

# Una introducción al análisis de datos circulares con algunas aplicaciones

Pedro Nel Jaimes Jaimes  
Silvia Johanna Pineda Garavito

Universidad Industrial de Santander  
Facultad de Ciencias  
Escuela de Matemáticas  
Licenciatura en Matemáticas  
Bucaramanga  
2012

# Una introducción al análisis de datos circulares con algunas aplicaciones

**Pedro Nel Jaimes Jaimes**  
**Silvia Johanna Pineda Garavito**

Trabajo de grado como requisito  
parcial para optar el título de  
*Licenciados en Matemáticas*

Director  
**Dr. Germán Moreno Arenas**

**Universidad Industrial de Santander**  
**Facultad de Ciencias**  
**Escuela de Matemáticas**  
**Licenciatura en Matemáticas**  
**Bucaramanga**  
**2012**

# Agradecimientos

*A nuestros padres porque gracias a su cariño y apoyo incondicional, hemos llegado a realizar una de nuestras grandes metas, la culminación de una carrera profesional.*

*Un agradecimiento especial a nuestro director de tesis Dr. Germán Moreno Arenas; porque gracias a su invaluable colaboración este trabajo fue realizado.*

*Al director del Centro de Investigaciones Epidemiológicas (CIE), Facultad de Salud, Universidad Industrial de Santander (UIS) el Dr. Luis Ángel Villar por su disposición y colaboración, al aportarnos ideas y la información necesaria para nuestra tesis. Al igual que el Observatorio de Salud Pública de Santander, Registros Individuales de Prestación de Servicios en Salud (RIPS) por la información suministrada.*

*A Ivonne Mildred Martínez por sus palabras de felicitaciones en los buenos momentos, por su comprensión y aliento en momentos difíciles.*

*A todos los compañeros y profesores que hicieron parte de este logro.*

# Índice general

<b>Introducción</b>	<b>14</b>
<b>1. Datos Circulares</b>	<b>19</b>
1.1. Análisis descriptivo de datos circulares . . . . .	23
1.1.1. Representación gráfica . . . . .	23
1.1.2. Medidas de localización . . . . .	25
1.1.3. Medidas de dispersión . . . . .	26
1.1.4. Momentos circulares . . . . .	28
1.2. Distribuciones de probabilidad circulares . . . . .	29
1.2.1. Distribución Uniforme . . . . .	30
1.2.2. Distribución Cardiode . . . . .	30
1.2.3. Distribución Wrapped . . . . .	31
1.2.4. Distribución von Mises . . . . .	32
<b>2. Dos ejemplos de aplicación con datos circulares</b>	<b>36</b>
2.1. El dengue . . . . .	36
2.1.1. Análisis descriptivo circular del dengue por años . . . . .	39
2.1.2. Ajuste de los casos de dengue a una distribución vM . . . . .	52
2.2. Asistencia a urgencias . . . . .	56
2.2.1. Asistencia a urgencias según el género . . . . .	60

<b>3. Conclusiones</b>	<b>63</b>
<b>4. Ilustración de un ejemplo usando Software R</b>	<b>66</b>
<b>5. Bibliografía</b>	<b>68</b>

# Índice de tablas

2.1. Datos de dengue clasificados por semanas para cada año . . . . .	41
2.2. Parámetros estadísticos sobre el dengue año 2003 . . . . .	42
2.3. Parámetros estadísticos sobre el dengue año 2004 . . . . .	43
2.4. Parámetros estadísticos sobre el dengue año 2005 . . . . .	45
2.5. Parámetros estadísticos sobre el dengue año 2006 . . . . .	46
2.6. Parámetros estadísticos sobre el dengue año 2007 . . . . .	47
2.7. Parámetros estadísticos sobre el dengue año 2008 . . . . .	48
2.8. Parámetros estadísticos de los casos de dengue por año . . . . .	50
2.9. Parámetros $k$ y $\mu$ para cada año . . . . .	52
2.10. Parámetros $k$ y $\mu$ para el dengue entre 2003-2008 . . . . .	55
2.11. Datos de los pacientes que asistieron a urgencias en el año 2009	58
2.12. Parámetros estadísticos de asistencia a urgencias en el año 2009	58
2.13. Parámetros estadísticos de los casos de urgencias por género . . . . .	61

# Índice de figuras

1.	Ángulos de vuelo de palomas liberadas . . . . .	16
1.1.	Representación de la dirección media de vuelo de dos aves . . .	20
1.2.	Selección de origen y sentido de rotación . . . . .	22
1.3.	Dirección media de datos circulares . . . . .	22
1.4.	Relación entre coordenadas rectangulares y polares . . . . .	23
1.5.	Gráfico de datos circulares en ramas . . . . .	24
1.6.	Histograma circular . . . . .	24
1.7.	Distancia circular $d_0$ . . . . .	27
1.8.	Representación de la f.d.p. Cardiode . . . . .	31
1.9.	Representación de la función de densidad von Mises . . . . .	33
2.1.	Aedes Aegypti y Aedes Albopictus. . . . .	37
2.2.	Gráfico de datos circulares en rama e histograma 2003 . . . . .	42
2.3.	Gráfico de datos circulares en rama e histograma 2004 . . . . .	44
2.4.	Gráfico de datos circulares en rama e histograma 2005 . . . . .	45
2.5.	Gráfico de datos circulares en rama e histograma 2006 . . . . .	47
2.6.	Gráfico de datos circulares en rama e histograma 2007 . . . . .	47
2.7.	Gráfico de datos circulares en rama e histograma 2008 . . . . .	49
2.8.	Gráfico de datos circulares en rama 2003 al 2008 . . . . .	51
2.9.	Ajuste gráfico de los casos de dengue en el año 2003 . . . . .	53

2.10. Ajuste gráfico de los casos de dengue en el año 2004 . . . . .	53
2.11. Ajuste gráfico de los casos de dengue en el año 2006 . . . . .	54
2.12. Ajuste gráfico de los casos de dengue en el año 2007 . . . . .	54
2.13. Ajuste gráfico de los casos de dengue en el año 2008 . . . . .	55
2.14. Ajuste gráfico de los casos de dengue entre el 2003-2008 . . . . .	56
2.15. Gráfico de datos circulares en ramas para urgencias . . . . .	59
2.16. Ajuste gráfico de urgencias a una distribución vM . . . . .	60
2.17. Gráfico de datos circulares en ramas por género . . . . .	61
2.18. Ajuste gráfico por género a una distribución vM . . . . .	62

**TITULO:** UNA INTRODUCCIÓN AL ANÁLISIS DE DATOS CIRCULARES CON ALGUNAS APLICACIONES <sup>1</sup>

**AUTORES:** Pedro Nel Jaimes Jaimes, Silvia Johanna Pineda Garavito  
2

**PALABRAS CLAVE:** Datos direccionales; Análisis descriptivo; Modelos probabilísticos.

## DESCRIPCIÓN

Los datos direccionales a pesar de ser un tema de gran utilidad en diferentes ciencias son poco usados, ya que es un área con solo tres décadas de estudio en la estadística por lo cual no muchas personas tienen conocimientos sobre este tema. Entonces este trabajo tiene como objetivo presentar algunos conceptos sobre datos circulares, así como hacer algunas aplicaciones de estos conceptos en algunos conjuntos de datos. Se presentan situaciones, en diferentes áreas de la ciencia, que utilizan datos circulares y se desarrolla la parte descriptiva de los modelos probabilísticos y de estimación de datos circulares.

El presente trabajo ha sido organizado de la siguiente manera. En la introducción se muestran algunos ejemplos de datos direccionales en áreas como: Meteorología, Biología, Geografía, Geología, Psicología y Medicina. El primer capítulo contiene un breve resumen de la historia de los datos circulares, la estadística descriptiva circular (representación gráfica, medidas de localización circular, medidas de concentración y dispersión circular y momentos circulares), también algunas distribuciones de probabilidad circulares con sus propiedades más importantes.

En el segundo capítulo presentamos dos ejemplos de aplicaciones de las técnicas de análisis de datos circulares, estas son: 1). La incidencia del dengue en el Área Metropolitana de Bucaramanga en el periodo 2003-2008; y 2). La asistencia a urgencias a la Clínica Chicamocha S.A de Bucaramanga en el año 2009. Finalmente, en el tercer capítulo encontramos las conclusiones que se dan de las aplicaciones dadas en el capítulo anterior y de los datos circulares en general.

---

<sup>1</sup>Tesis

<sup>2</sup>FACULTAD DE CIENCIAS, LICENCIATURA EN MATEMÁTICAS.  
DIRECTOR Dr. Germán Moreno Arenas

**TITLE:** AN INTRODUCTION TO DATA ANALYSIS OF CIRCULAR WITH SOME APPLICATIONS <sup>1</sup>

**AUTHORS:** Pedro Nel Jaimes Jaimes, Silvia Johanna Pineda Garavito<sup>2</sup>

**KEY WORDS:** Directional data; Descriptive analysis; Probabilistic models.

## DESCRIPTION

Directional data despite being a subject of great use in different sciences are rarely used because it is an area with only three decades of study in statistics so not many people have knowledge on this subject. So this document is to present some concepts of circular data and make some applications of these concepts in some data sets. Situations occur in different areas of science, using circular data and develops the narrative of the probabilistic models and estimation of circular data.

This document is organized as follows. In the introduction, some examples of directional data in areas such as Meteorology, Biology, Geography, Geology, Psychology and Medicine. The first chapter contains a brief summary of the history of the circular data, circular descriptive statistics (graphical representation, circular measures of location, measures of concentration and dispersion circular and circular moments), also some circular probability distributions with their most important properties.

In the second chapter we present two examples of applications of the techniques of circular data analysis, these are: 1). The incidence of dengue in the metropolitan area of Bucaramanga in the period 2003-2008, and 2). Assistance to emergency to the Clinic Chicamocha S.A. of Bucaramanga in 2009. Finally, the third chapter we find the conclusions are given of the applications given in the previous chapter and the generally circular data.

---

<sup>1</sup>Thesis

<sup>2</sup>FACULTY OF SCIENCES, DEGREE IN MATHEMATICS.  
DIRECTOR Dr. Germán Moreno Arenas

# Introducción

En algunas áreas de investigación y de aplicación es necesario utilizar métodos estadísticos para analizar conjuntos de datos donde algunas de sus variables tienen características circulares, por ejemplo, en el registro de la dirección de un desplazamiento, dado que indicar una dirección de  $1^\circ$  o  $359^\circ$  es equivalente a decir que hay un grado de desplazamiento con respecto a la dirección  $0^\circ$ .

De una forma más general, existen variables que pueden ser representados sobre la superficie de una hiperesfera de grado  $p$ , y estos datos son denominados direccionales [Mardia y Jupp, 2000].

