i] &= 0 \\ 2[nA_0 + X \sum c_i + Z \sum s_i - \sum y_i] &= 0 \end{aligned} \right\}. \quad (3.8)$$

Simplificando el sistema de ecuaciones (3.8) encontrando los estimadores de los parámetros:

$$\left. \begin{aligned} A_0 \sum c_i + X \sum c_i^2 + Z \sum s_i c_i &= \sum y_i c_i \\ A_0 \sum s_i + X \sum c_i s_i + Z \sum s_i^2 &= \sum y_i s_i \\ nA_0 + X \sum c_i + Z \sum s_i &= \sum y_i \end{aligned} \right\}. \quad (3.9)$$

Las ecuaciones que forman el sistema de ecuaciones (3.9) son conocidas como “**ecuaciones normales**”; se observa que estas ecuaciones son lineales, este sistema de ecuaciones se resolverá para obtener los estimadores de  $(A_0, X, Z)$ ; para facilitar la solución de este sistema se define lo siguiente:

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

de esta manera el sistema de ecuaciones (3.9) queda expresado de la siguiente manera:

$$\left. \begin{aligned} aA_0 + bX + cZ &= d \\ eA_0 + cX + fZ &= g \\ hA_0 + aX + eZ &= k \end{aligned} \right\}. \quad (3.10)$$

A continuación se encontrará la solución a este sistema, el método que se utilizará para encontrar la solución es el de eliminación.

- Tomaremos del sistema de ecuaciones (3.10) la primera y segunda ecuación; la primera ecuación es multiplicada por  $(f)$  y la segunda es multiplicada por  $(-c)$  obteniendo así el siguiente sistema de dos ecuaciones:

$$\left. \begin{aligned} afA_0 + bfX + cfZ &= df \\ -ecA_0 - c^2X - fcZ &= -gc \end{aligned} \right\}, \quad (3.11)$$

de este sistema es fácil ver que al aplicar el método de eliminación se obtiene la siguiente ecuación:

$$A_0[af - ec] + X[bf - c^2] = df - gc. \quad (3.12)$$

- Ahora tomaremos del sistema de ecuaciones (3.10), la primera y la tercera ecuación; la primera se multiplica por  $(e)$  y la segunda se multiplica por  $(-c)$  obteniendo así el siguiente sistema de ecuaciones:

$$\left. \begin{aligned} aeA_0 + beX + ceZ &= de \\ -hcA_0 - acX - ecZ &= -kc \end{aligned} \right\}, \quad (3.13)$$

de este sistema es fácil ver que al aplicar el método de eliminación se obtiene la siguiente ecuación:

$$A_0[ae - hc] + X[be - ac] = de - kc. \quad (3.14)$$

- Con las ecuaciones obtenidas (3.12) y (3.14), se forma un nuevo sistema de ecuaciones el cual queda de la siguiente forma:

$$\left. \begin{aligned} A_0[af - ec] + X[bf - c^2] &= df - gc \\ A_0[ae - hc] + X[be - ac] &= de - kc \end{aligned} \right\}. \quad (3.15)$$

A las ecuaciones de este sistema se les aplicará el método de eliminación; la primera ecuación se multiplica por  $[be - ac]$  y la segunda por  $[c^2 - bf]$ ; una vez aplicado el método de eliminación la ecuación resultante queda solo en términos de  $A_0$ ; a partir de ésta realizando los despejes correspondientes se obtiene la expresión del estimador del parámetro  $A_0$ , el cual queda de la siguiente forma:

$$A_0 = \frac{-dfa - gbe + gca + kbf - kc^2 + dec}{-a^2f - e^2b + hbf - hc^2 + 2aec}. \quad (3.16)$$

El procedimiento para obtener las expresiones de los estimadores de  $X$  y  $Z$  es el mismo que el aplicado para encontrar la expresión de  $A_0$  por lo cual los pasos se omitirán; los estimadores de los parámetros restantes son los siguientes:

$$X = \frac{gae - cgh - afk + hfd + eck - de^2}{-a^2f + 2aec + fhb - e^2b - hc^2}, \quad (3.17)$$

$$Z = \frac{kac - ekb + hgb - chd - a^2g + dea}{-a^2f + 2aec + fhb - e^2b - hc^2}. \quad (3.18)$$

Una vez encontrados los estimadores de los parámetros de interés en ocasiones puede ser útil o de importancia encontrar también los estimadores de  $\mathbf{A}_1$  y  $\phi$ , esto se lleva a cabo de la siguiente forma.

De las siguientes expresiones:

$$X = A_1 \cos(\phi), \quad Z = A_1 \sin(\phi),$$

utilizando la siguiente identidad trigonométrica:

$$x^2 + y^2 = r^2(\cos^2(\phi) + \operatorname{sen}^2(\phi)) = r^2$$

con

$$x = r\cos(\phi), \quad y = r\operatorname{sen}(\phi),$$

se obtiene el estimador de  $A_1$

$$A_1 = (X^2 + Z^2)^{\frac{1}{2}}, \quad (3.19)$$

Para el valor de  $\phi$  se tiene la siguiente identidad trigonométrica:

$$\tan(\phi) = \frac{\operatorname{sen}(\phi)}{\cos(\phi)}$$

haciendo uso de ella se obtiene la siguiente expresión para el valor de  $\phi$ :

$$\frac{Z}{X} = \frac{A_1 \operatorname{sen}(\phi)}{A_1 \cos(\phi)} = \frac{\operatorname{sen}(\phi)}{\cos(\phi)} = \tan(\phi),$$

y de esta expresión se deduce que:

$$\phi = \arctan(Z/X). \quad (3.20)$$

Para esta última expresión se tiene que la función  $\arctan$  toma valores entre  $-90^\circ$  y  $90^\circ$  por lo que es conveniente dar una expresión más explícita para encontrar el valor de  $\phi$  que es la siguiente:

$$\phi = \begin{cases} \arctan(Z/X), & \text{si } X > 0, \\ 180^\circ + \arctan(Z/X), & \text{si } X < 0. \end{cases}$$

### 3.1.2. Ejemplo

En 1977 Ehrhardt estudió la regulación del transporte de ácido 2aminoisobutírico (AIB) en el hígado de ratas, él encontró un ritmo circadiano, el cual está representado en la Figura 3.2 por ocho puntos. Ajustaremos una función cosenoidal a esta muestra de datos.

- Periodo (T) de 24 horas.
- Frecuencia angular ( $w$ ) de  $\frac{360^\circ}{24\text{horas}}$ .

Los valores  $\alpha_i$  son (6, 10, 14, 18, 22, 2) horas, usando las ecuaciones (3.6) se pueden calcular las cantidades  $c_i$  y  $s_i$ , y con los valores de  $y_i$  resulta la Tabla 3.1.



Tiempo	$w\alpha_i$	$c_i$	$s_i$	$y_i$
6 h	$90^\circ$	0	1.0	1.87
10 h	$150^\circ$	-0.866	0.5	1.50
14 h	$210^\circ$	-0.866	-0.5	1.64
18 h	$270^\circ$	0	-1.0	2.45
22 h	$330^\circ$	0.866	-0.5	2.14
2 h	$30^\circ$	0.866	0.5	2.15
Total	-	0	0	11.75

Tabla 3.1

Con los valores obtenidos en la Tabla 3.1 se calculan los siguientes valores:

$$\begin{aligned} \sum c_i &= 0, & \sum c_i^2 &= 3, & \sum s_i c_i &= 0, \\ \sum y_i c_i &= 0.996, & \sum s_i &= 0, & \sum s_i^2 &= 3, \\ \sum y_i s_i &= -0.645, & n &= 6, & \sum y_i &= 11.75, \end{aligned}$$

Sustituyendo estos valores en las ecuaciones normales de un polinomio con primer armónico se obtiene:

$$\begin{aligned} 3X &= 0.996, \\ 3Z &= -0.645, \\ 6A_0 &= 11.75, \end{aligned}$$

De las cuales se despeja el valor de los parámetros para poder encontrar el polinomio que ajusta a los datos, realizando los despejes correspondientes se obtiene:

$$X = 0.332, \quad Z = -0.215, \quad A_0 = 1.96,$$

de donde se obtienen los siguientes valores:

$$A_1 = 0.396, \quad \phi = 327.1^\circ,$$

y así con los valores de los parámetros se tiene la ecuación:

$$y = 1.96 + (0.396)(\cos(15\alpha - 327.1)). \quad (3.21)$$

De manera gráfica en la figura 3.2 se puede observar la representación de nuestros datos originales (línea con puntos) y el ajuste que se obtiene por medio de la ecuación (3.21).

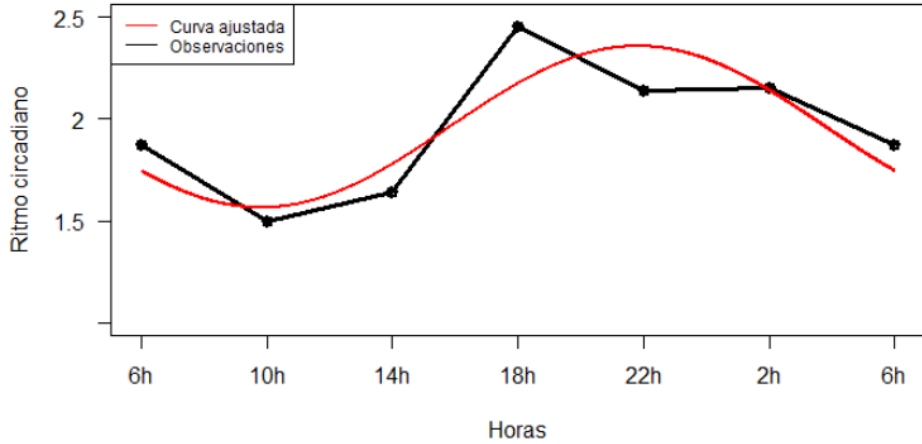


Figura 3.2: Ritmo circadiano a lo largo del día y ajuste de un polinomio trigonométrico

### 3.1.3. Polinomio con 1<sup>ero</sup> y 2<sup>do</sup> Armónico

En algunas ocasiones un modelo con primero y segundo armónico se ajustará de mejor manera a los datos y la ecuación del polinomio con estos dos armónicos será la siguiente:

$$g(\alpha) = A_0 + A_1 \cos(w\alpha - \phi_1) + A_2 \cos(2w\alpha - \phi_2), \quad (3.22)$$

definiendo:

$$X_1 = A_1 \cos(\phi_1), \quad Z_1 = A_1 \sin(\phi_1),$$

$$X_2 = A_2 \cos(\phi_2), \quad Z_2 = A_2 \sin(\phi_2),$$

la ecuación (3.22), se puede reescribir de la siguiente manera:

$$g(\alpha) = A_0 + X_1 \cos(w\alpha) + Z_1 \sin(w\alpha) + X_2 \cos(2w\alpha) + Z_2 \sin(2w\alpha). \quad (3.23)$$

se observa que  $X_1$ ,  $Z_1$ ,  $X_2$  y  $Z_2$  son parámetros y que dependen de la amplitud y del ángulo acrofase.

La ecuación (3.23) cuenta entonces con 6 parámetros ( $A_0$ ,  $X_1$ ,  $Z_1$ ,  $X_2$ ,  $Z_2$ ,  $w$ ); pero  $w$  se supone conocida por lo que solo resulta de interés estimar  $A_0$ ,  $X_1$ ,  $Z_1$ ,  $X_2$  y  $Z_2$ ; se observa que la ecuación (3.23) es lineal, y el método que se utilizará para estimar los parámetros de interés es el de mínimos cuadrados.

Sean  $(\alpha_1, y_1), (\alpha_2, y_2), \dots, (\alpha_n, y_n)$  los puntos dados y  $g(\alpha_i)$  los valores ajustados resultantes de la ecuación (3.4), los residuales del modelo están definidos de la siguiente manera:

$$\delta_i = g(\alpha_i) - y_i, \quad (3.24)$$

los parámetros  $A_0, X, Z$  podrán ser determinados minimizando la suma de cuadrados de residuales (desviaciones):

$$S = \sum_{i=1}^n (\delta_i)^2 = \sum_{i=1}^n [g(\alpha_i) - y_i]^2,$$

se sabe que:

$$g(\alpha_i) = A_0 + X_1 \cos(w\alpha_i) + Z_1 \sen(w\alpha_i) + X_2 \cos(2w\alpha_i) + Z_2 \sen(2w\alpha_i)$$

y como  $w$  es conocido se pueden calcular los siguientes valores:

$$c_i = \cos(w\alpha_i) \quad s_i = \sen(w\alpha_i) \quad f_i = \cos(2w\alpha_i) \quad r_i = \sen(2w\alpha_i), \quad (3.25)$$

con esto, la ecuación queda de la siguiente manera:

$$g(\alpha_i) = A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i,$$

sustituyendo  $g(\alpha_i)$  en la ecuación (3.24) se tiene

$$\delta_i = A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i - y_i,$$

de donde se plantea el problema de minimizar la suma de cuadrados de residuales siguiente:

$$S = \sum_{i=1}^n (\delta_i)^2 = \sum_{i=1}^n [A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i - y_i]^2. \quad (3.26)$$

Ahora aplicando algunas herramientas de cálculo diferencial encontraremos la estimación de lo parámetros  $(A_0, X_1, Z_1, X_2, Z_2)$  que minimizan a  $S$  en la ecuación (3.26).