Los datos medidos en forma de ángulos o en forma de orientaciones bidimensionales son encontrados en casi todas las áreas de la ciencia. Por lo general, surgen en la biología, geografía, geología, geofísica, medicina, meteorología, oceanografía y muchas otras áreas. A continuación se presentan algunos ejemplos de datos de carácter direccional, también llamados angulares o circulares.

- Meteorología: Una fuente natural de datos direccionales es la dirección del viento. Las direcciones son normalmente representadas en el mapa en el sentido horario desde el Norte ( $0^\circ$ ), al Este ( $90^\circ$ ), al Sur ( $180^\circ$ ), al oeste ( $270^\circ$ ) y de nuevo hacia el Norte ( $360^\circ$ ). También son registrados

como datos circulares la ocurrencia de truenos y la lluvia en ciertos períodos de tiempo. Considerando la conversión de las medidas del tiempo en ángulos, módulo algún período, la longitud del período de tiempo puede ser identificada por  $360^\circ$ . El período de tiempo usual es de un día o un año. Por ejemplo, si el período de tiempo es de un día, se puede registrar el número de veces que truena cada dos horas durante el día en épocas de verano. Para un año, se puede registrar el número de veces que llueve cada mes durante una década. El análisis de esta información es útil, por ejemplo, para ofrecer mayor seguridad en los vuelos aéreos.

- **Biología:** Existen varios experimentos importantes con aves. Por ejemplo, para determinar si las características celestes son utilizadas para orientar el vuelo, o si la orientación del vuelo es aleatoria bajo ciertas condiciones (por ejemplo, en la lluvia), o aún la dirección de migración, así como de retorno de las aves. Estos experimentos inducen el uso de datos circulares.

El análisis de estos datos es útil para determinar los movimientos migratorios de los animales y fijar estrategias de conservación de las especies. Por ejemplo, en un experimento sobre palomas mensajeras Schmidt-Koenig (1981) obtuvieron el siguiente conjunto de datos de ángulos de vuelo de 15 palomas puestas en libertad una por una (ordenados por dirección): 85, 135, 135, 140, 145, 150, 150, 150, 160, 185, 200, 210, 220, 225, 270. (ver Figura 1)

- **Geografía:** En el estudio de la ocurrencia de terremotos en una región se registran la longitud y la latitud del epicentro. Además, estos datos pueden ser cíclicos como en el estudio del número de terremotos en los

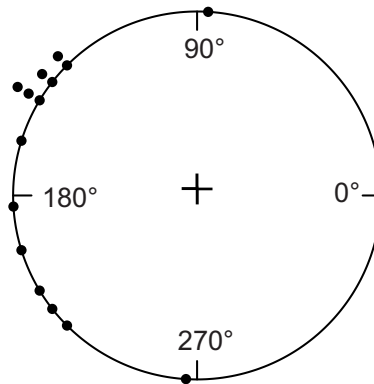


Figura 1: Ángulos de vuelo de palomas liberadas

últimos años o en los días siguientes a un terremoto más intenso.

- Geología: Los estudios de orientaciones de las grietas en las rocas son usados para interpretar las fuerzas tectónicas. Los estudios de orientaciones de las ondulaciones de la superficie terrestre son utilizados para interpretar las direcciones de las corrientes de viento o de agua.
- Psicología: Para simular gravedad cero en los viajes espaciales, se hacen experimentos en el agua con buzos y nadadores para evaluar su capacidad para percibir la verdadera dirección horizontal y vertical, en la falta de recurso visual. Los datos consisten en las desviaciones con respecto a la verdadera dirección.
- Medicina: El número de muertes debido a una enfermedad o el número de enfermos con una determinada enfermedad en cada mes a lo largo de unos años forman una distribución circular. La tasa de mortalidad mensual puede ser vista como una distribución circular bajo el supuesto de que la población-base es estacionaria, de modo que la velocidad es proporcional al número de muertes. Un médico podría estar interesado en el día de inicio de los síntomas en pacientes de una determinada

enfermedad que presenta repeticiones en el tiempo. Además del análisis del lugar de los inicios de una enfermedad en particular y la dirección de los contagios, proporciona información de hacia dónde se dirige dicha enfermedad.

Este trabajo se enfoca en el análisis de datos circulares, que se pueden representar como ángulos o puntos sobre la circunferencia de un círculo unitario. Una de las características de estos datos, por ser expresados en ángulos, es que la representación como un ángulo o un vector unitario no es necesariamente única ya que depende de la elección de la dirección cero y el sentido de rotación. Por lo tanto estos datos tienen rasgos únicos y nuevos en términos de modelación y tratamiento estadístico. Debido a que estos datos se consideran dentro de un ciclo, hace que muchas técnicas lineales usadas a menudo sean engañosas y no tengan sentido; por consiguiente, la media muestral, la varianza, la correlación, la regresión y otras medidas estadísticas, necesitaron ser reinventadas para analizar estos datos.

Nuestro objetivo es hacer que los lectores noten las limitaciones de los métodos estadísticos clásicos, presentado algunos conceptos básicos de análisis de datos circulares y algunas aplicaciones a conjunto de datos reales; esperamos que las personas interesadas obtengan información básica, permitiéndoles escoger un modelo correcto para analizar y hacer deducciones sobre los problemas de datos circulares que encuentren.

Este trabajo está diseñado en tres capítulos, en el primer capítulo se presenta una descripción de las técnicas y métodos de análisis de datos circulares como lo son las medidas de localización, concentración, dispersión circular y distribuciones de probabilidad para datos circulares. En el segundo capítulo se presentan dos ejemplos de aplicaciones de las técnicas de datos circulares, estas son:

1. El análisis de la incidencia del dengue en el Área Metropolitana de Bucaramanga (AMB) en el período 2003-2008.
2. El análisis descriptivo y ajuste de las asistencias a urgencias en la Clínica Chicamocha S.A de Bucaramanga en el 2009.

Finalmente presentamos una sección de conclusiones de este trabajo.

# Capítulo 1

## Datos Circulares

Las referencias más remotas de análisis de datos circulares datan de finales del siglo XVIII, de la mano del geólogo y astrónomo británico Rev. John Mitchell (considerado el padre de la sismología por sus estudios sobre terremotos) quien emplea estos análisis para estudiar las separaciones angulares entre estrellas con el objeto de probar la hipótesis de que sus direcciones estaban uniformemente distribuidas [Muñoz, 2008].

Los primeros trabajos en estadística circular tuvieron su origen en intentar comprender la habilidad de las palomas mensajeras para retornar a sus hogares. Posteriormente estos análisis se extienden a los estudios del empleo que hacen las aves migratorias de su reloj interno y de la dirección del sol para orientarse [Muñoz, 2008].

En diversos campos surgen problemas estadísticos donde la medida de los datos no es escalar sino angular, dando una orientación en el plano (datos circulares). Los datos circulares surgen principalmente de dos formas: la forma brújula llamados datos direccionales, por ejemplo, las direcciones del viento y las direcciones migratorias de los pájaros; y la forma reloj, por ejemplo,

horas de un día, días de una semana, meses de un año, etc.

Los datos circulares se encuentran presentes en muchos estudios científicos y requieren para su análisis el uso de herramientas y técnicas estadísticas especiales, que no son ampliamente conocidas, ni apreciadas debido a la falta de familiaridad con ellas. Generalmente se tratan estos datos con las técnicas usuales empleadas para los datos lineales (entiéndase como contrario a métodos de análisis de datos circulares), que son más comunes pero inapropiados para este tipo de datos.

Para ilustrar esta diferencia, suponga que dos aves salieron del mismo lugar y volaron a  $15^\circ$  y  $345^\circ$ , respectivamente. La media  $\mu = \sum_{i=1}^N X_i/N$  de estos dos valores indicaría que las aves volaron en dirección de  $180^\circ$ , pero note que esa dirección no representa la dirección promedio real en las que las aves volaron (ver Figura 1.1).

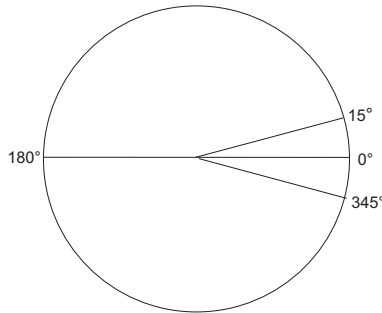


Figura 1.1: Representación de la dirección media de vuelo de dos aves

Un segundo ejemplo del uso inadecuado de la media  $\mu$  es usar el promedio de las horas de cuatro accidentes de tránsito que ocurrieron a las 23:00, 23:30, 00:30 y 1:00 horas como indicador de la hora promedio de accidentes, el promedio indica que es a medio día (12:00), cuando en realidad la mayoría de los accidentes ocurren a media noche.

A partir de estos ejemplos, es evidente que la media  $\mu$  no es una medida de tendencia central para este tipo de datos. La varianza que depende de la media también sufre el mismo problema y por lo tanto se necesita una medida alternativa de media y varianza cuando se trata de datos circulares.

Los datos circulares pueden ser considerados como un punto en una circunferencia de radio uno o un vector unitario en el plano. Una vez que la dirección inicial y la orientación del círculo han sido escogidas, cada observación circular puede ser especificada por el ángulo desde la dirección inicial hasta el punto en el círculo correspondiente a la observación. Los datos circulares se miden en grados, sin embargo, a veces es útil medirla en radianes. Hay que recordar que las medidas angulares se convierten de grados a radianes multiplicando por  $\frac{\pi}{180}$ .

La representación numérica de direcciones en dos dimensiones como un ángulo o un vector unitario no es necesariamente única, debido a que el valor angular depende de la elección de la dirección cero y el sentido de rotación. Lo que es considerado  $60^\circ$  para un matemático quien toma el Oriente como la dirección cero y el sentido contrario a las manecillas del reloj como la dirección positiva, para un geólogo es  $30^\circ$  debido a que toma el Norte como la dirección cero y el sentido de las manecillas del reloj como la dirección positiva (ver Figura 1.2).

Por ejemplo, supóngase que los datos están formados por los puntos  $50^\circ$ ,  $160^\circ$ ,  $210^\circ$  y  $300^\circ$  medidas de la manera habitual (el Oriente es cero, con valor positivo medido en sentido contrario a las manecillas del reloj), así su media es  $180^\circ$ , señalando que su dirección media es hacia el Occidente. Por otro lado, si con las mismas cuatro direcciones se miden los ángulos en el rango de  $-180^\circ$  a  $180^\circ$  (el Oriente es cero, con valor positivo medido en sentido contrario a

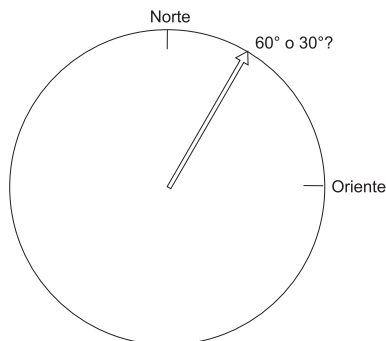


Figura 1.2: Selección de origen y sentido de rotación

las manecillas del reloj y negativo en sentido de las manecillas del reloj), sus valores se convierten en  $50^\circ$ ,  $160^\circ$ ,  $-150^\circ$  y  $-60^\circ$ , respectivamente. Su media es  $0^\circ$ , es decir, la dirección media es hacia el Oriente, que es exactamente lo opuesto a la dirección media anterior. Por último, con los mismos cuatro datos con ángulos que van desde  $-180^\circ$  hasta  $180^\circ$  tomando al Norte como cero y la rotación en sentido contrario a las manecillas del reloj como positivo, los datos son  $40^\circ$ ,  $150^\circ$ ,  $-70^\circ$  y  $-120^\circ$  su dirección media es hacia el Norte. Así, la dirección media puede variar para los mismos datos, dependiendo de la elección del origen y el sentido de rotación (ver Figura 1.3).

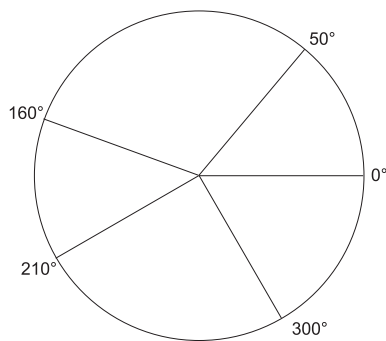


Figura 1.3: Dirección media de datos circulares dependiendo de la selección del origen

La metodología de datos circulares está basada en la relación entre coor-

denadas rectangulares y coordenadas polares. Cualquier punto  $P$  en el plano se puede representar como  $(X, Y)$  en coordenadas rectangulares o como  $(r, \alpha)$  en coordenadas polares donde  $r$  es la distancia al origen y  $\alpha$  se define como su dirección (ver Figura 1.4).

$$(r, \alpha) \leftrightarrow (r\cos(\alpha), r\sen(\alpha)) = (x, y)$$

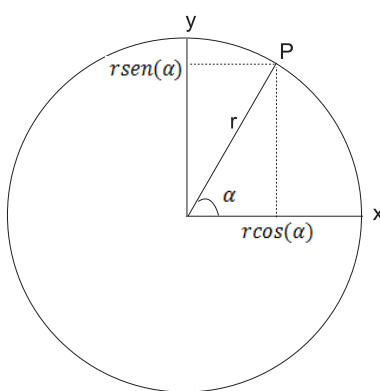


Figura 1.4: Relación entre coordenadas rectangulares y polares

## 1.1. Análisis descriptivo de datos circulares

El contenido de este capítulo está basado en la presentación que hace Jammalamadaka y SenGupta (2001) en su libro *Topics in Circular Statistics*. Presentamos las medidas de localización, concentración y dispersión y algunas distribuciones de probabilidad circulares.

### 1.1.1. Representación gráfica

*Gráfico de datos circulares en ramas:* Una forma sencilla de explorar las características de los datos circulares es representarlos por puntos en

una circunferencia del círculo unitario, un punto para cada observación (ver Figura 1.5). En el ejemplo descrito por Jammalamadaka y SenGupta (2001) se grafican las direcciones de vuelo de 15 palomas mensajeras.

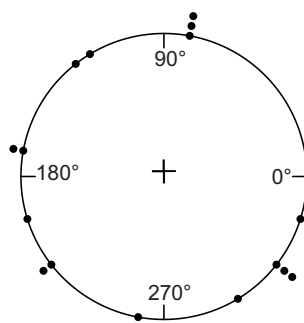


Figura 1.5: Gráfico de datos circulares en ramas

*Histograma circular:* Los datos circulares agrupados pueden ser representados por los histogramas, que son análogos a los histogramas en línea. Cada barra en un histograma circular está centrada en el punto medio del grupo correspondiente a los ángulos y la altura de la barra es proporcional a la frecuencia en ese grupo (véase Figura 1.6).

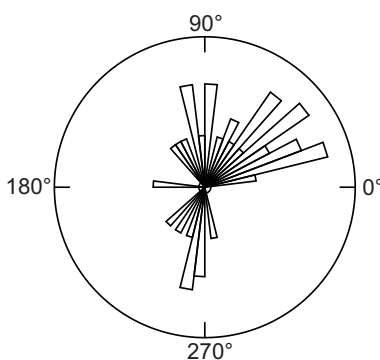


Figura 1.6: Histograma circular

### 1.1.2. Medidas de localización

*Dirección Media Circular:* Sea  $\alpha_1, \dots, \alpha_n$  un conjunto de observaciones circulares dadas en términos de ángulos y su transformación a coordenadas rectangulares, es decir,  $(\cos\alpha_i, \sen\alpha_i)$ ,  $i = 1, \dots, n$ . Con el vector suma resultante

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

donde  $R = \|\mathbf{R}\| = \sqrt{C^2 + S^2}$  es la longitud del vector resultante  $\mathbf{R}$ .

La dirección del vector resultante, se propone como la dirección media circular, se denota por  $\bar{\alpha}_0$  y se define teniendo en cuenta las siguientes ecuaciones:

$$\cos \bar{\alpha}_0 = \frac{C}{R}, \sen \bar{\alpha}_0 = \frac{S}{R},$$

entonces,

$$\bar{\alpha}_0 = \arctan \left( \frac{S}{C} \right)$$

Donde:

$$\bar{\alpha}_0 = \arctan \left( \frac{S}{C} \right) = \begin{cases} \arctan \left( \frac{S}{C} \right), & \text{si } C > 0, S \geq 0, \\ \frac{\pi}{2}, & \text{si } C = 0, S > 0, \\ \arctan \left( \frac{S}{C} \right) + \pi, & \text{si } C < 0, \\ \arctan \left( \frac{S}{C} \right) + 2\pi, & \text{si } C \geq 0, S < 0, \\ \text{sin definir,} & \text{si } C = 0, S = 0. \end{cases}$$

*Mediana Circular:* La dirección de la mediana del conjunto de  $\alpha_1, \dots, \alpha_n$  es cualquier ángulo  $\varphi$  de tal manera que la mitad de los puntos de datos se encuentran en el intervalo  $[\varphi, \varphi + \pi)$  y la mayoría de los puntos están más cerca de  $\varphi$  que de  $\varphi + \pi$ .

Cuando el tamaño de la muestra es impar, la mediana es uno de los puntos del conjunto de datos. Cuando  $n$  es par, se toma la mediana como el punto medio de dos puntos de datos adyacentes.

*Moda circular:* Al igual que en los datos lineales, la moda circular es el valor que más se repite, es decir, el valor con mayor frecuencia absoluta. Una distribución es bimodal cuando encontremos dos modas, es decir, dos datos que tengan la misma frecuencia absoluta máxima y una distribución es trimodal si encontramos tres modas, etc. Si todas las variables tienen la misma frecuencia diremos que no hay moda.

### 1.1.3. Medidas de dispersión

*Rango circular:* Una medida razonable de distancia circular entre dos puntos es tomar la menor distancia entre los puntos a lo largo de la circunferencia, es decir, para cualquier  $\alpha$  y  $\beta$  dos ángulos

$$d_0(\alpha, \beta) = \min(\alpha - \beta, 2\pi - (\alpha - \beta)) = \pi - |\pi - |\alpha - \beta||.$$

Por ejemplo, la distancia entre  $A$  y  $B$  puede ser la longitud del arco  $ANB$  o la longitud de  $ASB$  (ver Figura 1.7). Pero el arco  $ANB$  es más corto que el arco de  $ASB$ , se define la distancia circular como la longitud del arco  $ANB$ . Es evidente que no hay dos puntos de la circunferencia de un círculo que puedan estar más lejos de  $\pi$ , es decir, la distancia circular  $d_0(\alpha, \beta)$ , siempre se encuentra en  $[0, \pi]$ .

Por lo tanto, el rango circular es definido como el arco más corto en la

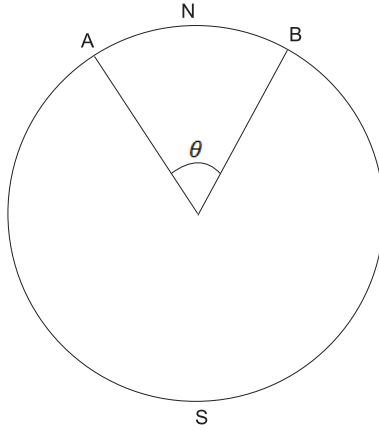


Figura 1.7: Distancia circular  $d_0$

circunferencia que contenga el conjunto de datos, y se denota

$$r_c = d_0(\alpha, \beta),$$

donde  $\alpha$  denota el menor ángulo y  $\beta$  denota el mayor ángulo del conjunto de datos.

*Varianza Circular:* Sea

$$\tilde{C} = \frac{1}{n} \sum_{i=1}^k (\cos \alpha_i), \tilde{S} = \frac{1}{n} \sum_{i=1}^k (\sen \alpha_i)$$

y

$$\tilde{R} = (\tilde{C}^2 + \tilde{S}^2)^{\frac{1}{2}}, 0 \leq \tilde{R} \leq 1,$$

donde  $\tilde{R}$  es la longitud media resultante. Si las direcciones  $\theta_1, \dots, \theta_n$  son cercanas entre sí entonces  $\tilde{R}$  será cercano a 1, por otro lado, si  $\theta_1, \dots, \theta_n$  están muy dispersas entonces  $\tilde{R}$  será casi 0, así  $\tilde{R}$  es una medida de concentración de un conjunto de datos. La longitud del vector resultante  $R$  y la longitud

media resultante se relacionan de la siguiente manera  $R = n\tilde{R}$ .

Para el propósito descriptivo e inferencial, la longitud media resultante  $\tilde{R}$  es una importante en las medidas de dispersión de los datos circulares. La más simple de ellas es la varianza de la muestra circular, se define como

$$V = 1 - \tilde{R} = 1 - \frac{R}{n},$$

donde  $0 \leq \tilde{R} \leq 1$ , entonces  $0 \leq V \leq 1$

*Desviación estándar circular:* A veces es útil tener un análogo de la desviación estándar lineal para los datos circulares. Una manera de obtenerla es transformando la varianza de la muestra ( $V$ ). Por lo tanto, la desviación estándar circular está dada por

$$s = \sqrt{-2\ln(1 - V)} = \sqrt{-2\ln(\tilde{R})},$$

donde  $\tilde{R} = \frac{R}{n}$  y  $n$  el tamaño de la muestra.

Note que  $s$  toma valores entre  $[0, \infty]$ , mientras que  $V$  toma valores entre  $[0, 1]$ .