Primero se calcularán las derivadas parciales de la ecuación (3.26) con respecto a los cinco parámetros:

- Derivada parcial con respecto a  $X_1$ :

$$\begin{aligned} \frac{\partial S}{\partial X_1} &= \frac{\partial \sum [A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i - y_i]^2}{\partial X_1} \\ &= 2 \left[ A_0 \sum c_i + X_1 \sum c_i^2 + Z_1 \sum s_i c_i + X_2 \sum f_i c_i + Z_2 \sum r_i c_i - \sum y_i c_i \right]. \end{aligned}$$

- Derivada parcial con respecto a  $Z_1$ :

$$\begin{aligned}\frac{\partial S}{\partial Z_1} &= \frac{\partial \sum [A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i - y_i]^2}{\partial Z_1} \\ &= 2 \left[ A_0 \sum s_i + X_1 \sum c_i s_i + Z_1 \sum s_i^2 + X_2 \sum f_i s_i + Z_2 \sum r_i s_i - \sum y_i s_i \right].\end{aligned}$$

- Derivada parcial con respecto a  $A_0$ .

$$\begin{aligned}\frac{\partial S}{\partial A_0} &= \frac{\partial \sum [A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i - y_i]^2}{\partial A_0} \\ &= 2 \left[ nA_0 + X_1 \sum c_i + Z_1 \sum s_i + X_2 \sum f_i + Z_2 \sum r_i - \sum y_i \right].\end{aligned}$$

- Derivada parcial con respecto a  $X_2$ :

$$\begin{aligned}\frac{\partial S}{\partial X_2} &= \frac{\partial \sum [A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i - y_i]^2}{\partial X_2} \\ &= 2 \left[ A_0 \sum f_i + X_1 \sum c_i f_i + Z_1 \sum s_i f_i + X_2 \sum f_i^2 + Z_2 \sum r_i f_i - \sum y_i f_i \right].\end{aligned}$$

- Derivada parcial con respecto a  $Z_2$ :

$$\begin{aligned}\frac{\partial S}{\partial Z_2} &= \frac{\partial \sum [A_0 + X_1 c_i + Z_1 s_i + X_2 f_i + Z_2 r_i - y_i]^2}{\partial Z_2} \\ &= 2 \left[ A_0 \sum r_i + X_1 \sum c_i r_i + Z_1 \sum s_i r_i + X_2 \sum f_i r_i + Z_2 \sum r_i^2 - \sum y_i r_i \right].\end{aligned}$$

Donde cada suma va de 1 a  $n$ . Igualando a cero cada una de las derivadas

$$\frac{\partial S}{\partial X_1} = 0, \quad \frac{\partial S}{\partial Z_1} = 0, \quad \frac{\partial S}{\partial A_0} = 0, \quad \frac{\partial S}{\partial X_2} = 0, \quad \frac{\partial S}{\partial Z_2} = 0,$$

Se forma un sistema de ecuaciones con las derivadas igualadas a cero, el cual queda de la siguiente manera:

$$\begin{array}{rcccccc} A_0 \sum c_i + & X_1 \sum c_i^2 + & Z_1 \sum s_i c_i + & X_2 \sum f_i c_i + & Z_2 \sum r_i c_i - & \sum y_i c_i & = 0 \\ A_0 \sum s_i + & X_1 \sum c_i s_i + & Z_1 \sum s_i^2 + & X_2 \sum f_i s_i + & Z_2 \sum r_i s_i - & \sum y_i s_i & = 0 \\ nA_0 + & X_1 \sum c_i + & Z_1 \sum s_i + & X_2 \sum f_i + & Z_2 \sum r_i - & \sum y_i & = 0 \\ A_0 \sum f_i + & X_1 \sum c_i f_i + & Z_1 \sum s_i f_i + & X_2 \sum f_i^2 + & Z_2 \sum r_i f_i - & \sum y_i f_i & = 0 \\ A_0 \sum r_i + & X_1 \sum c_i r_i + & Z_1 \sum s_i r_i + & X_2 \sum f_i r_i + & Z_2 \sum r_i^2 - & \sum y_i r_i & = 0 \end{array}$$

Las ecuaciones que forman el sistema de ecuaciones anterior son conocidas como “**ecuaciones normales**”; se observa que estas ecuaciones son lineales,

este sistema de ecuaciones se resolverá para obtener los estimadores de ( $A_0$ ,  $X_1$ ,  $Z_1$ ,  $X_2$  y  $Z_2$ ); para facilitar la solución de este sistema se define lo siguiente:

$$\begin{array}{llll}
 a = n & b = \sum c_i & c = \sum c_i & d = \sum f_i \\
 e = \sum r_i & f = \sum y_i & g = \sum c_i^2 & h = \sum s_i c_i \\
 i = \sum c_i f_i & j = \sum s_i r_i & k = \sum c_i y_i & l = \sum s_i^2 \\
 m = \sum f_i s_i & n = \sum r_i s_i & o = \sum s_i y_i & p = \sum f_i r_i \\
 q = \sum r_i^2 & r = \sum r_i y_i & s = \sum f_i^2 & t = \sum f_i y_i
 \end{array}$$

de esta manera el sistema de ecuaciones normales queda expresado de la siguiente manera:

$$\begin{array}{rcccccc}
 aA_0 & + & bX_1 & + & cZ_1 & + & dX_2 & + & eZ_2 & = & f \\
 bA_0 & + & gX_1 & + & hZ_1 & + & iX_2 & + & jZ_2 & = & k \\
 cA_0 & + & hX_1 & + & lZ_1 & + & mX_2 & + & nZ_2 & = & o \\
 eA_0 & + & jX_1 & + & nZ_1 & + & pX_2 & + & qZ_2 & = & r \\
 dA_0 & + & iX_1 & + & mZ_1 & + & sX_2 & + & pZ_2 & = & t
 \end{array}$$

La solución a este sistema, usando el método de Cramer es:

$$\widehat{A}_0 = \frac{\det \begin{bmatrix} f & b & c & d & e \\ k & g & h & i & j \\ o & h & l & m & n \\ r & j & n & p & q \\ t & i & m & s & p \end{bmatrix}}{\det \begin{bmatrix} a & b & c & d & e \\ b & g & h & i & j \\ c & h & l & m & n \\ e & j & n & p & q \\ d & i & m & s & p \end{bmatrix}}, \quad \widehat{X}_1 = \frac{\det \begin{bmatrix} a & f & c & d & e \\ b & k & h & i & j \\ c & o & l & m & n \\ e & r & n & p & q \\ d & t & m & s & p \end{bmatrix}}{\det \begin{bmatrix} a & b & c & d & e \\ b & g & h & i & j \\ c & h & l & m & n \\ e & j & n & p & q \\ d & i & m & s & p \end{bmatrix}},$$

$$\widehat{Z}_1 = \frac{\det \begin{bmatrix} a & b & f & d & e \\ b & g & k & i & j \\ c & h & o & m & n \\ e & j & r & p & q \\ d & i & t & s & p \end{bmatrix}}{\det \begin{bmatrix} a & b & c & d & e \\ b & g & h & i & j \\ c & h & l & m & n \\ e & j & n & p & q \\ d & i & m & s & p \end{bmatrix}}, \quad \widehat{X}_2 = \frac{\det \begin{bmatrix} a & b & c & f & e \\ b & g & h & k & j \\ c & h & l & o & n \\ e & j & n & r & q \\ d & i & m & t & p \end{bmatrix}}{\det \begin{bmatrix} a & b & c & d & e \\ b & g & h & i & j \\ c & h & l & m & n \\ e & j & n & p & q \\ d & i & m & s & p \end{bmatrix}},$$

$$\widehat{Z}_2 = \frac{\det \begin{bmatrix} a & b & c & d & f \\ b & g & h & i & k \\ c & h & l & m & o \\ e & j & n & p & r \\ d & i & m & s & t \end{bmatrix}}{\det \begin{bmatrix} a & b & c & d & e \\ b & g & h & i & j \\ c & h & l & m & n \\ e & j & n & p & q \\ d & i & m & s & p \end{bmatrix}}.$$

## 3.2. Correlación Circular-Lineal

En algunas ocasiones cuando una variable circular  $\phi$  está relacionada con una variable lineal  $y$ , por ejemplo la dirección del viento la cual es medida en ángulos está relacionada con variables como la humedad o cantidad de lluvia las cuales son variables lineales; debido a la existencia de este tipo de posibles relaciones es necesario contar con un valor que indique si estas variables presentan o no alguna posible relación, a este valor se le llama *coeficiente de correlación*.

El modelo de regresión es el siguiente:

$$y = A_0 + A_1 \cos(w(\alpha - \alpha_0)),$$

expresado de la siguiente forma:

$$y = A_0 + A_1 \cos(\phi - \phi_0), \quad (3.27)$$

con

$$\phi = w\alpha, \quad \phi_0 = w\alpha_0,$$

donde  $A_0$  es el mesor,  $A_1$  la amplitud y  $\phi_0$  el ángulo acrofase; los estimadores de estos parámetros ya se han obtenido, ahora el objetivo es saber si la variable  $y$  está o no correlacionada con  $\phi$ , es decir si  $y$  depende de  $\phi$ , para esto se definen las siguientes variables:

$$x = \cos(\phi - \phi_0), \quad x_i = \cos(\phi_i - \phi_0), \quad (3.28)$$

haciendo uso de estas variables podemos ver a la ecuación (3.2) como una ecuación lineal:

$$y = A_0 + A_1 x, \quad (3.29)$$

con pendiente  $A_1$ .

Debido a esto a las parejas  $(x_i, y_i)$  se les puede aplicar un método ordinario de correlación lineal, el coeficiente que utilizaremos será el **Coefficiente de Correlación de Pearson** dado por la siguiente ecuación:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}, \quad (3.30)$$

donde  $\bar{x}$  y  $\bar{y}$  son el promedio de dichos valores; todas las sumas con  $i = 1 \dots n$ .

### 3.3. Regresión Circular-Circular

Sean  $\alpha$  y  $\beta$  dos variables aleatorias con función de densidad conjunta  $f(\alpha, \beta)$ , con  $0 < \alpha, \beta \leq 2\pi$ ; para predecir el valor  $\beta$  para un  $\alpha$  dado consideremos la regresión o la esperanza condicional de un vector  $e^{i\beta}$  dado  $\alpha$  es decir:

$$E[e^{i\beta}|\alpha] = \rho(\alpha)e^{i\mu(\alpha)}, \quad (3.31)$$

de acuerdo con la identidad de Euler:

$$e^{ix} = \cos(x) + i\text{sen}(x),$$

la esperanza condicional puede ser escrita de la siguiente manera:

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

con

$$E[\cos(\beta)|\alpha] = g_1(\alpha),$$

$$E[\text{sen}(\beta)|\alpha] = g_2(\alpha),$$

de donde  $\hat{\beta}$  se determina como:

$$\hat{\beta} = \arctan\left(\frac{g_2(\alpha)}{g_1(\alpha)}\right). \quad (3.32)$$

A falta de más especificaciones acerca de la estructura de  $g_1(\alpha)$  y  $g_2(\alpha)$ , es en general difícil estimar a partir de los datos y en ocasiones aún con más especificaciones la determinación puede resultar difícil, por lo que  $g_1(\alpha)$  y  $g_2(\alpha)$ , serán aproximadas por funciones adecuadas, notemos que  $g_1(\alpha)$  y  $g_2(\alpha)$  son periódicas con periodo  $2\pi$ , por lo que pueden ser expresadas mediante una expansión en Serie de Fourier, como se muestra a continuación:

$$\begin{aligned} g_1(\alpha) &\approx \sum_{k=0}^m \left[ A_k \cos\left(\frac{2\pi k}{2\pi}\alpha\right) + B_k \text{sen}\left(\frac{2\pi k}{2\pi}\alpha\right) \right] \\ &\approx \sum_{k=0}^m [A_k \cos(k\alpha) + B_k \text{sen}(k\alpha)], \end{aligned}$$



$$\begin{aligned}
g_2(\alpha) &\approx \sum_{k=0}^m \left[ C_k \cos\left(\frac{2\pi k}{2\pi}\alpha\right) + D_k \operatorname{sen}\left(\frac{2\pi k}{2\pi}\alpha\right) \right] \\
&\approx \sum_{k=0}^m [C_k \cos(k\alpha) + D_k \operatorname{sen}(k\alpha)],
\end{aligned}$$

para una elección adecuada de  $m$ . A partir de esto se tiene el siguiente modelo

$$\cos(\beta) \approx \sum_{k=0}^m [A_k \cos(k\alpha) + B_k \operatorname{sen}(k\alpha)] + \epsilon_1$$

y

$$\operatorname{sen}(\beta) \approx \sum_{k=0}^m [C_k \cos(k\alpha) + D_k \operatorname{sen}(k\alpha)] + \epsilon_2$$

con  $(\epsilon_1, \epsilon_2)$  vector de errores con media cero y matriz de dispersión  $\mathbb{T}$  desconocida.

### 3.3.1. Estimación de los Coeficientes de Regresión

El problema de estimar los parámetros  $(A_k, B_k, C_k, D_k, \text{ para } k = 0, 1, \dots, m)$ , se resolverá por medio de mínimos cuadrados.

Sean  $(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)$  una muestra aleatoria de tamaño “ $n$ ”, las ecuaciones de las observaciones pueden ser escritas como:

$$Y_{1i} = \cos(\beta_i) \approx \sum_{k=0}^m [A_k \cos(k\alpha_i) + B_k \operatorname{sen}(k\alpha_i)] + \epsilon_{1i} \quad (3.33)$$

$$Y_{2i} = \operatorname{sen}(\beta_i) \approx \sum_{k=0}^m [C_k \cos(k\alpha_i) + D_k \operatorname{sen}(k\alpha_i)] + \epsilon_{2i} \quad (3.34)$$

para  $i = 1, \dots, n$ .

Se definen los siguientes vectores donde  $(X')$  denota al vector transpuesto de  $X$ ,

$$\begin{aligned}
Y^1 &= (Y_{11}, \dots, Y_{1n})', \\
Y^2 &= (Y_{21}, \dots, Y_{2n})', \\
\epsilon^1 &= (\epsilon_{11}, \dots, \epsilon_{1n})', \\
\epsilon^2 &= (\epsilon_{21}, \dots, \epsilon_{2n})', \\
\lambda^1 &= (A_0, A_1, \dots, A_m, B_1, \dots, B_m)',
\end{aligned}$$

$$\lambda^2 = (C_0, C_1, \dots, C_m, D_1, \dots, D_m)'$$

Se define la siguiente matriz:

$$X = \begin{bmatrix} 1 & \cos(\alpha_1) & \cdots & \cos(m\alpha_1) & \text{sen}(\alpha_1) & \cdots & \text{sen}(m\alpha_1) \\ 1 & \cos(\alpha_2) & \cdots & \cos(m\alpha_2) & \text{sen}(\alpha_2) & \cdots & \text{sen}(m\alpha_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos(\alpha_n) & \cdots & \cos(m\alpha_n) & \text{sen}(\alpha_n) & \cdots & \text{sen}(m\alpha_n) \end{bmatrix},$$

donde  $\text{sen}(0) = 0$  y  $\cos(0) = 1$  y la matriz  $X$  tiene los valores que acompañan a los parámetros a estimar, es por esto que la primera columna de la matriz está compuesta de valores 1.

Al haber definido los vectores  $(Y^1, Y^2, \epsilon^1, \epsilon^2, \lambda^1, \lambda^2)$  y la matriz  $(X)$ , las ecuaciones (3.33) y (3.34) pueden ser escritas de forma matricial quedando así:

$$Y^1 = X\lambda^1 + \epsilon^1 \quad (3.35)$$

$$Y^2 = X\lambda^2 + \epsilon^2 \quad (3.36)$$

se puede verificar la equivalencia en forma matricial realizando las operaciones correspondientes.

Poniendo las ecuaciones (3.35) y (3.36) en una sola expresión, obtenemos lo siguiente:

$$Y^* = X^*\lambda^* + \epsilon^*, \quad (3.37)$$

con

$$X^* = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix},$$

$$Y^{*'} = (Y^{1'} : Y^{2'}),$$

$$\lambda^{*'} = (\lambda^{1'} : \lambda^{2'}),$$

$$\epsilon^{*'} = (\epsilon^{1'} : \epsilon^{2'}).$$

De la ecuación (3.37) podemos calcular los mínimos cuadrados generalizados, sin embargo por la estructura de las ecuaciones (3.35), (3.36) y (3.37) y podemos utilizar la siguiente proposición:

**Proposición 3.1.** *Los estimadores de mínimos cuadrados para  $\beta$  en el modelo  $Y = X\beta + \epsilon$  son  $\hat{\beta} = (X'X)^{-1}X'Y$ .*

*Demostración.* Por medio del método de mínimos cuadrados se obtienen los valores  $\beta_1, \beta_2, \dots, \beta_k$  que minimizan la suma de cuadrados de los errores; en notación matricial la suma de los cuadrados de los errores:

$$\begin{aligned} R &= (Y - X\beta)'(Y - X\beta) \\ &= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta, \end{aligned}$$

ahora derivando  $R$  con respecto a  $\beta$  se tiene:

$$\frac{\partial R}{\partial \beta} = -2X'Y + 2X'X\beta$$

igualando a cero la derivada, suponiendo que  $X$  es de rango completo y despejando  $\beta$  se obtiene lo siguiente:

$$\hat{\beta} = [X'X]^{-1}X'Y.$$

□

Con la siguiente proposición se observa que obtener los mínimos cuadrados de las ecuaciones en un sólo vector es lo mismo que obtener los mínimos cuadrados de cada una de las ecuaciones.

**Proposición 3.2.** Sean  $\hat{\beta}_i = (X'_i X_i)^{-1} X'_i Y_i$  con  $i=1, 2, \dots, k$  los estimadores de cada  $\beta_i$  y sea  $\hat{\beta}_* = (X'_* X_*)^{-1} X'_* Y_*$  entonces:

$$\hat{\beta}_* = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}.$$

esto es, que los mínimos cuadrados aplicados a las ecuaciones combinadas es lo mismo que los aplicados a las ecuaciones por separado.

*Demostración.* se tiene que

$$\hat{\beta}_* = (X'_* X_*)^{-1} X'_* Y_*$$

lo cual es igual a:

$$\hat{\beta}_* = \begin{bmatrix} X'_1 X_1 & 0 & \dots & 0 & 0 \\ 0 & X'_2 X_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \dots & 0 & X'_k X_k \end{bmatrix}^{-1} \begin{bmatrix} X'_1 Y_1 \\ X'_2 Y_2 \\ \vdots \\ X'_k Y_k \end{bmatrix}$$

$$\hat{\beta}_* = \begin{bmatrix} (X_1'X_1)^{-1}X_1'Y_1 \\ (X_2'X_2)^{-1}X_2'Y_2 \\ \vdots \\ (X_k'X_k)^{-1}X_k'Y_k \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}.$$

□

Por lo que los estimadores de  $\lambda^1, \lambda^2$  son los siguientes:

$$\hat{\lambda}^1 = (X'X)^{-1}X'Y^1$$

$$\hat{\lambda}^2 = (X'X)^{-1}X'Y^2$$

### 3.3.2. Casos Especiales

1. Si para cada  $\alpha_i = \frac{2\pi(i-1)}{n}$ , para  $1 \leq i \leq n$ , se tienen los estimadores explícitos para  $1 \leq j \leq m$ :

$$\hat{A}_0 = \frac{1}{n} \sum \cos(\beta_i), \quad \hat{A}_j = \frac{1}{n} \sum \cos(\beta_i)\cos(j\alpha_i),$$

$$\hat{C}_0 = \frac{1}{n} \sum \sen(\beta_i), \quad \hat{C}_j = \frac{1}{n} \sum \sen(\beta_i)\cos(j\alpha_i),$$

$$\hat{B}_j = \frac{1}{n} \sum \cos(\beta_i)\sen(j\alpha_i), \quad \hat{D}_j = \frac{1}{n} \sum \sen(\beta_i)\sen(j\alpha_i).$$

con cada una de las sumas desde  $i = 1$  hasta  $n$ .

2. Si se tiene  $(\alpha, \beta_1), \dots, (\alpha, \beta_n)$ , es decir la muestra con cada  $\beta_i$  asociada a la misma  $\alpha$ , entonces los coeficientes de la ecuación de regresión serán cero, excepto  $A_0$  y  $C_0$  y estos estarán dados por:

$$\hat{A}_0 = \frac{1}{n} \sum_{j=1}^n \cos(\beta_j), \quad \hat{C}_0 = \frac{1}{n} \sum_{j=1}^n \sen(\beta_j).$$

### 3.3.3. Estimación de la matriz $\mathbb{T}$

La matriz de covarianzas ( $\mathbb{T}$ ) es estimada como sigue:

$$\begin{aligned} R_0(i, j) &= Y^{i'}Y^i - Y^{i'}X(X'X)^{-1}X'Y^j \\ &= Y^{i'}(I - M)Y^j \end{aligned}$$

donde:  $M = X(X'X)^{-1}X'$  y

$$R_0 = (R_0(i, j))_{2 \times 2}, \quad i = 1, 2, \quad j = 1, 2$$

entonces se tiene:

$$\widehat{\mathbb{T}} = \frac{R_0}{(n - 2(m + 1))},$$

que es un estimador insesgado de  $\mathbb{T}$ .

### 3.3.4. Determinación de $m$

Un problema general en el ajuste de cualquier polinomio de regresión es la determinación del grado, en el modelo de regresión Circular-Circular este valor  $m$  hace referencia al límite superior de la suma en las ecuaciones (3.33) y (3.34), esto se llevará a cabo utilizando el criterio de reducción en la suma de cuadrados del error al aumentar el valor de “ $m$ ”:

$$W = \begin{bmatrix} \cos((m + 1)\alpha_1) & \text{sen}((m + a)\alpha_1) \\ \vdots & \vdots \\ \cos((m + 1)\alpha_n) & \text{sen}((m + a)\alpha_n) \end{bmatrix}.$$

La matriz de diseño aumentada es escrita como sigue:

$$X_{(1)} = (X : W)P, \quad (3.38)$$

con  $P$  una matriz de permutación deseable que mantendrá las columnas de  $W$  en el lugar correcto.

Por lo que el modelo correspondiente es:

$$Y^{(1)} = X_{(1)}\lambda_{(1)}^{(1)} + \epsilon^{(1)}$$

$$Y^{(2)} = X_{(2)}\lambda_{(1)}^{(2)} + \epsilon^{(2)}$$

donde  $\lambda_{(1)}^{(i)}$  con  $i = 1, 2$  son los vectores de coeficientes que pueden ser estimados por el método de mínimos cuadrados; los cuales están dados por:

$$\lambda_{(1)}^{(i)} = [X'_{(1)}X_{(1)}]^{-1}X'_{(1)}Y^{(i)}, \quad i = 1, 2,$$

donde

$$X'_{(1)}X_{(1)} = P' \begin{bmatrix} X'X & X'W \\ W'X & W'W \end{bmatrix} P,$$

para invertir la matriz  $X'_{(1)}X_{(1)}$  se usa la siguiente preposición.

**Proposición 3.3.** *Sea  $G$  una matriz particionada,*

$$G = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

entonces  $G^{-1}$  queda de la siguiente forma:

$$G^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BF^{-1}CA^{-1} & -A^{-1}BF^{-1} \\ -F^{-1}CA^{-1} & F^{-1} \end{bmatrix}$$

con  $F = D - CA^{-1}B$ .

Se observa que la matriz  $X'_{(1)}X_{(1)}$  puede ser vista como una matriz particionada entonces se define lo siguiente:

$$A = X'X \quad B = X'W \quad C = W'X \quad D = W'W$$

haciendo uso del resultado anterior la matriz inversa de  $X'_{(1)}X_{(1)}$  tiene la siguiente forma

$$(X'_{(1)}X_{(1)})^{-1} = \begin{bmatrix} (X'X)^{-1} + (X'X)^{-1}(X'W)F(W'X)(X'X)^{-1} & -(X'X)^{-1}(X'W)F^{-1} \\ -F^{-1}(W'X)(X'X)^{-1} & F^{-1} \end{bmatrix},$$

con

$$\begin{aligned} F &= (W'W) - (W'X)(X'X)^{-1}(X'W) \\ &= W'(I - M)W \end{aligned}$$

donde  $M = X(X'X)^{-1}X'$ ; reescribiendo a la matriz  $(X'_{(1)}X_{(1)})^{-1}$  se obtiene:

$$\begin{aligned} (X'_{(1)}X_{(1)})^{-1} &= \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \\ &\quad + \begin{bmatrix} (X'X)^{-1}X'W \\ -I \end{bmatrix} F^{-1} \begin{bmatrix} (X'X)^{-1}X'W \\ -I \end{bmatrix}'; \end{aligned}$$

por lo que los estimadores por mínimos cuadrados quedan como sigue:

$$\begin{aligned} \hat{\lambda}_{(1)}^{(i)} &= P'(X'_{(1)}X_{(1)})^{-1}PP' \begin{bmatrix} X' \\ \dots \\ W' \end{bmatrix} Y^{(i)} \\ &= P' \begin{bmatrix} (X'X)^{-1}X'Y^{(i)} - (X'X)^{-1}X'N(I - M)Y^{(i)} \\ H^{-1}W'(I - M)Y^{(i)} \end{bmatrix}, \end{aligned}$$

con  $N = WH^{-1}W'$ , la suma de cuadrados del error es:

$$\begin{aligned} & Y^{(i)'}Y^{(i)} - Y^{(i)'}(X:W)P\hat{\lambda}_{(1)}^{(i)} \\ &= Y^{(i)'}(I - M)Y^{(i)} - Y^{(i)'}(I - M)N(I - M)Y^{(i)}. \end{aligned} \quad (3.39)$$

De la ecuación (3.39) se observa que la reducción en la suma de los cuadrados del error correspondiente a  $\hat{\lambda}^{(i)}$  por tomar términos adicionales es:

$$Y^{(i)'}(I - M)N(I - M)Y^{(i)} \geq 0. \quad (3.40)$$

Por lo tanto, antes de decidir incluir el término  $(m + 1)$ , primero se deberá calcular (3.40), y si ésta es lo suficientemente grande desde una perspectiva práctica, se podrá decidir incluir el término  $(m + 1)$ .

Un enfoque alternativo que es algebraicamente equivalente al anterior en la determinación del grado  $m$ , es evaluar el aumento en la estimación del parámetro de concentración de la distribución de  $\beta$  dado  $\alpha$ .

Una vez calculadas las aproximaciones de los polinomios trigonométricos de grado  $m$ -ésimo, es decir  $g_{im}(\alpha)$   $g_i(\alpha$  para  $i = 1, 2$  el parámetro de concentración está dado por

$$\rho_m(\alpha) = [g_{1m}^2(\alpha) + g_{2m}^2(\alpha)]^{1/2}. \quad (3.41)$$

Se tiene que  $0 \leq \rho_m(\alpha) \leq 1$  para cualquier  $m$  y  $\alpha$  dados, de esta manera a valores más cercanos a 1 se tiene un mejor ajuste.

Calculando  $\rho_m(\alpha_i)$  para cada  $\alpha_i$  con  $i = 1, \dots, n$  después de que se haya ajustado el polinomio trigonométrico de grado  $m$ -ésimo, se podrá decidir si se debe incluir un término adicional dependiendo de si hay un cambio significativo en la estimación del parámetro de concentración o no, es decir, si el valor de este parámetro aumenta significativamente, de esta manera  $\rho$  quedará dado por la siguiente ecuación:

$$\begin{aligned} \widehat{\rho_m^2} &= \frac{1}{n} \sum_{i=1}^n \rho_m^2(\alpha_i) = \frac{1}{n} \sum_{i=1}^n [g_{1m}^2(\alpha_i) + g_{2m}^2(\alpha_i)] \\ &= \frac{1}{n} \sum_{j=1}^2 Y^{(j)'}MY^{(j)} = \frac{1}{n} \sum_{j=1}^2 [Y^{(j)'}Y^{(j)} - R_0(j, j)]. \end{aligned}$$

### 3.4. *MAE* una Medida de Bondad de Ajuste

La medida de bondad de ajuste que será empleada para el modelo de regresión circular-circular y el modelo de regresión circular-lineal es *mean absolute error* (*MAE*) el cual está dado por la siguiente expresión:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n},$$

donde  $y_i$  es la observación  $i$ -ésima y  $\hat{y}_i$  es el valor  $i$ -ésimo estimado por el modelo. Se busca que el valor de esta medida de bondad de ajuste sea lo más cercana a cero y al realizar una comparación en modelos de regresión que busquen ajustar a un conjunto de datos el que tenga el menor valor para esta medida será el que mejor ajuste ofrece.



# Capítulo 4

## Aplicación de los Modelos

Las aplicaciones de los modelos presentados en este trabajo fueron implementados en el software *R* y el código de éstas se presentan en el anexo A.

### 4.1. Regresión Circular-Lineal

#### 4.1.1. Primera Aplicación

Para llevar a cabo la primera aplicación del modelo de regresión circular-lineal, se hará uso de la base de datos *AirPassengers* la cual se encuentra como un *dataset* precargado en el software *R*, este *dataset* contiene datos mensuales de doce años (1949-1960) de pasajeros en miles, de líneas aéreas internacionales; para realizar la aplicación del modelo se tomarán los datos de los últimos cuatro años (1957-1960). Se tendrán los valores  $\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = 3, \dots, \alpha_{12} = 12$  de la variable independiente siendo alfa el número de meses en un año para cada uno de los modelos, el periodo  $T = 12$  meses por lo que la frecuencia angular  $w$  es:

$$w = 360^\circ / T = 360^\circ / 12 \text{ meses} = 30^\circ / \text{mes}$$

la cual será la misma para todos los modelos.

- Año 1957.

La Tabla 4.1 muestra los valores obtenidos que son necesarios para estimar los parámetros del modelo.

$\alpha_i$	$w\alpha_i$	$c_i$	$s_i$	$y_i$
1	30°	.87	.50	315
2	60°	.50	0.87	301
3	90°	0	1	356
4	120°	-0.50	0.87	348
5	150°	-0.87	0.5	355
6	180°	-1	0	422
7	210°	-0.87	-0.5	465
8	240°	-0.50	-0.87	467
9	270°	0	-1	404
10	300°	0.50	-0.87	347
11	330°	0.87	-0.50	305
12	360°	1	0	336
Total	-	0	0	4421

Tabla 4.1: Valores necesarios para llevar a cabo los cálculos de un polinomio de primer armónico

Con los valores de la Tabla 4.1 se obtiene las siguientes cantidades definidas en la sección (3.1.1)

$$\begin{aligned}
 \sum c_i &= 0, & \sum c_i^2 &= 6.030, & \sum s_i c_i &= 0, \\
 \sum y_i c_i &= -343.5, & \sum s_i &= 0, & \sum s_i^2 &= 6.028, \\
 \sum y_i s_i &= -241.550, & n &= 12, & \sum y_i &= 4421,
 \end{aligned}$$

Sustituyendo estos valores en las ecuaciones normales de un polinomio con primer armónico y realizando los despejes correspondientes se obtiene el valor de los estimadores  $A_0$ ,  $X$  y  $Z$ :

$$A_0 = 368.417, \quad X = -56.965, \quad Z = -40.071,$$

de donde se obtienen los valores para  $A_1$  y  $\phi$ :

$$A_1 = 69.64694, \quad \phi = 215.1238,$$

y así con los valores de los parámetros se tiene la siguiente ecuación para el monto mensual de pasajeros de la aerolínea en el año 1957:

$$\hat{y} = 368.417 + 69.64694 \cos(30\alpha - 25.1238) \quad (4.1)$$

de manera gráfica en la Figura 4.1 se puede observar la representación de los datos originales y el ajuste que se obtiene por medio de la ecuación encontrada (4.1).

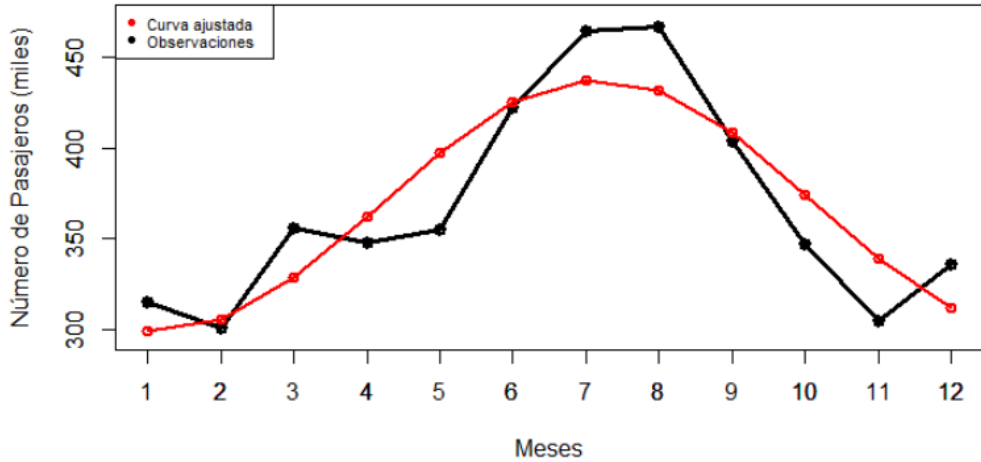


Figura 4.1: Ajuste de un polinomio trigonométrico con primer armónico año 1957

En la gráfica se puede observar un buen ajuste, sin embargo se desea un ajuste más preciso por lo que se llevará a cabo el ajuste de un modelo con 1<sup>ero</sup> y 2<sup>do</sup> armónico.