#### 1.1.4. Momentos circulares

Sea

$$\bar{C} = \frac{1}{n} \sum_{j=1}^n \cos \alpha_j$$

y

$$\bar{S} = \frac{1}{n} \sum_{j=1}^n \text{sen } \alpha_j$$

además como  $e^{i\alpha} = \cos \alpha + i \operatorname{sen} \alpha$ , entonces

$$\bar{C} + i\bar{S} = \frac{1}{n} \sum_{j=1}^n e^{i\alpha_j}$$

De acuerdo con lo anterior el momento muestral de orden  $p$  viene dado por:

$$\frac{1}{n} \sum_{j=1}^n e^{ip\alpha_j} = \frac{1}{n} \sum_{j=1}^n \cos p\alpha_j + i \frac{1}{n} \sum_{j=1}^n \operatorname{sen} p\alpha_j$$

donde  $p = 0, 1, 2, \dots$

## 1.2. Distribuciones de probabilidad circulares

Una distribución circular es una distribución de probabilidad cuyo total de probabilidad se concentra en la circunferencia de un círculo unitario. Dado que cada punto de la circunferencia representa una dirección, tal distribución es una forma de asignar probabilidades a las diferentes direcciones. El rango de una variable aleatoria circular (va)  $\theta$ , medido en radianes, puede ser tomado como  $[0, 2\pi)$  o  $[-\pi, \pi)$ .

Las distribuciones circulares son de dos tipos:

- i. Discretas: Asigna probabilidad sólo a un número contable de direcciones.
- ii. Absolutamente continuas.

En las distribuciones absolutamente continuas, la función de densidad de probabilidad (f.d.p.) existe y tiene las siguientes características básicas:

- i.  $f(\theta) \geq 0$

ii.  $\int_0^{2\pi} f(\theta) d\theta = 1$

iii.  $f(\theta) = f(\theta + k \cdot 2\pi)$  para cualquier  $k$  entero (es decir,  $f$  es periódica)

A continuación se dan las características de algunas distribuciones absolutamente continuas.

### 1.2.1. Distribución Uniforme

La distribución uniforme es la distribución más básica en el círculo y a menudo se utiliza como modelo nulo. Esta es la única distribución sobre el círculo que es invariante bajo rotación y reflexión. Su función de densidad es:

$$f(\theta) = \frac{1}{2\pi}$$

### 1.2.2. Distribución Cardiode

La perturbación de la densidad uniforme por la función coseno da lugar a una distribución denominada cardiode  $C(\mu, \rho)$ , cuya función de densidad es:

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

Otras características de esta distribución son:

- a. La longitud media resultante es  $\rho$ .
- b. La media de la dirección es  $\mu$ .
- c. La distribución es simétrica y unimodal en  $\mu$  si  $\rho > 0$  (ver Figura 1.8).
- d. Si  $\rho = 0$  la distribución se reduce a distribución uniforme (ver Figura 1.8).

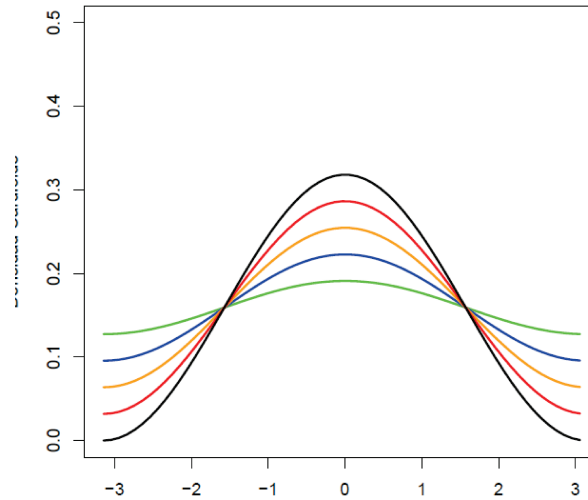


Figura 1.8: Representación de la f.d.p. Cardiode  $C(o, \rho)$ ,  $\rho = 0,1$  (verde),  $\rho = 0,2$  (azul),  $\rho = 0,3$  (naranja),  $\rho = 0,4$  (rojo),  $\rho = 0,5$  (negro).

### 1.2.3. Distribución Wrapped

Dada una variable aleatoria  $X$  en la recta real con función de densidad  $f(X)$  y función de distribución  $F(X)$ , se puede obtener la variable aleatoria circular  $X_w$ , definida por

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

Entonces, la función de densidad de probabilidad se puede obtener “envolviendo” o “enrollando” (wrapping)  $F(X)$  alrededor de un círculo de radio uno.

La función de distribución Wrapped de  $X_w$  viene dada por

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

De esta manera tenemos que si  $x$  tiene función de densidad  $f$ , entonces

la función de densidad  $f_w$  de  $x_w$  es

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

Hay diversos tipos de distribuciones Wrapped, una de ellas es la distribución normal Wrapped (envuelta)  $NE(\mu, \rho)$  la cual se obtiene al envolver la distribución  $N(\mu, \sigma^2)$  en el círculo, donde  $\sigma^2 = -2 \ln \rho$ , es decir,  $\rho = e^{-\frac{\sigma^2}{2}}$ .

Por consiguiente, la función de densidad de  $NE(\mu, \rho)$  viene dada por

$$\varphi_w(\theta; \mu, \rho) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu) \right\}, \quad 0 \leq \theta \leq 2\pi, \quad 0 \leq \rho \leq 1.$$

#### 1.2.4. Distribución von Mises

Entre las distribuciones de probabilidad circulares existentes, una de las más usadas para la modelización de variables circulares es la distribución de von Mises. En el análisis direccional, la distribución de von Mises juega un papel similar al de la distribución normal en el análisis lineal. Por esta razón, también es conocida como “distribución circular normal”. La función de densidad es:

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

donde

- i.  $I_0$  denota la función de Bessel modificada de primer tipo y orden cero,

$$I_0(k) = \frac{1}{2\pi} \int_0^{2\pi} e^{k \cos \theta} d\theta$$

- ii. El parámetro  $\mu$  es la media de las direcciones.
- iii.  $k$  es el parámetro de concentración.

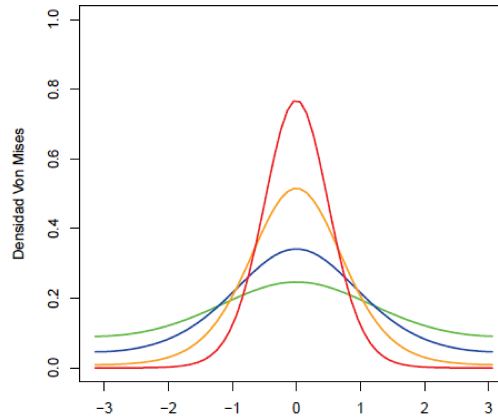


Figura 1.9: Representación de la función de densidad von Mises  $vM(0, k)$ ,  $k = 0,5$  (verde),  $k = 1$  (azul),  $k = 2$  (naranja),  $k = 4$  (rojo).

Como se puede ver en la Figura 1.9 la distribución von Mises es unimodal y es simétrica sobre  $\theta = \mu$ , así mismo, cuanto más grande sea el parámetro  $k$  su función de densidad estará más concentrada.

La distribución von Mises juega un papel crucial en el campo de los datos circulares, de hecho, juega el mismo papel que la distribución gaussiana en los datos lineales. La relación de la distribución von Mises con otras distribuciones es la siguiente: Si  $k = 0$ , entonces  $vM(\mu, k)$  es la distribución uniforme. Además, para  $x$  próximos a cero la aproximación  $\exp(x) \simeq 1 + x$ , muestra que para  $k$  pequeños  $vM(\mu, k) \simeq C(\mu, \frac{k}{2})$ , donde  $C(\mu, \frac{k}{2})$  denota una distribución cardiode. Esto quiere decir que la distribución cardiode es usada, principalmente, como una aproximación a la distribución von Mises cuando hay poca concentración.

## Estimación de los parámetros de una distribución von Mises

Sean  $\theta_1, \theta_2, \dots, \theta_n$  un conjunto de observaciones de una muestra aleatoria simple de una distribución von Mises con parámetros  $\mu$  y  $k$ . A partir de dichas observaciones, podemos plantearnos estimar los parámetros que caracterizan nuestro modelo  $vM(\mu, k)$ . Para ello, utilizaremos el método de los momentos.

Los momentos de la distribución von Mises generalmente se calculan como los momentos de la variable  $z = e^{i\theta}$ , es decir, representando los datos como puntos en el plano. A estos momentos se les denomina *momentos circulares*.

Si definimos la función de Bessel modificada de orden  $p$  como

$$I_p(k) = \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) \exp(k \cos \theta) d\theta, p = 0, 1, 2, \dots$$

y tenemos en cuenta la relación:

$$\frac{1}{2\pi} \int_0^{2\pi} \text{sen}(p\theta) \exp(k \cos \theta) d\theta = 0,$$

el momento de orden  $p$  de la densidad von Mises está dado por

$$\begin{aligned} \frac{1}{2\pi I_0(k)} \int_0^{2\pi} e^{ip\theta} \exp(k \cos(\theta - \mu)) d\theta &= \frac{1}{2\pi I_0(k)} \int_0^{2\pi} e^{ip(\omega + \mu)} \exp(k \cos \omega) d\omega \\ &= \frac{e^{ip\mu}}{2\pi I_0(k)} \int_0^{2\pi} [\cos(p\omega) + i \text{sen}(p\omega)] \exp(k \cos \omega) d\omega = \frac{I_p(k)}{I_0(k)} e^{ip\mu} \\ &\quad \frac{I_p(k)}{I_0(k)} [\cos(p\mu) + i \text{sen}(p\mu)]. \end{aligned}$$

El momento muestral de orden  $p$  de la variable  $z = e^{i\theta}$  viene dado por:

$$\frac{1}{n} \sum_{j=1}^n (e^{i\theta_j})^p = \frac{1}{n} \sum_{j=1}^n \cos(p\theta_j) + i \frac{1}{n} \sum_{j=1}^n \text{sen}(p\theta_j).$$

Igualando los momentos poblacionales a los momentos muestrales para  $p = 1$  se obtienen las ecuaciones que definen los estimadores  $\hat{\mu}$  y  $\hat{k}$  por el método de los momentos para  $\mu$  y  $k$ , respectivamente:

$$A(\hat{k}) \cos \hat{\mu} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i, \quad (1.1)$$

$$A(\hat{k}) \sen \hat{\mu} = \frac{1}{n} \sum_{i=1}^n \sen \theta_i, \quad (1.2)$$

donde  $A(\hat{k}) = \frac{I_1(\hat{k})}{I_0(\hat{k})}$ .

Suponiendo que  $\sum_{i=1}^n \cos \theta_i \neq 0$ , haciendo el cociente de las ecuaciones (1.1) y (1.2) resulta:

$$\hat{\mu} = \tan^{-1} \left( \frac{\sum_{i=1}^n \sen \theta_i}{\sum_{i=1}^n \cos \theta_i} \right), \quad (1.3)$$

donde  $\tan^{-1}$  denota la función arcotangente.

El estimador por el método de los momentos de  $\mu$  es la dirección de la media muestral

$$\frac{1}{n} \sum_{j=1}^n \cos(\theta_j) + i \frac{1}{n} \sum_{j=1}^n \sen(\theta_j).$$

Multiplicando (1.1) por  $\cos \hat{\mu}$  y (1.2) por  $\sen \hat{\mu}$  y sumando se obtiene que  $\hat{k}$  es solución de

$$A(\hat{k}) = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \hat{\mu}). \quad (1.4)$$

## Capítulo 2

# Dos ejemplos de aplicación con datos circulares

### 2.1. El dengue

El dengue es la enfermedad viral con más alta prevalencia transmitida por mosquitos y considerada una de las enfermedades infecciosas más graves en todo el mundo. Es transmitido esencialmente por el mosquito *Aedes Aegypti* como vector principal (ver Figura 2.1), que se encuentra en las regiones tropicales y subtropicales. Esto incluye países en África, norte de Australia, Sudamérica, Centroamérica y México.

La enfermedad se propaga por la picadura de la hembra del mosquito *Aedes Aegypti* o *Aedes albopictus*, que ha adquirido inicialmente sangre de una persona infectada por alguno de los cuatro serotipos del virus del dengue. Luego de la ingestión de sangre infectante, el mosquito puede transmitir el virus después de un período de 8 a 12 días de incubación dependiendo de la temperatura ambiental. Los microorganismos que causan esta enfermedad

son los serotipos: DEN-1, DEN-2, DEN-3 y DEN-4. Cada serotipo crea inmunidad contra el mismo serotipo (homólogo) e inmunidad cruzada de corto plazo contra los otros tres, la cual puede durar varios meses; por eso es posible que una persona con dengue sea nuevamente contagiada después de un período tiempo de haber padecido la enfermedad.

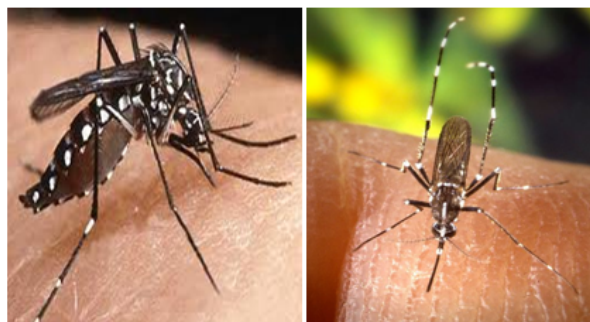


Figura 2.1: Aedes Aegypti y Aedes Albopictus.

*Estas imágenes son una obra de los Centros para el Control y la Prevención de Enfermedades, parte de los Departamento de Salud y Servicios Humanos de los Estados Unidos, adoptadas o realizados durante el desempeño de funciones oficiales de un empleado.*

La infección por cualquiera de los serotipos del virus del dengue puede resultar en diferentes grados de severidad de la enfermedad, que va desde una relativa fiebre benigna, llamado Dengue Clásico (DC) hasta manifestaciones hemorrágicas llamada Fiebre Hemorrágica de Dengue (FHD).

El DC es la manifestación más común de esta enfermedad y su agente causal puede ser cualquier serotipo de dengue y se manifiesta con las siguientes características: se inicia con una fiebre alta y repentina, a menudo de 40 a 40,5 grados Celsius; del segundo al quinto día puede aparecer una erupción plana y roja sobre casi todo el cuerpo y posteriormente, se presenta una

segunda erupción parecida al sarampión. Las personas infectadas pueden experimentar una mayor sensibilidad en la piel, dolores musculares y articulares y dolor retrocular.

La FHD es una complicación del DC. Hay dos teorías que tratan de explicar la forma como se origina este tipo de dengue. Una de ellas plantea que la FHD se presenta como una infección secundaria; esto ocurre cuando circula un serotipo después de otro. La otra teoría afirma que es una infección primaria, la cual se desarrolla debido a la presencia de anticuerpos. Se presenta como una enfermedad febril aguda acompañada de manifestaciones hemorrágicas y síntomas tales como: aumento súbito de temperatura, rubor facial encendido, anorexia, vómito, dolor de cabeza, dolores musculares y articulares. La temperatura se mantiene alta de 2 a 7 días y puede llegar a 41 grados Celsius, presentándose convulsiones febriles [Gonzales y Velasquez, 1995].

Anualmente ocurren alrededor de 100 millones de casos de DC y medio millón de casos de FHD, con presencia en más de 100 países en todo el mundo, que además de requerir eventualmente cuidados intensivos ocasiona aproximadamente 24.000 muertes anuales [*Guía de atención clínica integral del paciente con dengue, Ministerio de Protección Social, 2010*].

La primera epidemia de dengue documentada en laboratorios afectó al Caribe y a Venezuela en el período 1963-1964. Posteriormente, en la década del 70 del siglo XX la epidemia se propagó a Colombia y para ese entonces, el dengue comenzaba a hacerse endémico en la zona caribeña. Inicialmente aparecen casos de dengue reportados en Jamaica y por la intensificación de los viajes turísticos de la zona, se propaga a las demás islas del Caribe causando brotes explosivos en otros países cercanos: Venezuela, Guyana, Suriname,

Guayana Francesa, Belice, Honduras, El Salvador, Guatemala y México. Recientemente (en el siglo XXI) existen registros de epidemia de dengue iniciado en Río de Janeiro-Brasil y propagado a otros estados brasileños. Otras epidemias registradas son Bolivia en 1987, Paraguay y Ecuador en 1988 y Perú en 1990. Los estudios sugieren que varios millones de personas han sido afectadas. La preocupación en Colombia es que durante esos mismos años se registró un aumento notable de la ocurrencia de la FHD, que en muchos casos llevó a la muerte de los enfermos. [*Antecedentes de Dengue en Latinoamérica, Medicina y Prevención, 2008*]

### **2.1.1. Análisis descriptivo circular del dengue por años**

El principal propósito en esta sección es realizar un análisis descriptivo sobre la estacionalidad del dengue, en el área metropolitana de Bucaramanga, durante el período 2003 al 2008.

Los datos fueron aportados por el Centro de Investigaciones Epidemiológicas (CIE), Facultad de Salud, Universidad Industrial de Santander (UIS). Para poder llevar a cabo este análisis estadístico circular, los datos fueron previamente analizados y clasificados por semanas, esto se hizo de la siguiente manera:

- i. En la base de datos original encontramos dos tipos de enfermedades virales agudas: el Dengue y la Influenza. Se hizo una selección tomando sólo los datos de dengue, que son 659 casos entre DC y FHD.
- ii. Dado que los datos deben estar distribuidos alrededor de una circunferencia de radio 1, dividimos los  $360^\circ$  de la circunferencia entre 52

(número de semanas por año); este valor es aproximadamente  $6,92^\circ$  .

- iii. Los datos semanales son fijados en la mitad de cada intervalo, dado que no se conoce como es la distribución de los datos dentro de cada semana.

La clasificación de los datos de dengue por semanas para cada año se muestra en la Tabla 2.1, los grados que aparecen en esta tabla corresponden a la ubicación de los datos en cada intervalo (o cada semana).

Antes de presentar los análisis, es necesario aclarar que el rango circular calculado con el software estadístico R y su librería *CircStats* se disminuye una semana, por ejemplo, si los datos se ubican de la semana 21 a la semana 32 el programa lee esta información en grados como el intervalo ( $141,92^\circ - 218,07^\circ$ ) pero la longitud de la semana 21 es ( $138,46^\circ - 145,38^\circ$ ) y la de la semana 32 es ( $214,61^\circ - 221,53^\circ$ ), por lo tanto los intervalos ( $138,46^\circ - 141,92^\circ$ ) y ( $218,07^\circ - 221,53^\circ$ ) no se están teniendo en cuenta y la suma de los rangos de estos intervalos es de  $6,92^\circ$ , lo cual corresponde a una semana. Entonces para el análisis de los datos al valor obtenido con el software como rango circular se le sumará  $6,92^\circ$  .

A continuación se presenta el análisis descriptivo por cada año, en donde se van a detallar algunas características de los datos, además se hace una presentación, organización y resumen de los datos, usando tablas, gráficos y medidas de resumen con los siguientes parámetros: la dirección media, la mediana circular, la moda circular y la varianza circular.

Tabla 2.1: Datos de dengue clasificados por semanas para cada año

Semana	Representante	2003	2004	2005	2006	2007	2008
1	3,46°		5				
2	10,38°		1			4	1
3	17,30°		3			5	1
4	24,23°		2			4	4
5	31,15°		1			5	2
6	38,07°		3			5	6
7	45,00°		3			3	1
8	51,92°					4	5
9	58,84°		1		2	7	1
10	65,76°			3	2	5	2
11	72,69°				1	7	1
12	79,61°		6		2	5	
13	86,53°			1	4	3	4
14	93,46°		1			2	1
15	100,38°	1		1	2	10	4
16	107,30°	1			8	8	
17	114,23°	1		1	4	12	1
18	121,15°	4			4	5	1
19	128,07°	2	3		2	7	1
20	135,00°	6	3		4	5	
21	141,92°	5			2	9	
22	148,84°	1			3	10	
23	155,76°	1	2		3	9	
24	162,69°	5		1	5	6	
25	169,61°	3	3		3	9	
26	176,53°	2	4		2	6	1
27	183,46°	9	3		8	11	
28	190,38°	3	3		5	9	
29	197,30°	7	3		5	3	
30	204,23°	4	4		4	11	
31	211,15°	3	1			7	
32	218,07°	3				8	
33	225,00°	6			5	4	1
34	231,92°	5			4	6	
35	238,84°	5	1		1	10	1
36	245,76°	4	2		5	3	
37	252,69°	2	2		4	7	1
38	259,61°	7	1		4	4	4
39	266,53°	2	2		5	4	3
40	273,46°	5	2		1	2	
41	280,38°	5	1		3	2	1
42	287,30°	3	2		3	5	
43	294,23°	2	1		4	1	
44	301,15°	2				5	
45	308,07°		1		4	2	
46	315,00°		4		2	7	
47	321,92°	2	1		1	4	
48	328,84°	1	2		2	1	
49	335,76°	1	1		1	3	
50	342,69°	1	2		3	1	
51	349,61°						
52	356,53°						
TOTAL		117	80	7	127	275	53

## Año 2003

La Tabla 2.2, muestra las medidas de resumen de los casos de dengue en el año 2003, en donde se puede inferir que en la semana donde se obtuvieron más casos fue en la 27, con un total de 9 casos.

Además, en promedio los casos de dengue en el año 2003, tienden a presentarse en la semana 31 y se distribuyen en un amplio intervalo de 36 semanas (ver Tabla 2.2 y Figura 2.2).