La tabla 4.2 muestra los valores obtenidos que son necesarios para estimar los parámetros del modelo.

Con los valores de la Tabla 4.2 se obtienen las siguientes cantidades:

$$\begin{array}{lll}
 n & = & 12, \\
 \sum f_i & = & 0, \\
 \sum c_i^2 & = & 6.03, \\
 \sum c_i r_i & = & 0, \\
 \sum f_i s_i & = & 0, \\
 \sum f_i r_i & = & 0, \\
 \sum f_i^2 & = & 6, \\
 \sum c_i & = & 0, \\
 \sum r_i & = & 0, \\
 \sum c_i s_i & = & 0, \\
 \sum c_i y_i & = & -343.5, \\
 \sum r_i s_i & = & 0, \\
 \sum r_i^2 & = & 6.06, \\
 \sum f_i y_i & = & -13.5, \\
 \sum s_i & = & 0, \\
 \sum y_i & = & 4421, \\
 \sum c_i f_i & = & 0, \\
 \sum s_i^2 & = & 6.03, \\
 \sum s_i y_i & = & -241.55, \\
 \sum r_i y_i & = & 167.91,
 \end{array}$$

Sustituyendo estos valores en las ecuaciones normales de un polinomio con primer y segundo armónico y realizando los despejes correspondientes se obtiene el valor de los estimadores  $A_0$ ,  $X_1$ ,  $Z_1$ ,  $X_2$  y  $Z_2$ :

$$A_0 = 368.416, \quad X_1 = -56.96, \quad Z_1 = -40.058, \quad X_2 = -2.25, \quad Z_2 = 27.707,$$

---

$\alpha_i$	$w\alpha_i$	$c_i$	$s_i$	$f_i$	$r_i$	$y_i$
1	30°	.87	.50	0.5	0.87	315
2	60°	.50	0.87	-0.50	0.87	301
3	90°	0	1	-1.0	0	356
4	120°	-0.50	0.87	-0.5	-0.87	348
5	150°	-0.87	0.5	0.5	-0.87	355
6	180°	-1	0	1	0	422
7	210°	-0.87	-0.5	0.5	0.87	465
8	240°	-0.50	-0.87	-0.5	0.87	467
9	270°	0	-1	-1	0	404
10	300°	0.50	-0.87	-0.5	-0.87	347
11	330°	0.87	-0.50	0.5	-0.87	305
12	360°	1	0	1	0	336
Total	-	0	0	4421		

---

Tabla 4.2: Valores necesarios para llevar a cabo los cálculos de un polinomio con primer y segundo armónico

y así con estos valores obtenidos se tiene la siguiente ecuación para el monto mensual de pasajeros de la aerolínea en el año 1957:

$$\hat{y} = 368.416 - 56.96 \cos(30\alpha) - 40.058 \sin(30\alpha) - 2.25 \cos(60\alpha) + 27.707 \sin(60\alpha), \quad (4.2)$$

de manera gráfica en la Figura 4.2 se puede observar la representación de los datos originales y el ajuste que se tiene por medio de la ecuación (4.2).

En la gráfica se puede observar que el ajuste a mejorado considerablemente; a continuación se calculará el MAE para obtener una medida de bondad de ajuste para cada uno de los modelos

- MAE Polinomio con primer armónico.

$$\text{MAE} = 21.794$$

- MAE Polinomio con primer y segundo armónico.

$$\text{MAE} = 12.58$$

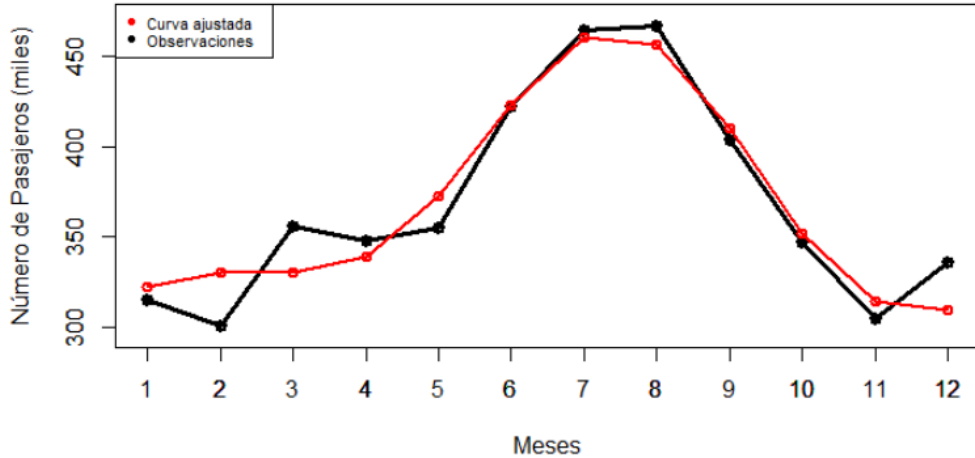


Figura 4.2: Ajuste de un polinomio trigonométrico con primer y segundo armónico año 1957

Recordemos que esta medida de bondad de ajuste se busca que sea lo más cercana a cero, por lo que realizando una comparación de los valores obtenidos nos damos cuenta que al realizar un ajuste con polinomio de segundo armónico el ajuste obtenido es bastante bueno y se ha logrado una mejora significativa con respecto al obtenido con el polinomio ajustado de primer armónico.

Se utilizará el coeficiente de correlación para poder tener este valor como referencia de la posible correlación que existe entre el número de pasajeros y el mes.

Para calcular el coeficiente se calcula la variable  $x_i = \cos(\phi_i - \phi_0)$  la cual se encuentra detallada en la sección 3.2, aplicando la fórmula con los mismos datos ( $y_i$ ) de la tabla se obtiene el siguiente valor para el coeficiente de correlación:

$$r = 0.8907$$

el cual resulta alto, por lo que se puede concluir que las variables *pasajeros mensuales* y *mes del año* se encuentran correlacionados.

Aplicando el mismo procedimiento, se obtendrán los modelos que se ajustan, el valor de *MAE* y los valores de los coeficientes de correlación para los años restantes.

■ Año 1958.

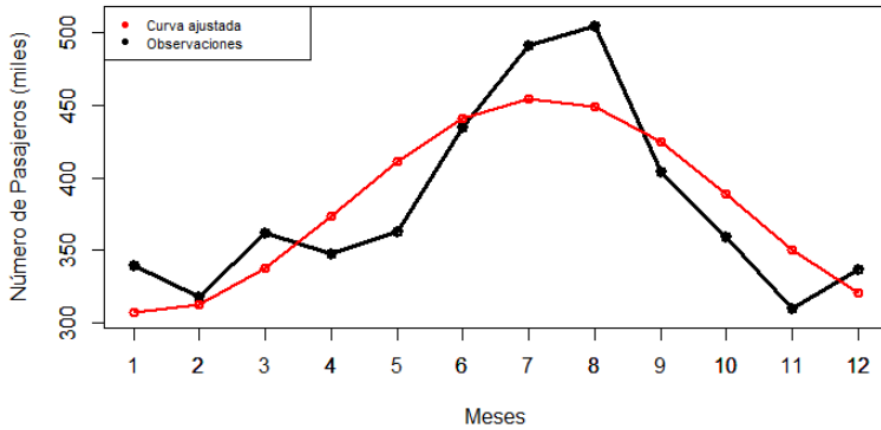


Figura 4.3: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> armónico del año 1958

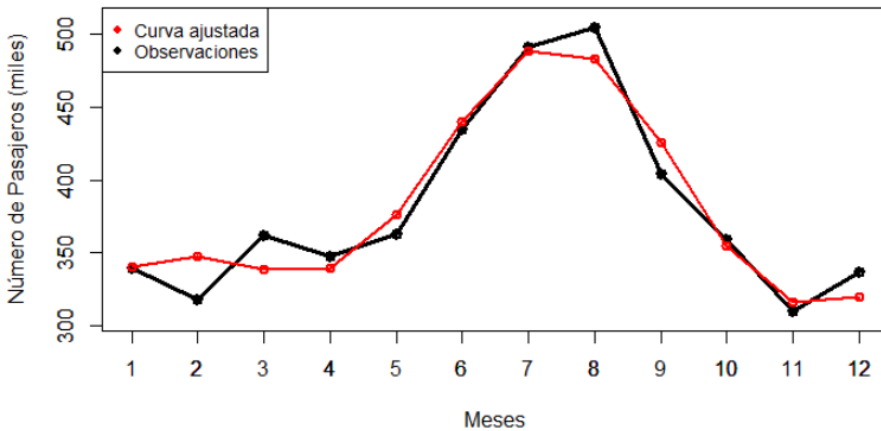


Figura 4.4: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico del año 1958

Ecuación de un polinomio trigonométrico con 1<sup>er</sup> armónico y  $MAE$  para el monto mensual de pasajeros de la aerolínea en el año 1958 representada en la Figura 4.3

$$\hat{y} = 381 + 74.4371\cos(30\alpha - 215.9239) \quad MAE = 28.492$$

Ecuación de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico y *MAE* para el monto mensual de pasajeros de la aerolínea en el año 1958 representada en la Figura 4.4

$$\hat{y} = 381 - 60.28 \cos(30\alpha) - 43.66 \sin(30\alpha) \\ - 1.17 \cos(60\alpha) + 39.34 \sin(60\alpha),$$

$$MAE = 12.798$$

Coefficiente de correlación para las variables *pasajeros mensuales* y *mes del año*

$$r = 0.854$$

■ Año 1959.

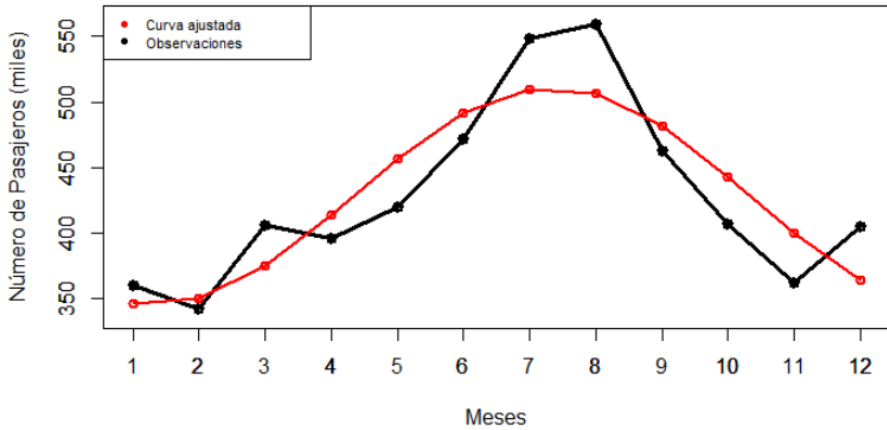


Figura 4.5: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico del año 1959

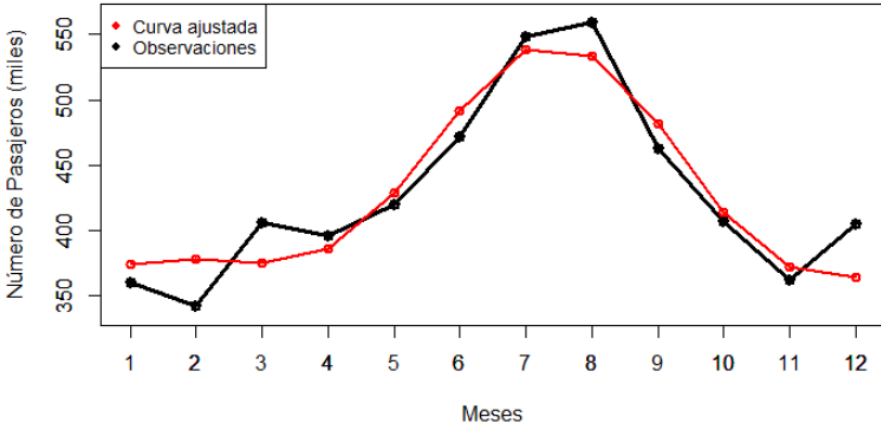


Figura 4.6: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> armónico del año 1959

Ecuación de un polinomio trigonométrico con 1<sup>er</sup> armónico y  $MAE$  para el monto mensual de pasajeros de la aerolínea en el año 1959 representada en la Figura 4.5

$$\hat{y} = 428.33 + 82.9467\cos(30\alpha - 219.8449) \quad MAE = 29.215$$



Ecuación de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico y *MAE* para el monto mensual de pasajeros de la aerolínea en el año 1959 representada en la Figura 4.6

$$\hat{y} = 428.33 - 63.684 \cos(30\alpha) - 53.127 \sin(30\alpha) \\ + 0.167 \cos(60\alpha) + 32.158 \sin(60\alpha),$$

$$MAE = 19.32$$

Coefficiente de correlación para las variables *pasajeros mensuales* y *mes del año*

$$r = 0.8792$$

■ Año 1960.

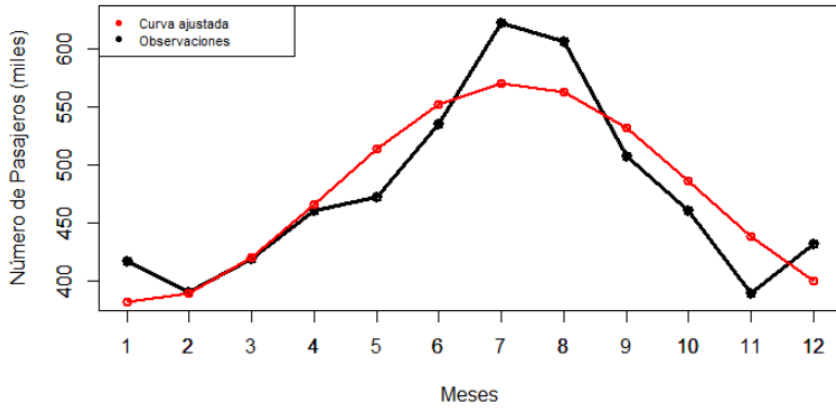


Figura 4.7: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> armónico del año 1960

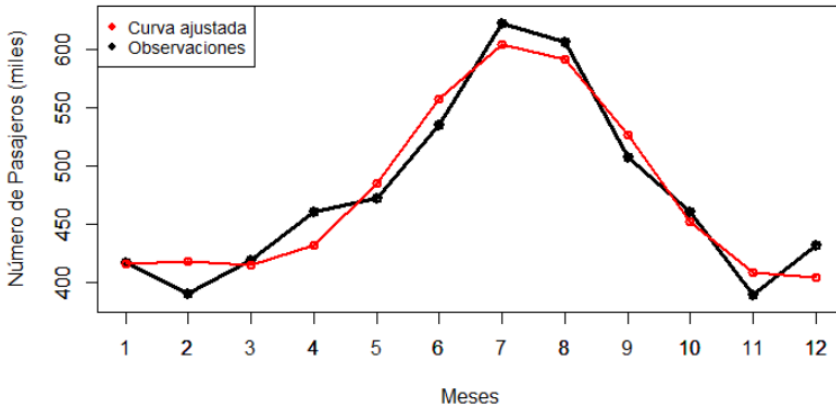


Figura 4.8: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico del año 1960

Ecuación de un polinomio trigonométrico con 1<sup>er</sup> armónico y *MAE* para el monto mensual de pasajeros de la aerolínea en el año 1960 representada en la Figura 4.7

$$\hat{y} = 476.167 + 94.657\cos(30\alpha - 216.2691) \quad MAE = 27.24$$

Ecuación de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico y *MAE* para el monto mensual de pasajeros de la aerolínea en el año 1960 representada en la Figura 4.8

$$\hat{y} = 476.16 - 76.31 \cos(30\alpha) - 55.978 \sin(30\alpha) \\ + 5.16 \cos(60\alpha) + 36.178 \sin(60\alpha),$$

$$MAE = 16.95$$

Coefficiente de correlación para las variables *pasajeros mensuales* y *mes del año*

$$r = 0.9014$$

### 4.1.2. Segunda Aplicación

Para llevar a cabo la segunda aplicación del modelo de regresión circular-lineal se hará uso de la base de datos de temperatura que fue obtenida de la siguiente página web:

- <https://datosclima.es/index.htm>

La base de datos es proporcionada por la **Agencia Estatal de Meteorología**<sup>1</sup> (AEMET), de esta base se tomaron los valores diarios de la temperatura máxima y mínima del 01-01-2016 al 31-12-2016, la cual está medida en grados centígrados ( $^{\circ}$  C) de la estación de Madrid; con estos valores se realizó un promedio mensual obteniendo los valores presentados en la Tabla 4.3.