Tabla 2.2: Parámetros estadísticos sobre el dengue año 2003

Año 2003	
<b>Total</b>	117
<b>Dirección Media</b>	242,55°
<b>Moda</b>	183,46°
<b>Rango</b>	249,23°
<b>Varianza</b>	0,40

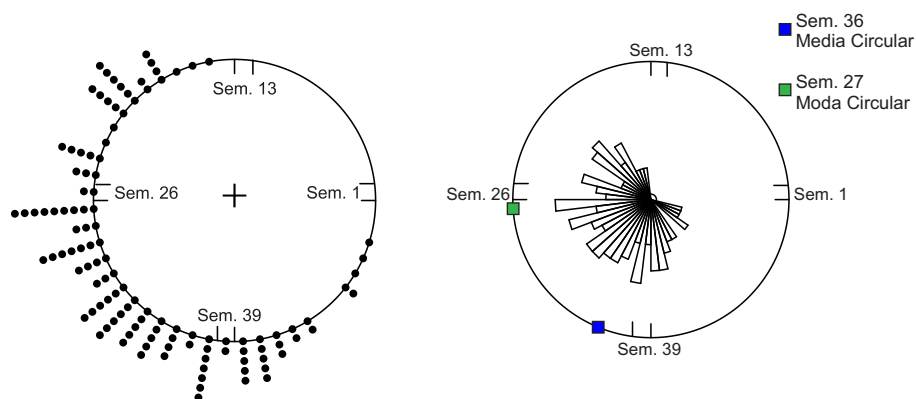


Figura 2.2: Gráfico de datos circulares en rama e histograma de los casos de dengue en el año 2003

Luego, en la Figura 2.2, podemos ver el inicio y la culminación de esta enfermedad en el año 2003, también, la ausencia de casos de dengue en las

primeras 14 semanas la cuales van de enero a principios de abril, también al analizar la variable semanas nos conduce a considerar una distribución sesgada a la derecha (esto se presenta debido a que los casos de dengue disminuyeron a medida que transcurrieron los meses), lo cual era de esperarse ya que en este año Bucaramanga tuvo temperaturas en promedio de  $25^{\circ}$  C y precipitación entre los meses de julio a septiembre, condiciones favorables para el desarrollo del agente transmisor de esta enfermedad.

### **Año 2004**

En la Tabla 2.3 y la Figura 2.3 se muestran los resultados obtenidos al hacer el análisis descriptivo a los 80 casos de dengue en el año 2004. Así, en la Tabla 2.3 se muestran las medidas descriptivas y en la Figura 2.3 su histograma circular y su gráfico en ramas. En promedio los casos de dengue en este año se presentaron en la semana 37 ( $256,85^{\circ}$ ), con una considerable dispersión de los casos ( $V = 0,96$ ) y además podemos inferir que hay poca asimetría como se pone de manifiesto en el histograma circular (ver Figura 2.3).

Tabla 2.3: Parámetros estadísticos sobre el dengue año 2004

<b>Año 2004</b>	
<b>Total</b>	80
<b>Dirección Media</b>	$256,85^{\circ}$
<b>Moda</b>	$79,61^{\circ}$
<b>Rango</b>	$332,31^{\circ}$
<b>Varianza</b>	0,96

Una característica bastante importante para resaltar es que durante todo el año hubo gran incidencia de la enfermedad, esto es debido al rango obtenido de 48 semanas ( $332,31^{\circ}$ ).

Teniendo en cuenta la misma figura los casos de dengue tienden a aglomerarse a finales de julio y principios de agosto, debido a las condiciones presentadas en este año que fueron muy favorables para el desarrollo de esta enfermedad. Además, los casos de dengue más frecuentes se presentaron en la semana 12 ( $79,61^\circ$ ) con un valor de 6 casos.

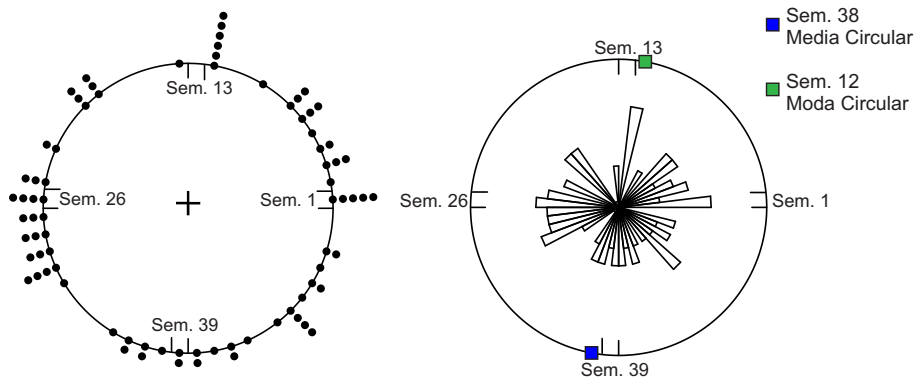


Figura 2.3: Gráfico de datos circulares en rama e histograma de los casos de dengue en el año 2004

### Año 2005

Este fue el año en el que se presentaron menos casos de dengue, tan solo 7 casos los cuales se distribuyeron en un rango de  $103,85^\circ$  que corresponde a 15 semanas que van de principios de marzo a mediados de junio (ver Figura 2.4). Además, la poca incidencia de casos de dengue se debe a que en el año 2005 las dos condiciones necesarias para la reproducción del mosquito no fueron muy frecuentes, debido a la ola invernal que predominó durante gran parte del año.

Teniendo en cuenta la información contenida en la Tabla 2.4 podemos inferir que los casos de dengue en promedio se presentan en la semana 14 con una alta concentración ( $V = 0,15$ ), dado a que los casos se distribuyen en un amplio intervalo de 15 semanas.

Tabla 2.4: Parámetros estadísticos sobre el dengue año 2005

Año 2005	
<b>Total</b>	7
<b>Dirección Media</b>	92,44°
<b>Moda</b>	65,76°
<b>Rango</b>	103,85°
<b>Varianza</b>	0,15

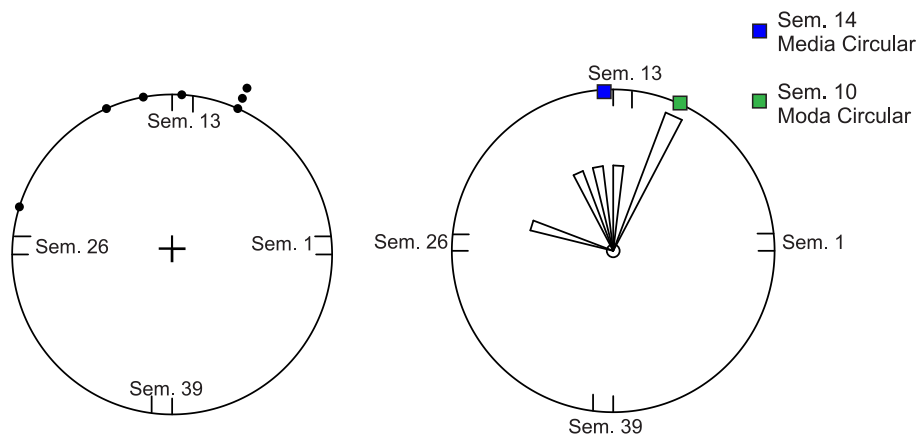


Figura 2.4: Gráfico de datos circulares en rama e histograma de los casos de dengue en el año 2005

También se observa un dato atípico o lejano que representa un solo caso de dengue en la semana 24 que es poco probable, entonces se puede pensar que donde se contagió esta persona, otras también lo hicieron solo que no fueron reportados (ver Figura 2.4).

### **Año 2006**

El número total de casos de dengue en este año fue de 127, su distribución bimodal (semana 16 y 27) sobre la circunferencia se puede ver en la Figura 2.5. En promedio los casos de dengue se presentaron en la semana 28 y se distribuyeron en un amplio intervalo de 42 semanas comprendidas entre marzo y diciembre (ver Tabla 2.5).

Tabla 2.5: Parámetros estadísticos sobre el dengue año 2006

<b>Año 2006</b>	
<b>Total</b>	127
<b>Dirección Media</b>	192,58°
<b>Moda 1</b>	107,30°
<b>Moda 2</b>	183,46°
<b>Rango</b>	290,77°
<b>Varianza</b>	0,65

### **Año 2007**

En este año se presentaron 275 casos de dengue de los cuales el mayor número de estos (12 casos) se presentaron en la semana 17. Además, en promedio esta enfermedad tiene alta incidencia en la semana 24, lo cual corresponde al mes de junio (ver Tabla 2.6).

Una característica bastante notable a diferencia de los otros años es que la enfermedad tuvo una gran prevalencia en gran parte del año con un rango

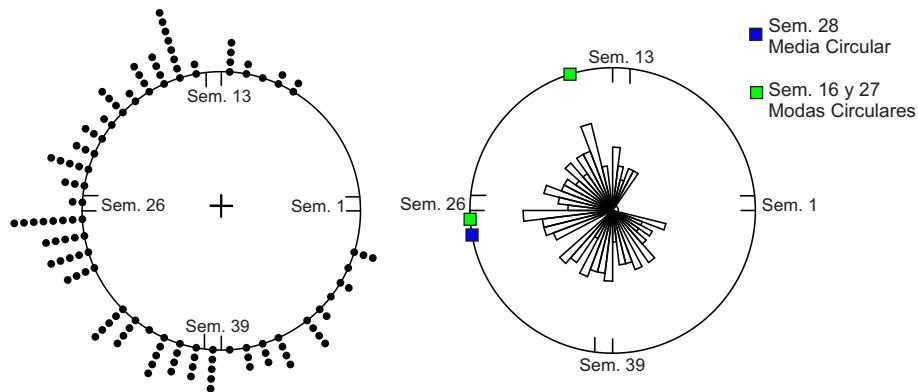


Figura 2.5: Gráfico de datos circulares en rama e histograma de los casos de dengue en el año 2006

Tabla 2.6: Parámetros estadísticos sobre el dengue año 2007

Año 2007	
<b>Total</b>	275
<b>Dirección Media</b>	159,56°
<b>Moda</b>	114,23°
<b>Rango</b>	339,23°
<b>Varianza</b>	0,72

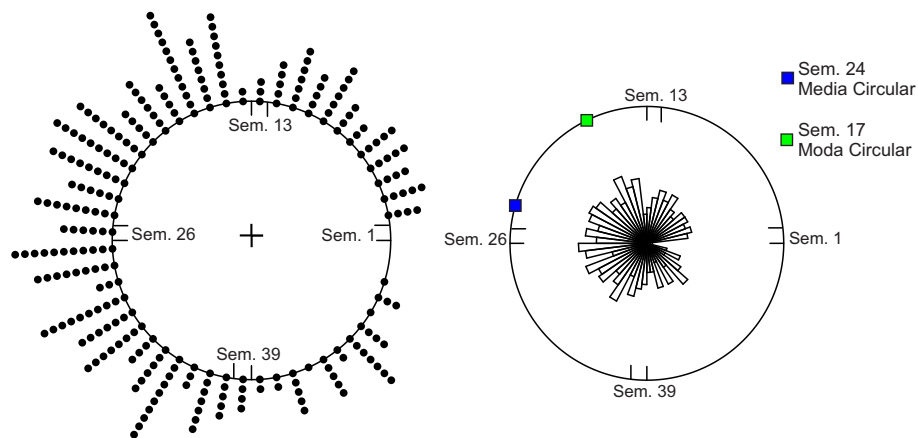


Figura 2.6: Gráfico de datos circulares en rama e histograma de los casos de dengue en el año 2007

de 49 semanas ( $339,23^\circ$ ) que va desde la semana 2 a la 50, esto es, por que las condiciones fueron excelentes para la propagación del dengue. Además se presentó dispersión de los casos ( $V = 0,72$ ).