Mes	Temperatura Máxima	Temperatura Mínima
Enero	9.51 $^{\circ}$ C	3.19 $^{\circ}$ C
Febrero	13.21 $^{\circ}$ C	3.67 $^{\circ}$ C
Marzo	14.11 $^{\circ}$ C	5.8 $^{\circ}$ C
Abril	22.17 $^{\circ}$ C	11.27 $^{\circ}$ C
Mayo	24.70 $^{\circ}$ C	13.55 $^{\circ}$ C
Junio	29.11 $^{\circ}$ C	16.44 $^{\circ}$ C
Julio	31.82 $^{\circ}$ C	18.31 $^{\circ}$ C
Agosto	32.50 $^{\circ}$ C	19.23 $^{\circ}$ C
Septiembre	28.85 $^{\circ}$ C	16.55 $^{\circ}$ C
Octubre	22.32 $^{\circ}$ C	12.03 $^{\circ}$ C
Noviembre	14.06 $^{\circ}$ C	8.24 $^{\circ}$ C
Diciembre	10.77 $^{\circ}$ C	3.08 $^{\circ}$ C

Tabla 4.3: Temperaturas máximas y mínimas mensuales

Se tendrán los valores  $\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = 3, \dots, \alpha_{12} = 12$  de la variable independiente siendo  $\alpha$  el número de meses en un año para los dos modelos, el periodo  $T = 12$  meses por lo que la frecuencia angular ( $w$ ) es:

$$w = 360^{\circ}/T = 360^{\circ}/12 \text{ meses} = 30^{\circ}/\text{mes},$$

la cual será la misma para los dos modelos.

<sup>1</sup>Es una agencia estatal de España, cuyo objetivo básico es la prestación de servicios meteorológicos, que sean competencia del Estado. Fue creada el 8 de febrero de 2008, sustituyendo al antiguo Instituto Nacional de Meteorología.

### ■ Temperatura Máxima

La Tabla 4.4 muestra los valores obtenidos que son necesarios para estimar los parámetros del modelo.

$\alpha_i$	$w\alpha_i$	$c_i$	$s_i$	$y_i$
1	30°	.87	.50	9.51
2	60°	.50	0.87	13.21
3	90°	0	1	14.11
4	120°	-0.50	0.87	22.17
5	150°	-0.87	0.5	24.70
6	180°	-1	0	29.11
7	210°	-0.87	-0.5	31.82
8	240°	-0.50	-0.87	32.50
9	270°	0	-1	28.85
10	300°	0.50	-0.87	22.32
11	330°	0.87	-0.50	14.06
12	360°	1	0	10.77
Total	-	0	0	253.13

Tabla 4.4: Valores necesarios para llevar a cabo los cálculos de un polinomio de primer armónico

Con los valores de la Tabla 4.4 se obtienen las siguientes cantidades definidas en la sección (3.1.1)

$$\begin{array}{rcl}
 \sum c_i & = & 0, \\
 \sum y_i c_i & = & -56.576, \\
 \sum y_i s_i & = & -37.488,
 \end{array}
 \quad
 \begin{array}{rcl}
 \sum c_i^2 & = & 6, \\
 \sum s_i & = & 0, \\
 n & = & 12,
 \end{array}
 \quad
 \begin{array}{rcl}
 \sum s_i c_i & = & 0, \\
 \sum s_i^2 & = & 6, \\
 \sum y_i & = & 253.13,
 \end{array}$$

Sustituyendo estos valores en las ecuaciones normales de un polinomio con primer armónico y realizando los despejes correspondientes se obtiene el valor de los estimadores de  $A_0$ ,  $X$  y  $Z$ :

$$\widehat{A}_0 = 21.094, \quad \widehat{X} = -9.382, \quad \widehat{Z} = -6.219,$$

de donde se obtienen los valores para  $A_1$  y  $\phi$ :

$$\widehat{A}_1 = 11.25602, \quad \widehat{\phi} = 213.539,$$

y así con los valores de los parámetros se tiene la siguiente ecuación para el promedio de la temperatura máxima mensual:

$$\hat{y} = 21.094 + 11.25602\cos(30\alpha - 213.539). \quad (4.3)$$

De manera gráfica en la Figura 4.9 se puede observar la representación de los datos originales y el ajuste que se obtiene por medio de la ecuación estimada (4.3).

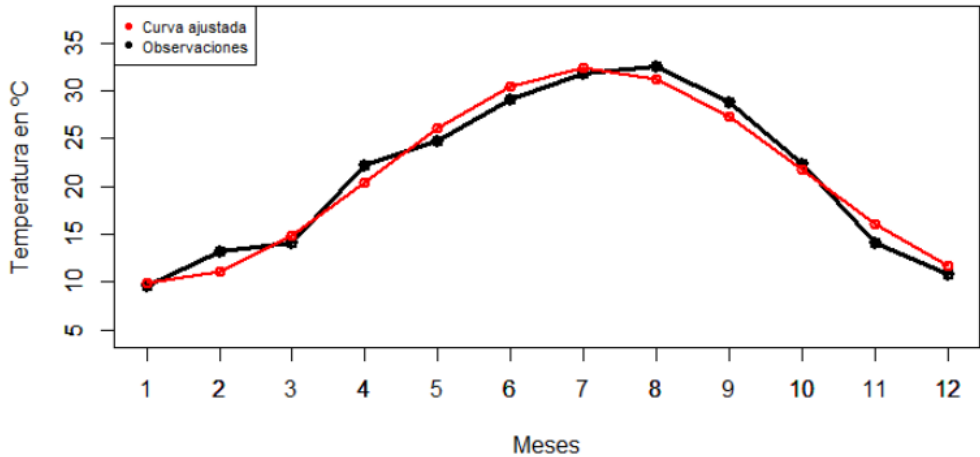


Figura 4.9: Datos de temperatura máxima y ajuste de un polinomio trigonométrico con primer armónico

De manera gráfica en la Figura 4.9 se puede observar un ajuste bastante bueno, sin embargo se desea un ajuste más preciso, por lo que se llevará a cabo el ajuste de un modelo con 1<sup>ero</sup> y 2<sup>do</sup>.

La tabla 4.5 muestra los valores obtenidos que son necesarios para estimar los parámetros del modelo.

Con los valores de la Tabla 4.5 se obtienen las siguientes cantidades:

$$\begin{array}{llll}
 n & = & 12, & \sum c_i & = & 0, & \sum s_i & = & 0, \\
 \sum f_i & = & 0, & \sum r_i & = & 0, & \sum y_i & = & 253.13, \\
 \sum c_i^2 & = & 6.03, & \sum c_i s_i & = & 0, & \sum c_i f_i & = & 0, \\
 \sum c_i r_i & = & 0, & \sum c_i y_i & = & -56.58, & \sum s_i^2 & = & 6.03, \\
 \sum f_i s_i & = & 0, & \sum r_i s_i & = & 0, & \sum s_i y_i & = & -37.49, \\
 \sum f_i r_i & = & 0, & \sum r_i^2 & = & 6.06, & \sum r_i y_i & = & 3.3, \\
 \sum f_i^2 & = & 6, & \sum f_i y_i & = & -8.14
 \end{array}$$

$\alpha_i$	$w\alpha_i$	$c_i$	$s_i$	$f_i$	$r_i$	$y_i$
1	30°	.87	.50	0.5	0.87	9.51
2	60°	.50	0.87	-0.50	0.87	13.21
3	90°	0	1	-1.0	0	14.11
4	120°	-0.50	0.87	-0.5	-0.87	22.17
5	150°	-0.87	0.5	0.5	-0.87	24.70
6	180°	-1	0	1	0	29.11
7	210°	-0.87	-0.5	0.5	0.87	31.82
8	240°	-0.50	-0.87	-0.5	0.87	32.50
9	270°	0	-1	-1	0	28.85
10	300°	0.50	-0.87	-0.5	-0.87	22.32
11	330°	0.87	-0.50	0.5	-0.87	14.06
12	360°	1	0	1	0	10.77
Total	-	0	0	0	0	253.13

Tabla 4.5: Valores necesarios para llevar a cabo los cálculos de un polinomio con primer y segundo armónico.

Sustituyendo estos valores en las ecuaciones normales de un polinomio con primer y segundo armónico y realizando los despejes correspondientes se obtiene el valor de los estimadores  $A_0$ ,  $X_1$ ,  $Z_1$ ,  $X_2$  y  $Z_2$ :

$$A_0 = 21.0941, \quad X_1 = -9.3830, \quad Z_1 = -6.2172, \quad X_2 = -1.35, \quad Z_2 = 0.5445,$$

y así con estos valores obtenidos se tiene la siguiente ecuación para el monto mensual de pasajeros de la aerolínea en el año 1957:

$$\hat{y} = 21.0941 - 9.3830 \cos(30\alpha) - 6.2172 \sin(30\alpha) - 1.35 \cos(60\alpha) + 0.5445 \sin(60\alpha), \quad (4.4)$$

de manera gráfica en la Figura 4.10 se puede observar la representación de los datos originales y el ajuste que se tiene por medio de la ecuación (4.4).

En la Figura 4.10 se puede observar qué el ajuste a mejorado considerablemente; a continuación se calculará el MAE para obtener una medida de bondad de ajuste para cada uno de los modelos.

- MAE Polinomio con primer armónico.

$$\text{MAE} = 1.2266$$

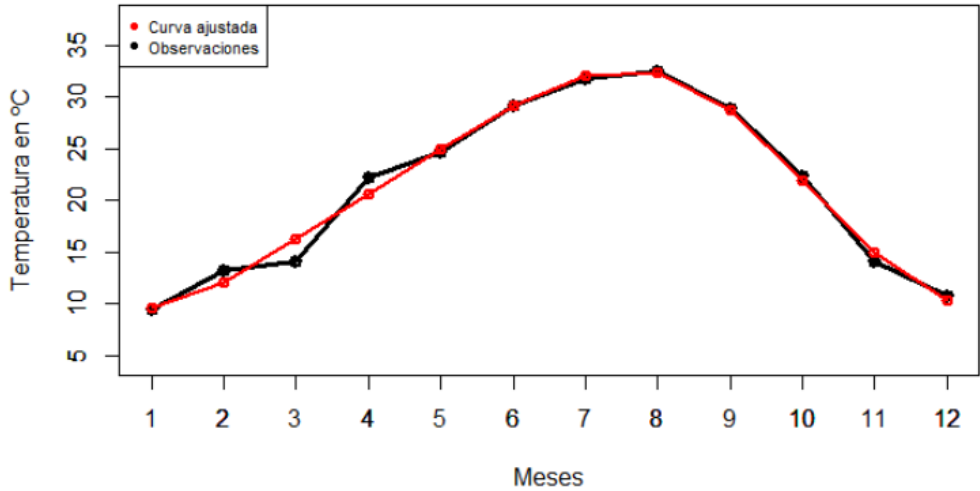


Figura 4.10: Ajuste de un polinomio trigonométrico con primer y segundo armónico temperatura máxima.

- MAE Polinomio con primer y segundo armónico.

$$\text{MAE} = 0.6979$$

Recordemos que esta medida de bondad de ajuste se busca que sea lo más cercana a cero, por lo que realizando una comparación de los valores obtenidos nos damos cuenta que al realizar un ajuste con polinomio de segundo armónico el ajuste obtenido es bastante bueno y se ha logrado una mejora significativa con respecto al obtenido con el polinomio ajustado de primer armónico.

Se utilizará el coeficiente de correlación para tener este valor como referencia de la correlación existente entre la temperatura y el mes en curso. Para calcular el coeficiente de correlación se calcula la variable  $x_i = \cos(\phi_i - \phi_0)$ , aplicando la fórmula del coeficiente de correlación con los datos observados  $y_i$  se obtiene el siguiente valor:

$$r = 0.9858,$$

el cual resulta bastante alto; por lo que se puede concluir que las variables *Temperatura máxima promedio* y *mes del año* se encuentran fuertemente correlacionadas.



Aplicando el mismo procedimiento, se obtendrá el modelo que se ajusta y el valor del coeficiente de correlación para la temperatura mínima promedio.

■ Temperatura Mínima.

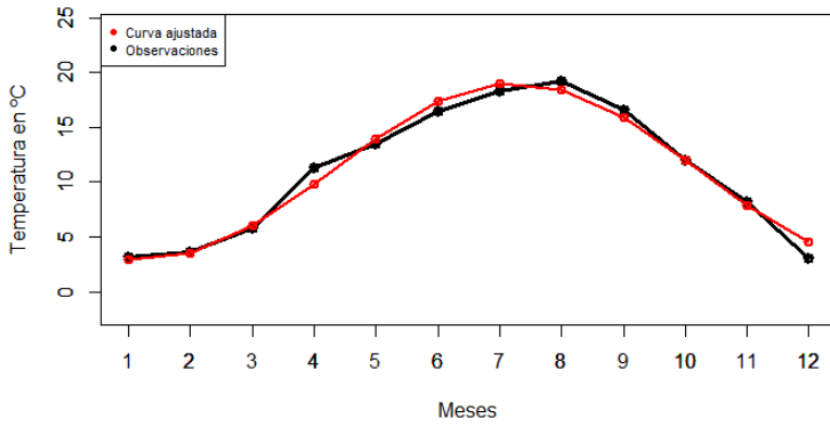


Figura 4.11: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> armónico temperatura máxima.