### **Año 2008**

En este año se presentaron 53 casos de dengue, estos se distribuyeron en un intervalo de 40 semanas (rango de  $276,92^\circ$ ) las cuales van desde la semana 2 a la semana 41, es decir, desde principios de enero a mediados de octubre (ver Tabla 2.7 y Figura 2.7).

Tabla 2.7: Parámetros estadísticos sobre el dengue año 2008

<b>Año 2008</b>	
<b>Total</b>	53
<b>Dirección Media</b>	$46,37^\circ$
<b>Moda 1</b>	$17,30^\circ$
<b>Moda 2</b>	$38,07^\circ$
<b>Rango</b>	$276,92^\circ$
<b>Varianza</b>	0,53

Teniendo en cuenta la información contenida en la Tabla 2.7 podemos afirmar que en promedio la enfermedad tiene incidencia en la semana 7 (esto corresponde al mes de febrero), su distribución se dio en dos grandes grupos, el primero va desde la semana 2 a la 19 (es decir, de enero a mayo con 41 casos de dengue) y el segundo la semana 33 a la 41 (es decir de agosto a octubre con 11 casos de dengue), ver Figura 2.7.

Además, en la misma figura podemos notar que las semanas con más casos frecuentes de dengue fueron la 3 y la 6 cada una con un total de 6 casos, debido a esta característica la distribución de los datos se le adjudica el nombre de distribución bimodal.

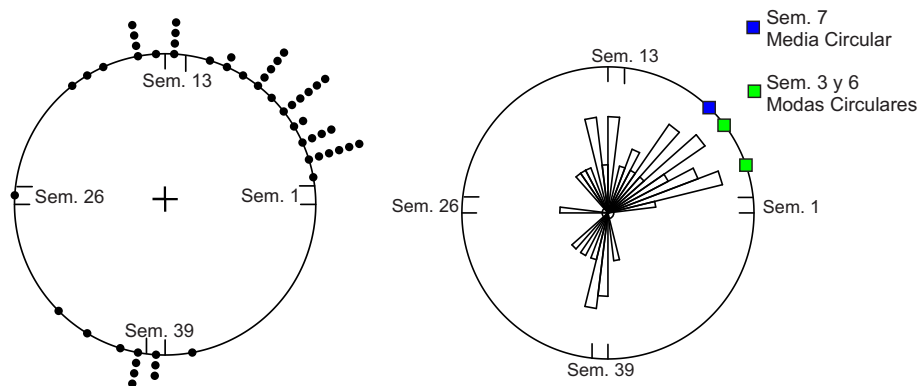


Figura 2.7: Gráfico de datos circulares en rama e histograma de los casos de dengue en el año 2008

### Comparación del análisis descriptivo del dengue

En la Tabla 2.8 se muestran los parámetros correspondientes a los casos de dengue por cada año, podemos notar que en el año 2005 se presentaron menos casos que en el año 2007, en el primer año se dieron 7 casos y en el segundo 275 casos respectivamente, esto se debe a que en el año 2005 las dos condiciones necesarias para la reproducción del mosquito no se dieron, por el contrario en el año 2007 las condiciones fueron excelentes.

También es interesante observar que el número de casos de dengue y el rango tienen un comportamiento directamente proporcional, es decir, a medida que aumentó el número de casos también aumentó el rango (2005 rango 15 semanas y 2007 rango 49 semanas), así de esta manera se puede mostrar que no existe una incidencia del dengue en alguna semana o mes específico, en cierto sentido la incidencia de esta enfermedad depende de las condiciones necesarias para la reproducción del mosquito transmisor.

Al observar el número de casos de dengue presentados en los diferentes

Tabla 2.8: Parámetros estadísticos de los casos de dengue por año

	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
<b>Total</b>	117	80	7	127	275	53
<b>D. Media</b>	242,55°	226,32°	92,44°	192,58°	159,56°	46,37°
<b>Moda 1</b>	183,46°	79,61°	65,76°	107,30°	114,23°	17,30°
<b>Moda 2</b>				183,46°		38,07°
<b>Rango</b>	249,23°	332,30°	103,85°	290,77°	339,23°	276,92°
<b>Varianza</b>	0,40	0,96	0,15	0,65	0,72	0,53

años, se nota que no es posible predecir el número de casos de dengue en los próximos años, ya que es muy variable (debido a los cambios climáticos presentes en cada año y como sabemos es muy importante que se den las condiciones climáticas favorables para la propagación y desarrollo de esta enfermedad), esta característica la podemos observar en la Figura 2.8 en donde los años 2003, 2006 y 2007 los casos de dengue se presentaron de una forma continua semana tras semana, en cambio en los años 2004, 2005 y 2008 tuvieron un comportamiento distinto ya que se nota una división de los casos en grupos.

En la Figura 2.8 podemos observar el inicio y la culminación de la enfermedad en cada uno de los años, claramente vemos que en el año 2003 la presencia de casos de dengue comenzaron en el mes de abril a diferencia de los otros años, en donde los casos iniciaron a principios de año, más explícitamente en el mes de enero. Así mismo, observamos que la incidencia de la enfermedad se dio hasta finales de cada año, a excepción del año 2005 en donde los casos de dengue se presentaron hasta finales del mes de junio.

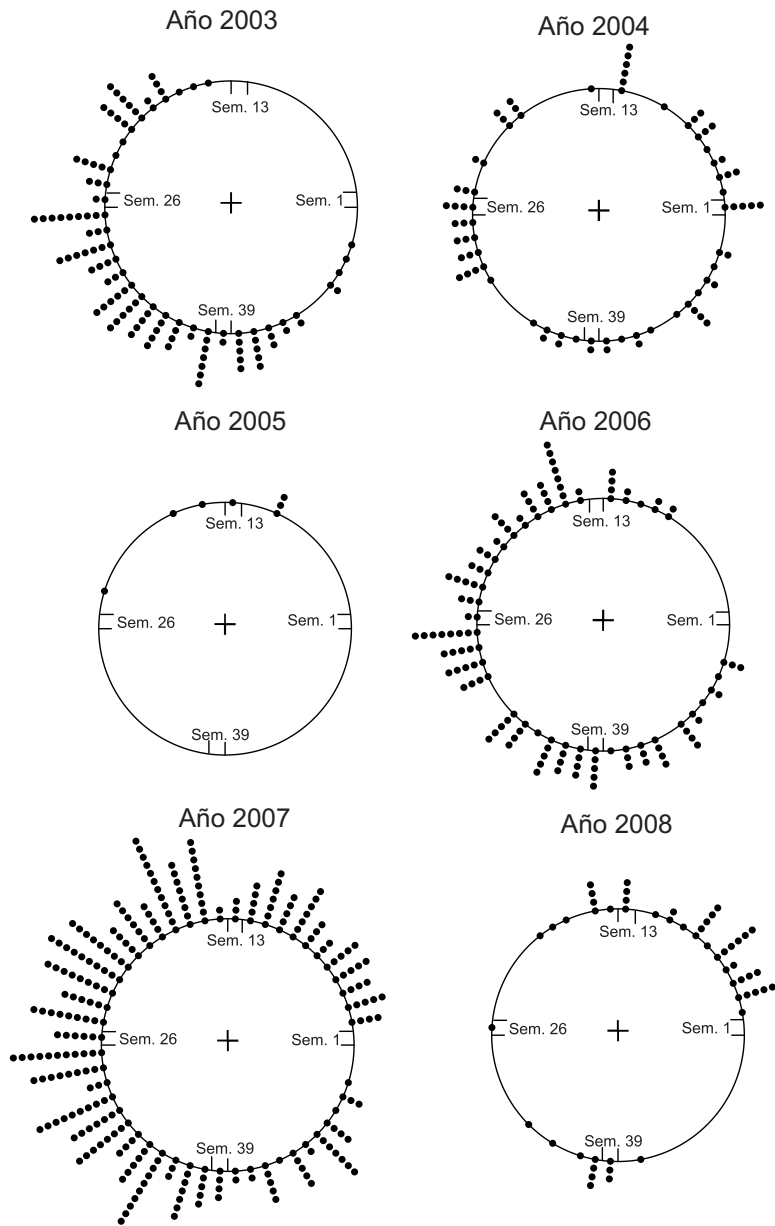


Figura 2.8: Gráficos de datos circulares en rama de los casos de dengue del 2003 al 2008

### 2.1.2. Ajuste de los casos de dengue a una distribución vM

Teniendo en cuenta las respectivas distribuciones de los datos de dengue para cada año y los análisis descriptivos correspondientes, nuestro propósito es confirmar si estos datos se pueden ajustar a una distribución circular. El proceso de ajuste se hará a una distribución von Mises, primero para cada año y luego en forma general, haciendo uso del software R y su librería *CircStats* para hallar los parámetros  $k$  y  $\mu$ . Debido a que en el año 2005 se presentaron 7 casos de dengue no es pertinente hacer un ajuste a la distribución von Mises, por lo tanto este año fue omitido en el siguiente análisis.

En la Tabla 2.9 se muestra los parámetros  $k$  y  $\mu$  para cada año, en donde se nota una variación en los valores de  $k$ , recordemos que este parámetro nos indica la concentración de los datos alrededor de la media circular ( $\mu$ ).

Tabla 2.9: Parámetros  $k$  y  $\mu$  para cada año

Año	$\mu$	$k$
2003	212,55	1,4896
2004	256,85	0,0742
2006	192,58	0,7431
2007	159,56	0,5771
2008	46,37	1,0403

Para el año 2003, tenemos que el parámetro de concentración fue  $k = 1,4896$  y  $\mu = 212,55^\circ$ . Después de obtener estos valores se procede a graficar la distribución de los datos contra la distribución von Mises. Así podemos inferir que para los valores de  $k$  y  $\mu$  los datos de dengue en el año 2003 se ajustan razonablemente a una distribución von Mises (ver Figura 2.9).

Para el año 2004, se obtuvieron valores de  $k = 0,0742$  y  $\mu = 256,85^\circ$ .

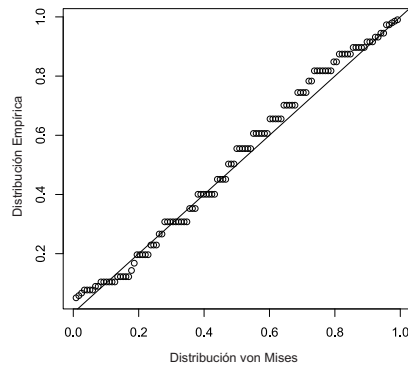


Figura 2.9: Ajuste gráfico de los casos de dengue en el año 2003 a una distribución von Mises. (La distribución empírica es la distribución de los datos).