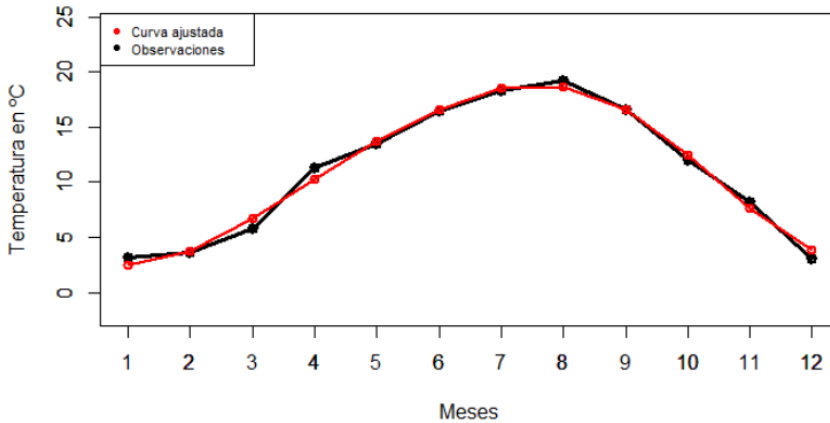


Figura 4.12: Ajuste de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico temperatura máxima.

Ecuación de un polinomio trigonométrico con 1<sup>er</sup> armónico y  $MAE$  para la temperatura mínima promedio mensual representada en la Figura 4.11

$$\hat{y} = 10.947 + 8.084207\cos(30\alpha - 217.7743) \quad MAE = 0.6132$$

Ecuación de un polinomio trigonométrico con 1<sup>er</sup> y 2<sup>do</sup> armónico y *MAE* para la temperatura mínima promedio mensual representada en la Figura 4.12

$$\hat{y} = 476.16 - 76.31 \cos(30\alpha) - 55.978 \sin(30\alpha) \\ + 5.16 \cos(60\alpha) + 36.178 \sin(60\alpha),$$

$$MAE = 0.4701$$

Coefficiente de correlación para las variables *pasajeros mensuales* y *mes del año*

$$r = 0.9913$$

## 4.2. Regresión Circular-Circular

### 4.2.1. Primera Aplicación

Para ilustrar el modelo de *regresión circular-circular* se hará uso de los datos presentados en la Tabla 4.6 los cuales están medidos en grados:

$\phi$	$\theta$	$\phi$	$\theta$
356	199	85	45
97	162	324	47
211	221	340	108
232	259	157	221
343	270	238	270
292	29	254	119
157	97	146	248
302	292	232	270
335	40	122	45
302	313	329	23
324	94		

Tabla 4.6: Parejas de direcciones del viento

Los datos son parejas de medidas de la dirección del viento a las 6:00 am ( $\phi$ ) y

12:00 pm ( $\theta$ ) de 21 días consecutivos, de una estación en Milwaukee.<sup>2</sup> Para obtener los valores estimados del modelo se hará uso de la paquetería *Circular* del lenguaje *R*; de esta paquetería se utilizará la función *lm.circular.cc*<sup>3</sup>, a continuación se muestran los valores obtenidos por medio de dicha función.

- Matriz cuyas columnas son los valores de los datos expresados en radianes:

	x	y		
[1,]	6.213372	2.0769418	[11,]	5.654867 1.6406095
[2,]	1.692969	2.8274334	[12,]	1.483530 0.7853982
[3,]	3.682645	3.8571776	[13,]	5.654867 0.8203047
[4,]	4.049164	4.5204028	[14,]	5.934119 1.8849556
[5,]	5.986479	4.7123890	[15,]	2.740167 3.8571776
[6,]	5.096361	0.5061455	[16,]	4.153884 4.7123890
[7,]	2.740167	1.6929694	[17,]	4.433136 2.0769418
[8,]	5.270894	5.0963614	[18,]	2.548181 4.3284165
[9,]	5.846853	0.6981317	[19,]	4.049164 4.7123890
[10,]	5.270894	5.4628806	[20,]	2.129302 0.7853982
			[21,]	5.742133 0.4014257

---

<sup>2</sup>Jammalamadaka, S. R. and Sengupta, A. (2001). *Topics in Circular Statistics*. World Scientific, Singapore.

<sup>3</sup>El código de esta función se encuentra en el Anexo A

- Matriz cuyas entradas son los coeficientes estimados del modelo. La primera columna corresponde a los coeficientes del modelo que predice el coseno de  $\beta$ , mientras que la segunda columna contiene las estimaciones para el modelo que predice el seno de  $\beta$ , el número de filas de la matriz obtenida dependerá del grado elegido del polinomio trigonométrico en este caso el grado elegido es 1:

```

$coefficients
              [,1]      [,2]
(Intercept) 0.01441183 0.1145101
cos.x        0.33465348 0.5494121
sin.x        -0.07478683 0.4821004

```

Con estos valores se obtienen las siguientes ecuaciones con los coeficientes estimados del modelo:

$$\widehat{\cos(\beta)} = 0.0144 + 0.3346(\cos(\alpha)) - 0.0747(\sin(\alpha)),$$

$$\widehat{\sin(\beta)} = 0.1145 + 0.5494(\cos(\alpha)) + 0.4821(\sin(\alpha)),$$

- Los valores ajustados de la variable  $\beta$ , los cuales son obtenidos aplicando la ecuación (3.33) son los siguientes:

	1	2	3	4	5	6	7
1.0587932	1.7597507	4.3432914	4.4960167	0.9506717	5.7385487	3.7025659	
	8	9	10	11	12	13	14
6.2706841	0.8635808	6.2706841	0.6971781	1.6188771	0.6971781	0.9204126	
	15	16	17	18	19	20	21
3.7025659	4.5440431	4.7005150	3.3716169	4.4960167	2.3434161	0.7815045	

- Valor del parámetro  $\rho$

```

$rho
      [,1]
[1,] 0.502289

```

Con este valor de  $\rho$  se sugiere que el ajuste obtenido es moderadamente bueno.

De manera gráfica en la Figura 4.13 se puede observar la representación de los datos originales (puntos negros) y los datos ajustados (puntos rojos), en la cual podemos observar un ajuste regular de los datos.

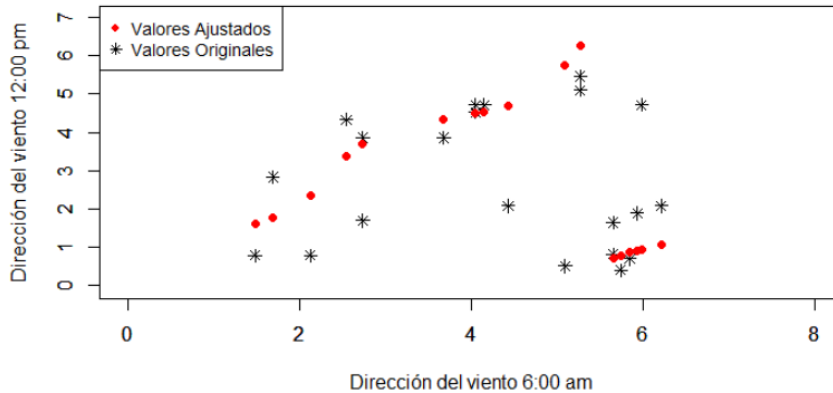


Figura 4.13: Valores ajustados de un modelo de regresión con  $m = 1$

Buscando un mejor ajuste se llevó a cabo el proceso anterior de manera iterativa modificando el valor para el grado del polinomio ( $m$ ) para observar si el valor de  $\rho$  aumentaba de manera significativa. El valor de  $\rho$  no aumentó significativamente a partir de  $m = 7$  por lo que para obtener un mejor ajuste el valor final de  $m$  fue de 6. De manera explícita se puede observar en la Tabla 4.7 los valores obtenidos para  $\rho$  con estos aumentos iterativos.

- Valor de  $\rho$  con ( $m = 6$ )

$$\rho = 0.89493$$

este valor de  $\rho$  indica que se obtiene un buen ajuste ya que está muy cercano a 1; de manera gráfica en la Figura 4.14 es posible observar que el ajuste se ha mejorado considerablemente

$m$	$\rho$
1	0.5022
2	0.5688
3	0.6137
4	0.6658
5	0.7928
6	0.89493
7	0.9019

Tabla 4.7: Valor de  $\rho$  para distintos valores de  $m$

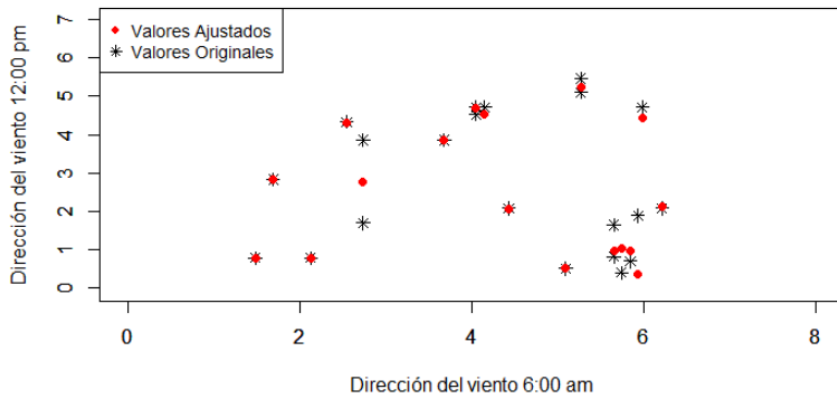


Figura 4.14: Valores ajustados de un modelo de regresión con  $m = 6$

Se calculará el MAE para obtener una medida de bondad de ajuste para cada uno de los modelos

- MAE modelo de regresión con valor  $m = 1$ .

$$\text{MAE} = 1.1748$$

- MAE modelo de regresión con valor  $m = 6$ .

$$\text{MAE} = 0.3124$$

Esta medida de bondad de ajuste se busca que sea lo más cercana a cero, por lo que realizando una comparación de los valores obtenidos nos damos cuenta que

al realizar un ajuste con un modelo de regresión con  $m = 6$  el ajuste que se obtiene es bastante bueno, pues es muy cercano a cero a diferencia del ajuste realizado con un modelo de regresión con  $m = 1$ .



## 4.2.2. Segunda Aplicación

Para la segunda ilustración del modelo de regresión circular-circular se realizará una simulación de valores para las variables  $\alpha$  y  $\beta$  por medio de una distribución *uniforme circular* y una distribución *von Mises* a continuación se muestran los valores obtenidos al aplicar el modelo de regresión circular-circular por medio de la función *lm.circular.cc* a los valores simulados.<sup>4</sup>

- Matriz cuyas columnas son los valores de los datos expresados en radianes:

	x	y
[1,]	0.2439525	1.1905100
[2,]	5.6883516	2.8625758
[3,]	1.1439644	0.9362814
[4,]	2.9520415	0.8720798
[5,]	0.9331646	6.1843456
[6,]	4.8656502	3.6670007
[7,]	3.7498669	1.7584181
[8,]	4.5142764	3.2984451

- Matriz cuyas entradas son los coeficientes estimados del modelo; el grado elegido del polinomio trigonométrico es 1:

	[,1]	[,2]
(Intercept)	0.1126461	0.4539954
cos.x	-0.2942487	-0.2628216
sin.x	1.0626641	0.3954773

con estos valores se obtienen las siguientes ecuaciones con los coeficientes estimados del modelo:

$$\widehat{\cos(\beta)} = 0.1126 - 0.29424(\cos(\alpha)) + 1.0626(\sin(\alpha)),$$

$$\widehat{\sin(\beta)} = 0.4539 - 0.2628(\cos(\alpha)) + 0.3954(\sin(\alpha)),$$

- Los valores ajustados de la variable  $\beta$ :

	1	2	3	4	5	6	7
	1.2935999	3.1213606	0.6344702	0.9176986	0.6609025	3.1181554	2.0892851
	8						
	3.0070048						

<sup>4</sup>El código de esta función se encuentra en el Anexo A

- Valor del parámetro  $\rho$ :

$$\begin{aligned} \rho &= [., 1] \\ &= [1., ] 0.8686024 \end{aligned}$$

con este valor de  $\rho$  se sugiere que el ajuste obtenido es bueno. De manera gráfica en la figura 4.15 se puede observar la representación de los datos originales y los datos ajustados, en esta se observa un buen ajuste.

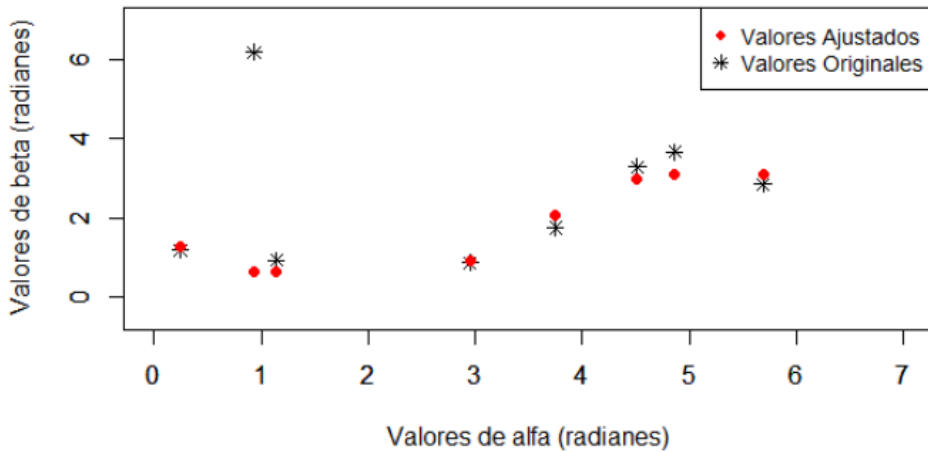


Figura 4.15: Valores ajustados de un modelo de regresión con  $m = 1$

Buscando un mejor ajuste se llevó a cabo el proceso anterior de manera iterativa modificando el valor para el grado del polinomio ( $m$ ) para así observar si el valor de  $\rho$  aumentaba de manera significativa y se pudiera obtener un valor muy cercano a 1; el valor de  $\rho$  no aumentó significativamente a partir de  $m=3$  por lo que para obtener un mejor ajuste el valor final de  $m$  fue de 2.

De manera explícita se puede observar en la Tabla 4.8 los valores obtenidos para  $\rho$  con estos aumentos iterativos.

- Valor de  $\rho$  con ( $m=2$ )

$$\rho = 0.941065$$

este valor de  $\rho$  indica que se obtiene un muy buen ajuste, porque el valor de  $\rho$  está muy cercano a 1; de manera gráfica en la Figura 4.16 es posible observar que el ajuste ha mejorado.

$m$	$\rho$
1	0.8686
2	0.9410
3	0.9740

Tabla 4.8: Valor de  $\rho$  para distintos valores de  $m$

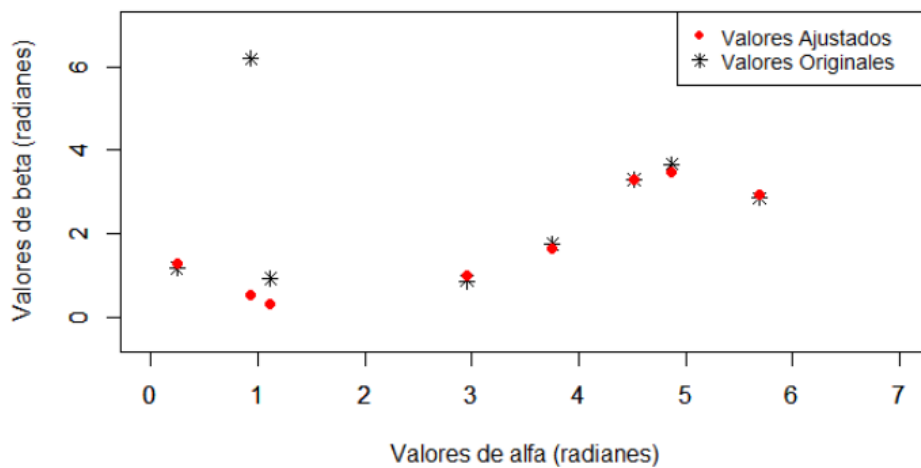


Figura 4.16: Valores ajustados de un modelo de regresión con  $m = 2$

Se calculará el MAE para obtener una medida de bondad de ajuste para cada uno de los modelos

- MAE modelo de regresión con valor  $m = 1$ .

$$\text{MAE} = 0.9232$$

- MAE modelo de regresión con valor  $m = 2$ .

$$\text{MAE} = 0.8578$$

Esta medida de bondad de ajuste se busca que sea lo más cercana a cero, por lo que realizando una comparación de los valores obtenidos nos damos cuenta que al realizar un ajuste con un modelo de regresión con  $m = 2$  el ajuste que se obtiene

es bastante bueno, pues es muy cercano a cero a diferencia del ajuste realizado con un modelo de regresión con  $m = 1$ .

Las gráficas 4.14, 4.13, 4.15 y 4.16, proporcionan la representación gráfica de los datos así como sus valores ajustados; recordamos que en este modelo los datos son del tipo circular por lo que los datos realmente viven en un toroide unitario, por lo que las gráficas fueron obtenidas realizando un primer corte vertical en el origen del toroide para formar un cilindro, posteriormente se cortó ese cilindro horizontalmente en el origen y al aplanar esta superficie se produce la representación plana de los datos; en este caso las gráficas están dadas en radianes.

# Conclusiones

En la presente tesis se trabajaron los modelos de regresión circular-circular y circular-lineal los cuales fueron desarrollados incluyendo dos aplicaciones para cada uno de ellos con datos reales y simulados, por lo anterior se considera que este texto puede ser utilizado como un material de consulta complementaria para un primer acercamiento a los modelos de regresión para datos circulares proporcionando al lector un panorama general del uso y desarrollo de estos modelos.

Los modelos que se han obtenido se han llevado a cabo con el fin de ilustrar:

1. El proceso a realizar en cada uno de ellos si se desean implementar.
2. Algunas de las áreas en las que los modelos de regresión para datos circulares tienen cabida.

## Modelo de regresión circular-lineal

- En la primera aplicación se ajustaron cuatro modelos de regresión con su respectivo coeficiente de correlación para la posible correlación entre el mes del año y el número de pasajeros que viajan en ese mes en una determinada aerolínea. Los ajustes obtenidos fueron buenos, pero estos se mejoraron con un modelo de dos o más armónicos; los cuatro coeficientes de correlación calculados tuvieron un valor muy cercano a 1, indicando una fuerte correlación entre las variables estudiadas; por lo que realizar un mejor ajuste resulto necesario sin importar que el modelo resultará poco parsimonioso.
- En la segunda aplicación se ajustaron dos modelos de regresión uno para la temperatura máxima y el segundo para la temperatura mínima ambas de la ciudad de Madrid, cada uno con su coeficiente de correlación entre la temperatura y el mes en curso del año 2016. Los ajustes obtenidos fueron bastante buenos, por lo que estos modelos proporcionan una idea del comportamiento de la temperatura a lo largo del año; el mejorar el ajuste con un modelo de dos o más armónicos se llevo a cabo obteniendo un ajuste

muy bueno el cual puede ser comprobado al observar las gráficas que lo representan.

### **Modelo de regresión circular-circular**

- En la primera aplicación se ajustaron dos modelos a las parejas de medidas de la dirección del viento a dos horas distintas del día, el primer modelo obtenido proporciona un ajuste regular, por lo que se siguió el criterio de incrementar el grado del polinomio  $m$  para que  $\rho$  aumentara significativamente y así encontrar un mejor ajuste, esto se logró hasta que  $m$  tomó el valor de 6.
- En la segunda aplicación se realizó una simulación con ayuda de dos distribuciones circulares, se ajustaron dos modelos a los datos obtenidos por medio de la simulación, en esta ocasión el grado del polinomio del segundo modelo se incrementó en una unidad  $m = 2$  y esto resultó suficiente, ya que al elevar el grado del polinomio a  $m = 3$ ,  $\rho$  no presentó un aumento significativo.

Los valores que  $\rho$  tomó en los dos modelos que se obtienen como el mejor ajuste fueron cercanos a 1; si se buscara un mejor ajuste aumentando a  $m$  lo suficiente para que  $\rho$  fuese lo más cercano a 1 sin importar que éste no aumentara significativamente. Esto estaría limitado por la complejidad al realizar los cálculos, sin importar si se llevan acabo de forma manual o con ayuda de algún software; en los modelos presentados se buscó realizar esto y con la ayuda del software  $R$  en el primer modelo el software fue incapaz de llevar acabo los cálculos con el valor  $m = 8$  y en el segundo modelo con  $m = 3$  presentando aumentos no significativos en el valor de  $m$ ; por lo que se reitera el uso del criterio mencionado para asignar el valor de  $m$ .

# Apéndice A

## Códigos en R

### A.1. Gráficas

#### A.1.1. Figura 2.1, 2.2 y 2.3

```
library(CircStats)
library(circular)
#Graficas Von Mises
f1 <- function(x) dvonmises(x, mu=circular(0), kappa=2)
#Grafica en circulo
curve.circular(f1, join=TRUE, xlim=c(-.5, .5), axes = TRUE, ticks = TRUE,
              ylim=c(-1.1, 1.1), shrink=1.35, nosort=TRUE,
              main="", lwd=2, col="red", lty=1, tcl=0.08, tcl.text=.3)

legend("top", c("Densidad von Mises con \n mu=0 kappa=2"),
      cex=0.8, col=c("red"), lwd = 2, lty = c(1), border = "white", box.col="white")
#Gráfica Uniforme
f2<-function(x) dcardioid(x,mu=circular(2),rho=0)
#Gráfica en circulo
curve.circular(f2, join=TRUE, xlim=c(-1.2,1.2),
              ylim=c(-1.2, 1.2), main="", lwd=2, col="red"
              , tcl=0.08, tcl.text=.3, shrink=1.15, lty=1)

legend("top", c("Densidad Uniforme"),
      cex=0.8, col=c("red"), lwd = 2, lty = c(1), border = "white", box.col="white")
#gráfica Cardioide
f3<-function(x) dcardioid(x,mu=circular(5),rho=.4)
```

```
curve.circular(f3, join=TRUE, xlim=c(-1.2,1.2),
              ylim=c(-1.2, 1.2), main="",lwd=2,col="red"
              , tcl=0.08,tcl.text=.3, shrink=1.15,lty=1)

legend("top", c("Densidad Cardioide"),
      cex=0.8, col=c("red"),lwd = 2, lty = c(1),border = "white",box.col="white")
```

### A.1.2. Figura 3.2

```
x<-c(6, 10, 14, 18, 22, 26,30)
y<-c(1.87, 1.50, 1.64, 2.45, 2.14,2.15,1.87)
x0<-seq(6,30,by=0.01)
x0
f<-c(1,1.5,2,2.5)
z<-c(((1.96+(.396*cos((15*x0-327.1)*(pi/180))))))
plot(x,y,type="o",xlim=c(6,30), ylim=c(1,2.5),
     lwd=3,xlab="", ylab="",axes = FALSE)
par(new=TRUE)
plot(x0,z, type="l",xlim=c(6,30),
     ylim=c(1,2.5),lwd=2,col="black", xlab="",
     ylab="",axes = FALSE)
axis(1, at=x, lab=c("6h", "10h", "14h", "18h", "22h", "2h", "6h" )
     ,cex.axis=1)
axis(2,at=f,lab=c("", "1.5", "2", "2.5"),las=2)
box()
legend("topleft", c("Curva ajustada","Observaciones"),
      cex=0.75, col=c("red","black"), lty = c(1))
```

## A.2. Aplicaciones

### A.2.1. 1<sup>ra</sup> Aplicación Regresión C-L

```
##### Modelo de Regresión primer armónico #####
View(AirPassengers)
data("AirPassengers")
cont<-c(97,109,121,133)
ti<-c(1:12)
w<-360/12
wti<-ti*w
```



```

for(i in cont){
  pasajeros<-c(AirPassengers[c(i:(i+11))])
  parametro1<-max(pasajeros)
  parametro2<-min(pasajeros)
  ci<-round(cos(wti*(2*pi/360)),2)
  si<-round(sin(wti*(2*pi/360)),2)
  a<-round(sum(ci),2)
  b<-round(sum(ci^2),2)
  c<-round(sum(si*ci),2)
  d<-round(sum(ci*pasajeros),3)
  e<-round(sum(si),2)
  f<-round(sum(si^2),3)
  g<-round(sum(si*pasajeros),3)
  h<-round(length(ti),3)
  k<-round(sum(pasajeros),3)
  A_0<-round((((k*b*f)-(d*f*a)-(g*b*e)+(g*c*a)-
              (k*c^2)+(d*e*c))/((2*a*e*c)-(a^2*f)-
              (e*e*b)+(h*b*f)-(h*c*c))),3)
  X<- round((((g*a*e)-(c*g*h)-(a*f*k)+(h*f*d)+
              (e*c*k)-(d*e*e))/((2*a*e*c)-(a^2*f)+
              (f*h*b)-(e^2*b)-(h*c^2))),3)
  Z<- round((((k*a*c)-(e*k*b)+(h*g*b)-(c*h*d)-
              (a^2*g)+(d*e*a))/((2*a*e*c)-(a^2*f)+
              (f*h*b)-(e^2*b)-(h*c^2))),3)
  A_1<-(X^2+Z^2)^(1/2)
  phi<-180+atan(Z/X)*(180/pi)
  regresion<-A_0+(A_1*cos(((w*ti)-phi)*(pi/180)))
  plot(ti,pasajeros,xlim=c(1,12), ylim=c(parametro2-5,parametro1+5)
       ,lwd=3,xlab="", ylab="",type="o")
  par(new=TRUE)
  plot(ti,regresion,xlim=c(1,12), ylim=c(parametro2-5,parametro1+5)
       ,lwd=2,xlab="Meses", ylab="Número de Pasajeros (miles)",
       type="l",col=c("red"))
  axis(1, at=ti, lab=c("1", "2", "3","4","5","6","7","8"
                      ,"9","10","11","12"),cex.axis=1)
  box()
  legend("topleft", c("Ecuación ajustada"), cex=0.7, col=c("red"),
        lty = c(1))
}
MAE<-sum(abs(pasajeros-regresion))/length(pasajeros)

```

```
##### Modelo de Regresión primer y segundo armónico #####
data("AirPassengers")
cont<-c(97,109,121,133)
ti<-c(1:12)
w<-360/12
wti<-ti*w
for(i in cont){
  pasajeros<-c(AirPassengers[c(i:(i+11))])
  parametro1<-max(pasajeros)
  parametro2<-min(pasajeros)
  ci<-round(cos(wti*(2*pi/360)),2)
  si<-round(sin(wti*(2*pi/360)),2)
  fi<-round(cos(2*wti*(2*pi/360)),2)
  cantidades<-matrix(c(ci,si,fi,ri),12,4)
  cantidades
  ri<-round(sin(2*wti*(2*pi/360)),2)
  a<-round(length(ti),2)
  b<-round(sum(ci),2)
  c<-round(sum(si),2)
  d<-round(sum(fi),2)
  e<-round(sum(ri),2)
  f<-round(sum(pasajeros),2)
  g<-round(sum(ci^2),2)
  h<-round(sum(si*ci),2)
  i<-round(sum(ci*fi),2)
  j<-round(sum(ci*ri),2)
  k<-round(sum(ci*pasajeros),2)
  l<-round(sum(si^2),2)
  m<-round(sum(fi*si),2)
  n<-round(sum(ri*si),2)
  o<-round(sum(si*pasajeros),2)
  p<-round(sum(fi*ri),2)
  q<-round(sum(ri^2),2)
  r<-round(sum(ri*pasajeros),2)
  s<-round(sum(fi^2),2)
  t<-round(sum(fi*pasajeros),2)
  la <- rbind(c(a,b,c,d,e),
             c(b,g,h,i,j),
             c(c,h,l,m,n),
             c(e,j,n,p,q),
             c(d,i,m,s,p))
}
```

```

lo <- c(f,k,o,r,t)
estimadores<-c(solve(la, lo))
estimadores
A_0<-estimadores[1]
X_1<-estimadores[2]
Z_1<-estimadores[3]
X_2<-estimadores[4]
Z_2<-estimadores[5]
#A_1<-(X^2+Z^2)^(1/2)
#phi<-180+atan(Z/X)*(180/pi)
regresion<-(A_0+X_1*cos((w*ti)*(pi/180))
            +Z_1*sin((w*ti)*(pi/180))+X_2*cos((2*w*ti)
            *(pi/180))+Z_2*sin((2*w*ti)*(pi/180)))
#regresion<-A_0+(A_1*cos(((w*ti)-phi)*(pi/180)))
plot(ti,pasajeros,xlim=c(1,12), ylim=c(parametro2-5,parametro1+5)
     ,lwd=3,xlab="", ylab="",type="o")
par(new=TRUE)
plot(ti,regresion,xlim=c(1,12), ylim=c(parametro2-5,parametro1+5)
     ,lwd=2,xlab="Meses", ylab="Número de Pasajeros (miles)",
     type="o",col=c("red"))
axis(1, at=ti, lab=c("1", "2", "3","4","5","6","7","8"
                    ,"9","10","11","12"),cex.axis=1)

box()
legend("topleft", c("Curva ajustada"
                    ,"Observaciones"), cex=0.85, col=c("red","black"), pch = c(19))
}

```

```
MAE2<-sum(abs(pasajeros-regresion))/length(pasajeros)
```

```

##### Coeficiente de Correlación #####
data("AirPassengers")
cont<-c(1,13,25,37,49,61,73,85,97,109,121,133)
ti<-c(1:12)
w<-360/12
wti<-ti*w
for(i in cont){
  pasajeros<-c(AirPassengers[c(i:(i+11))])
  ci<-cos(wti*(2*pi/360))
  si<-sin(wti*(2*pi/360))
  a<-sum(ci)
}

```

```

b<-sum(ci^2)
c<-sum(si*ci)
d<-sum(ci*pasajeros)
e<-sum(si)
f<-sum(si^2)
g<-sum(si*pasajeros)
h<-length(ti)
k<-sum(pasajeros)
A_0<-(((k*b*f)-(d*f*a)-(g*b*e)+(g*c*a)-(k*c^2)+(d*e*c))
/(2*a*e*c)-(a^2*f)-(e*e*b)+(h*b*f)-(h*c*c)))
X<-(((g*a*e)-(c*g*h)-(a*f*k)+(h*f*d)+(e*c*k)-(d*e*e))
/(2*a*e*c)-(a^2*f)+(f*h*b)-(e^2*b)-(h*c^2)))
Z<-(((k*a*c)-(e*k*b)+(h*g*b)-(c*h*d)
-(a^2*g)+(d*e*a))/(2*a*e*c)-
(a^2*f)+(f*h*b)-(e^2*b)-(h*c^2)))
A_1<-(X^2+Z^2)^(1/2)
phi<-180+atan(Z/X)*(180/pi)
xi<-c(cos(((w*ti)-phi)*(pi/180)))
xi_barra<-c(xi-mean(xi))
yi_barra<-c(pasajeros-mean(pasajeros))
parte_superior<-sum(xi_barra*yi_barra)
xi_barra2<-c(xi_barra^2)
yi_barra2<-c(yi_barra^2)
sxi_barra2<-sum(xi_barra2)
syi_barra2<-sum(yi_barra2)
parte_inf<-(((sxi_barra2)^(1/2))*((syi_barra2)^(1/2)))
coeficiente<-round(parte_superior/parte_inf,4)
print(coeficiente)
}

```

### A.2.2. 2<sup>da</sup> Aplicación Regresión C-L (Temperatura Máxima).

```

##### Modelo de Regresión primer armónico #####
temperaturasmax<-c(9.51,13.21,14.11,22.17,
24.70,29.11,31.82,32.50,28.85,22.32,
14.06,10.77)

ti<-c(1:12)
w<-360/12
wti<-ti*w

```

```

parametro1<-max(temperaturasmax)
parametro2<-min(temperaturasmax)
ci<-round(cos(wti*(2*pi/360)),2)
si<-round(sin(wti*(2*pi/360)),2)
a<-round(sum(ci),2)
b<-round(sum(ci^2),2)
c<-round(sum(si*ci),2)
d<-round(sum(ci*temperaturasmax),3)
e<-round(sum(si),2)
f<-round(sum(si^2),3)
g<-round(sum(si*temperaturasmax),3)
h<-round(length(ti),3)
k<-round(sum(temperaturasmax),3)
A_0<-round((((k*b*f)-(d*f*a)-(g*b*e)+(g*c*a)-
              (k*c^2)+(d*e*c))/((2*a*e*c)-(a^2*f)-
              (e*e*b)+(h*b*f)-(h*c*c))),3)
X<- round((((g*a*e)-(c*g*h)-(a*f*k)+(h*f*d)+
            (e*c*k)-(d*e*e))/((2*a*e*c)-(a^2*f)+
            (f*h*b)-(e^2*b)-(h*c^2))),3)
Z<- round((((k*a*c)-(e*k*b)+(h*g*b)-(c*h*d)-
            (a^2*g)+(d*e*a))/((2*a*e*c)-(a^2*f)+
            (f*h*b)-(e^2*b)-(h*c^2))),3)
A_1<-(X^2+Z^2)^(1/2)
phi<-180+atan(Z/X)*(180/pi)
regresion<-A_0+(A_1*cos(((w*ti)-phi)*(pi/180)))
plot(ti,temperaturasmax,xlim=c(1,12),
      ylim=c(parametro2-5,parametro1+5),lwd=3,xlab="",
      ylab="",type="o")
par(new=TRUE)
plot(ti,regresion,xlim=c(1,12),
      ylim=c(parametro2-5,parametro1+5),lwd=2,xlab="Meses",
      ylab="Temperatura en °C",type="l",col=c("red"))
axis(1, at=ti, lab=c("1", "2", "3","4","5",
                    "6","7","8","9","10","11","12")
      ,cex.axis=1)
box()
legend("topleft", c("Curva ajustada","Observaciones"),
      cex=0.7, col=c("red","black"), pch = c(19))
MAETM<-sum(abs(temperaturasmax-regresion))/length(temperaturasmax)

##### Modelo de Regresión primer y segundo armónico #####

```

```
temperaturamax<-c(9.51,13.21,14.11,22.17,
                  24.70,29.11,31.82,32.50,28.85,22.32,
                  14.06,10.77)

ti<-c(1:12)
w<-360/12
wti<-ti*w
parametro1<-max(temperaturamax)
parametro2<-min(temperaturamax)
ci<-round(cos(wti*(2*pi/360)),2)
si<-round(sin(wti*(2*pi/360)),2)
fi<-round(cos(2*wti*(2*pi/360)),2)
cantidades<-matrix(c(ci,si,fi,ri),12,4)
cantidades
ri<-round(sin(2*wti*(2*pi/360)),2)
a<-round(length(ti),2)
b<-round(sum(ci),2)
c<-round(sum(si),2)
d<-round(sum(fi),2)
e<-round(sum(ri),2)
f<-round(sum(temperaturamax),2)
g<-round(sum(ci^2),2)
h<-round(sum(si*ci),2)
i<-round(sum(ci*fi),2)
j<-round(sum(ci*ri),2)
k<-round(sum(ci*temperaturamax),2)
l<-round(sum(si^2),2)
m<-round(sum(fi*si),2)
n<-round(sum(ri*si),2)
o<-round(sum(si*temperaturamax),2)
p<-round(sum(fi*ri),2)
q<-round(sum(ri^2),2)
r<-round(sum(ri*temperaturamax),2)
s<-round(sum(fi^2),2)
t<-round(sum(fi*temperaturamax),2)
list(a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t)
la <- rbind(c(a,b,c,d,e),
            c(b,g,h,i,j),
            c(c,h,l,m,n),
            c(e,j,n,p,q),
            c(d,i,m,s,p))
lo <- c(f,k,o,r,t)
```

```

estimadores<-c(solve(la, lo))
estimadores
A_0<-estimadores[1]
X_1<-estimadores[2]
Z_1<-estimadores[3]
X_2<-estimadores[4]
Z_2<-estimadores[5]
regresion<-(A_0+X_1*cos((w*ti)*(pi/180))
            +Z_1*sin((w*ti)*(pi/180))+X_2*cos((2*w*ti)
            *(pi/180))+Z_2*sin((2*w*ti)*(pi/180))
plot(ti,temperaturasmax,xlim=c(1,12),
      ylim=c(parametro2-5,parametro1+5),lwd=3,xlab="",
      ylab="",type="o")
par(new=TRUE)
plot(ti,regresion,xlim=c(1,12),
      ylim=c(parametro2-5,parametro1+5),lwd=2,xlab="Meses",
      ylab="Temperatura en °C",type="o",col=c("red"))
axis(1, at=ti, lab=c("1", "2", "3","4","5",
                    "6","7","8","9","10","11","12")
      ,cex.axis=1)
box()
legend("topleft", c("Curva ajustada","Observaciones"),
      cex=0.7, col=c("red","black"), pch = c(19))

MAETM2<-sum(abs(temperaturasmax-regresion))/length(temperaturasmax)

##### Coeficiente de Correlación #####
xi<-c(cos(((w*ti)-phi)*(pi/180)))
xi_barra<-c(xi-mean(xi))
yi_barra<-c(temperaturasmax-mean(temperaturasmax))
parte_superior<-sum(xi_barra*yi_barra)
xi_barra2<-c(xi_barra^2)
yi_barra2<-c(yi_barra^2)
sxi_barra2<-sum(xi_barra2)
syi_barra2<-sum(yi_barra2)
parte_inf<-(((sxi_barra2)^(1/2))*((syi_barra2)^(1/2)))
coeficiente<-round(parte_superior/parte_inf,4)
print(coeficiente)

```

### A.2.3. 2<sup>da</sup> Aplicación Regresión C-L (Temperatura Mínima).

```
##### Modelo de Regresión #####
temperaturasmin<-c(3.19,3.67,5.80,11.27,13.55,
                  16.44,18.31,19.23,16.55,12.03,
                  8.24,3.08)

ti<-c(1:12)
w<-360/12
wti<-ti*w
parametro1<-max(temperaturasmin)
parametro2<-min(temperaturasmin)
ci<-round(cos(wti*(2*pi/360)),2)
si<-round(sin(wti*(2*pi/360)),2)
a<-round(sum(ci),2)
b<-round(sum(ci^2),2)
c<-round(sum(si*ci),2)
d<-round(sum(ci*temperaturasmin),3)
e<-round(sum(si),2)
f<-round(sum(si^2),3)
g<-round(sum(si*temperaturasmin),3)
h<-round(length(ti),3)
k<-round(sum(temperaturasmin),3)
A_0<-round((((k*b*f)-(d*f*a)-(g*b*e)+(g*c*a)-
              (k*c^2)+(d*e*c))/((2*a*e*c)-(a^2*f)-
              (e*e*b)+(h*b*f)-(h*c*c))),3)
X<- round((((g*a*e)-(c*g*h)-(a*f*k)+(h*f*d)+
            (e*c*k)-(d*e*e))/((2*a*e*c)-(a^2*f)+
            (f*h*b)-(e^2*b)-(h*c^2))),3)
Z<- round((((k*a*c)-(e*k*b)+(h*g*b)-(c*h*d)-
            (a^2*g)+(d*e*a))/((2*a*e*c)-(a^2*f)+
            (f*h*b)-(e^2*b)-(h*c^2))),3)
A_1<-(X^2+Z^2)^(1/2)
phi<-180+atan(Z/X)*(180/pi)
regresion<-A_0+(A_1*cos(((w*ti)-phi)*(pi/180)))
plot(ti,temperaturasmin,xlim=c(1,12),
      ylim=c(parametro2-5,parametro1+5),lwd=3,xlab="", ylab="",type="o")
par(new=TRUE)
plot(ti,regresion,xlim=c(1,12),
      ylim=c(parametro2-5,parametro1+5),lwd=2,xlab="Meses",
      ylab="Temperatura en °C",type="l",col=c("red"))
```



---

```

axis(1, at=ti, lab=c("1", "2", "3","4","5","6","7","8","9",
"10","11","12"),cex.axis=1)
box()
legend("topleft", c("Ecuación ajustada"),
cex=0.7, col=c("red"), lty = c(1))

##### Coeficiente de Correlación #####
xi<-c(cos(((w*ti)-phi)*(pi/180)))
xi_barra<-c(xi-mean(xi))
yi_barra<-c(temperaturasmin-mean(temperaturasmin))
parte_superior<-sum(xi_barra*yi_barra)
xi_barra2<-c(xi_barra^2)
yi_barra2<-c(yi_barra^2)
sxi_barra2<-sum(xi_barra2)
syi_barra2<-sum(yi_barra2)
parte_inf<-(((sxi_barra2)^(1/2))*((syi_barra2)^(1/2)))
coeficiente<-round(parte_superior/parte_inf,4)
print(coeficiente)

##### Modelo de Regresión primer y segundo armónico #####
temperaturasmin<-c(3.19,3.67,5.80,11.27,13.55,
16.44,18.31,19.23,16.55,12.03,
8.24,3.08)

ti<-c(1:12)
w<-360/12
wti<-ti*w
parametro1<-max(temperaturasmin)
parametro2<-min(temperaturasmin)
ci<-round(cos(wti*(2*pi/360)),2)
si<-round(sin(wti*(2*pi/360)),2)
fi<-round(cos(2*wti*(2*pi/360)),2)
ri<-round(sin(2*wti*(2*pi/360)),2)
a<-round(length(ti),2)
b<-round(sum(ci),2)
c<-round(sum(si),2)
d<-round(sum(fi),2)
e<-round(sum(ri),2)
f<-round(sum(temperaturasmin),2)
g<-round(sum(ci^2),2)
h<-round(sum(si*ci),2)
i<-round(sum(ci*fi),2)

```

---

```
j<-round(sum(ci*ri),2)
k<-round(sum(ci*temperaturasmin),2)
l<-round(sum(si^2),2)
m<-round(sum(fi*si),2)
n<-round(sum(ri*si),2)
o<-round(sum(si*temperaturasmin),2)
p<-round(sum(fi*ri),2)
q<-round(sum(ri^2),2)
r<-round(sum(ri*temperaturasmin),2)
s<-round(sum(fi^2),2)
t<-round(sum(fi*temperaturasmin),2)
la <- rbind(c(a,b,c,d,e),
            c(b,g,h,i,j),
            c(c,h,l,m,n),
            c(e,j,n,p,q),
            c(d,i,m,s,p))
lo <- c(f,k,o,r,t)
estimadores<-c(solve(la, lo))
estimadores
A_0<-estimadores[1]
X_1<-estimadores[2]
Z_1<-estimadores[3]
X_2<-estimadores[4]
Z_2<-estimadores[5]
regresion<-(A_0+X_1*cos((w*ti)*(pi/180))
            +Z_1*sin((w*ti)*(pi/180))+X_2*cos((2*w*ti)
            *(pi/180))+Z_2*sin((2*w*ti)*(pi/180))
plot(ti,temperaturasmin,xlim=c(1,12),
     ylim=c(parametro2-5,parametro1+5),lwd=3,xlab="",
     ylab="",type="o")
par(new=TRUE)
plot(ti,regresion,xlim=c(1,12),
     ylim=c(parametro2-5,parametro1+5),lwd=2,xlab="Meses",
     ylab="Temperatura en °C",type="o",col=c("red"))
axis(1, at=ti, lab=c("1", "2", "3","4","5",
                    "6", "7", "8", "9", "10", "11", "12"))
     ,cex.axis=1)
box()
legend("topleft", c("Curva ajustada","Observaciones"),
     cex=0.7, col=c("red","black"), pch = c(19))
```

---

```
MAETMi2<-sum(abs(temperaturasmin-regresion))/length(temperaturasmin)
```

#### A.2.4. 1<sup>ra</sup> Aplicación Regresión C-C

```
psigrad<-c(356,97,211,232,343,292,157,302,335,302,
           324,85,324,340,157,238,254,146,232,122,329)
thetagrado<-c(119,162,221,259,270,29,97,292,40,313,
              94,45,47,108,221,270,119,248,270,45,23)
cpsirad<-circular(psigrad*(pi/180))
cthetaad<-circular(thetagrado*(pi/180))
regresion<-lm.circular(type="c-c",y=cthetaad,x=cpsirad,order=1)
plot.default(cpsirad,cthetaad, ylim=c(-.05,7),xlim = c(0,8),
             xlab="", ylab="",lwd=1, pch=(16))
par(new=TRUE)
plot.default(cpsirad,regresion$fitted,ylim=c(-.05,7),xlim = c(0,8),
             col="red",xlab="Dirección del viento 6:00 am",
             ylab="Dirección del viento 12:00 pm",lwd=1, pch=(16))
legend("topleft", c("Valores Ajustados","Valores Originales")
      , cex=.9, col=c("red","black"),pch= c(16,8))
MAECC<-sum(abs(cthetaad-regresion$fitted))/length(cthetaad)
```

#### A.2.5. 2<sup>da</sup> Aplicación Regresión C-C

```
x <- circular(runif(8, 0, 2*pi))
y <- abs(atan2(0.15*cos(x) + 0.25*sin(x), 0.35*sin(x)) +
         rvonmises(n=8, mu=circular(0), kappa=5))
modregresion<-lm.circular(y, x, order=1)
modregresion
lambda<-c(modregresion$fitted)
plot.default(x, y,ylim = c(-.5,7),xlim =c(0,7),
             xlab="", ylab="",lwd=1, pch=(8))
par(new=TRUE)
plot.default(x,lambda,ylim = c(-.5,7),xlim=c(0,7)
             ,col="red",xlab="Valores de alfa(radianes)",
             ylab="Valores de beta(radianes)",lwd=1, pch=(16))
legend("topleft", c("Valores Ajustados","Valores Originales")
      , cex=.9, col=c("red","black"),pch= c(16,8))
MAECC2<-sum(abs(y-modregresion$fitted))/length(y)
```

# Bibliografía

- [1] BATCHELET, E. *Circular Statistics in Biology*. London: Academic Press, 1981.
- [2] FISHER, N. *Statistical Analysis of Circular Data*. Cambridge University Press, 1993.
- [3] JAMMALAMADAKA, S., AND SENGUPTA, A. *Topics in Circular Statistics*. World Scientific, 2001.
- [4] MARDIA, K. *Statistics of Directional Data*. London: Academic Press, 1975.
- [5] PEWSEY, A., AND RUXTON, G. *Circular Statistics in R*. Oxford:University Press, 2013.
- [6] SARMA, Y., AND JAMMALAMADAKA, S. *Statistical Science and Data Analysis: Proceeding of the Third Pacific Area Statistical Conference*. VSP: Utrecht, Netherladns, 1993.
- [7] SCHMIDT, P. *Econometrics*. Marcel Dekker, 1976.