Teniendo en cuenta estos resultados en la Figura 2.10 podemos observar que el ajuste de estos datos no es razonable, debido a que se encuentran alejados de la línea de ajuste.

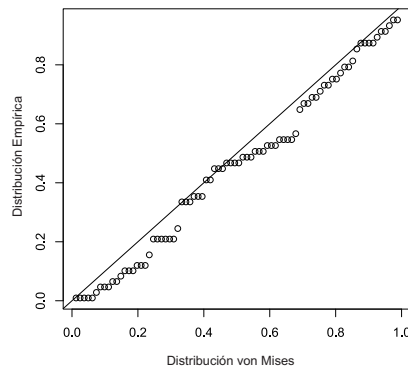


Figura 2.10: Ajuste gráfico de los casos de dengue en el año 2004 a una distribución von Mises.

En el año 2006, se presentó una media circular de  $\mu = 192,58^\circ$  y una concentración alrededor de esta medida de  $k = 0,7431$ . Teniendo en cuenta estos parámetros en la Figura 2.11 podemos observar que los datos tienen un ajuste razonable con la distribución von Mises.

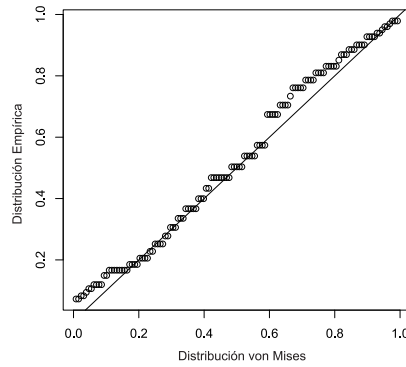


Figura 2.11: Ajuste gráfico de los casos de dengue en el año 2006 a una distribución von Mises.

En el año 2007, es donde se puede considerar que el ajuste a una distribución von Mises es muy bueno, con  $\mu = 159,56^\circ$  y  $k = 0,5771$ , esto es debido al número total de casos y a la poca dispersión que se presentaron en los datos (ver Figura 2.12).

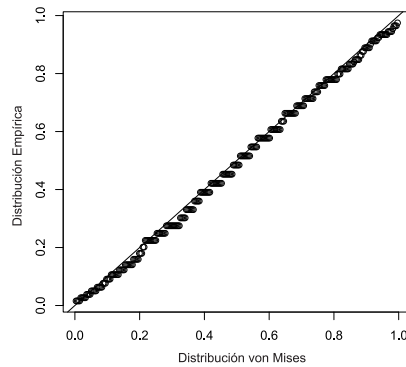


Figura 2.12: Ajuste gráfico de los casos de dengue en el año 2007 a una distribución von Mises.

Para el año 2008, tenemos que el ajuste de los datos a una distribución von Mises es poco razonable en comparación con el año anterior esto se puede observar en la Figura 2.13; los parámetros obtenidos en este año fueron  $k = 1,0403$  y  $\mu = 46,37^\circ$ .

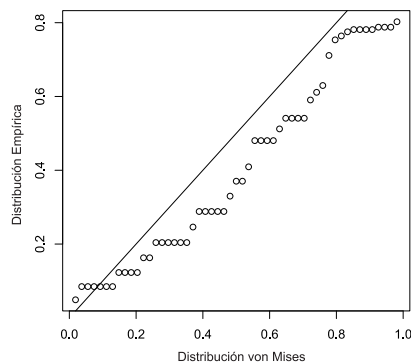


Figura 2.13: Ajuste gráfico de los casos de dengue en el año 2008 a una distribución von Mises.

La Tabla 2.10 muestra los parámetros: media circular y concentración de los casos de dengue entre el período 2003-2008, esto con el fin de poderlos ajustar a una distribución von Mises.

Tabla 2.10: Parámetros  $k$  y  $\mu$  para el dengue entre 2003-2008

Parámetros	Valor
$\mu$	179,68
k	0,4944

A pesar de que los ajustes hechos anteriormente para cada año no fueron completamente adecuados, asumimos de forma conveniente que los datos en cada año se ajustaron razonablemente a una distribución von Mises, esto con el propósito de hacer un ajuste usando todos los datos (659 casos de dengue). Para todos los casos de dengue, se obtuvo una media circular de  $\mu = 179,68^\circ$  y una concentración alrededor de esta medida de  $k = 0,4944$ . Al realizar el ajuste usando el software estadístico R con la librería *CircStats* podemos observar en la Figura 2.14 que los 659 casos de dengue registrados en los seis años (2003-2008) se ajustan razonablemente a una distribución von Mises.

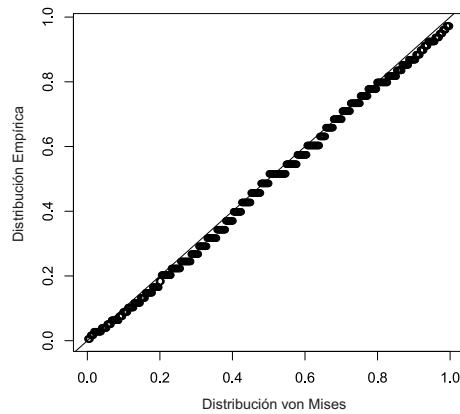


Figura 2.14: Ajuste gráfico de los casos de dengue durante el período 2003-2008 a una distribución von Mises.

De acuerdo con lo anterior en donde todos los casos de dengue se ajustaron a una distribución von Mises, podemos inferir que en promedio esta enfermedad tienen una incidencia entre finales de junio y principios de julio, aunque esta incidencia se podría considerar baja debido a que el parámetro de concentración fue pequeño (0,4944), es decir, los casos de dengue se encuentran dispersos con respecto a la media circular.

## 2.2. Asistencia a urgencias

Los datos de las personas atendidas por urgencias en el año 2009 en la Clínica Chicamocha S.A. fueron aportados por el Observatorio de Salud Pública de Santander, Registros Individuales de Prestación de Servicios en Salud (RIPS). Para poder llevar a cabo este análisis, los datos fueron previamente analizados y clasificados por horas, esto se hizo de la siguiente manera:

1. En la base de datos original encontramos varios centros salud (clínicas y hospitales) en donde los pacientes asistieron a urgencias en el año 2009, en total habían 22470 datos. Se hizo una selección tomando sólo

los datos de la Clínica Chicamocha S.A de Bucaramanga, que son 958.

2. Dado que los datos deben estar distribuidos alrededor de una circunferencia de radio 1, dividimos los  $360^\circ$  de la circunferencia entre 24 (número de horas en un día); este valor es  $15^\circ$  exactamente.
3. Los datos son redondeados a la hora, dado que no se conoce como es la distribución de los datos dentro de cada *intervalo de hora*.
4. El análisis de los datos se va hacer teniendo en cuenta el reloj, por lo tanto la dirección cero u *origen* en esta ocasión será el norte (00:00 a.m.) y dirección positiva va en el sentido de las manecillas del reloj.

La clasificación de los datos de los pacientes que asistieron a urgencias a la Clínica Chicamocha en el año 2009 se muestra en la Tabla 2.11, los grados que aparecen en esta tabla corresponden a la ubicación de los datos en cada hora.

Los 958 casos de asistencia a urgencias en la Clínica Chicamocha S.A tuvieron una distribución unimodal (0:00 am con 146 datos) la cual podemos ver en la Figura 2.15, así mismo, podemos inferir que los pacientes en promedio asisten a urgencias entre las 6:00 pm y las 7:00 pm, con una razonable dispersión de los casos ( $V = 0,80$ ), esto se debe a la alta asistencia de los pacientes a urgencias en las horas de la noche (ver Tabla 2.12).

Ahora, teniendo en cuenta que los datos se redondean a la hora más cercana. Esto no es visto como un problema importante: “Fisher (1993, p. 35) señala que los errores de redondeo no se convierten en un problema cuando la estimación estadística de la muestra de los datos circulares es asignada a cerca de 12 grupos (es decir, el redondeo con una precisión de 2 h), además, él cita específicamente el redondeo a la hora más cercana como no problemático”.

Tabla 2.11: Datos de los pacientes que asistieron a urgencias en el año 2009

<b>Hora</b>	<b>N° de pacientes</b>	<b>Grados</b>
00:00 am	146	0
01:00 am	19	15
02:00 am	15	30
03:00 am	17	45
04:00 am	10	60
05:00 am	12	75
06:00 am	16	90
07:00 am	22	105
08:00 am	39	120
09:00 am	42	135
10:00 am	48	150
11:00 am	53	165
12:00 pm	33	180
01:00 pm	23	195
02:00 pm	34	210
03:00 pm	39	225
04:00 pm	59	240
05:00 pm	39	255
06:00 pm	52	270
07:00 pm	70	285
08:00 pm	61	300
09:00 pm	44	315
10:00 pm	37	330
11:00 pm	28	345
<b>Total</b>	<b>958</b>	

Tabla 2.12: Parámetros estadísticos de asistencia a urgencias en el año 2009

<b>Urgencias</b>	
<b>Total</b>	958
<b>Dirección Media</b>	281,28°
<b>Moda</b>	0°
<b>Rango</b>	360°
<b>Varianza</b>	0,80

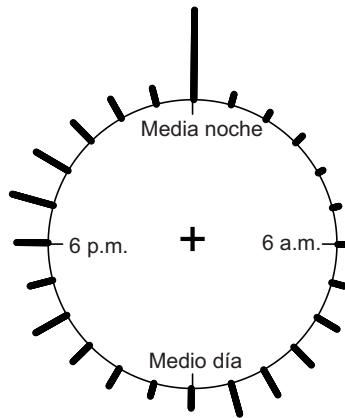


Figura 2.15: Gráfico de datos circulares en ramas para los casos de asistencia a urgencias

Así de esta manera nuestro propósito es comprobar si estos datos si se pueden ajustar a una distribución circular.

Haciendo uso del software R y su librería *CircStats* tenemos que el parámetro de concentración obtenido a partir de los datos es  $k = 0,3993$  y su media circular es  $\mu = 281,28^\circ$ . Teniendo en cuenta que el valor de  $k$  que es pequeño podemos concluir que los datos están dispersos con respecto a  $\mu$ , como se muestra en la Figura 2.15.

Ahora, después de obtener los parámetros  $k$  y  $\mu$  procedemos a graficar la distribución de los datos la cual llamamos *distribución empírica* contra la distribución von Mises, como se observa en la Figura 2.16. Así podemos inferir que para los valores de  $k$  (0,3993) y  $\mu$  (281,28°) los datos de urgencias de la Clínica Chicamocha S.A en el año 2009 se ajustan razonablemente a una distribución von Mises.

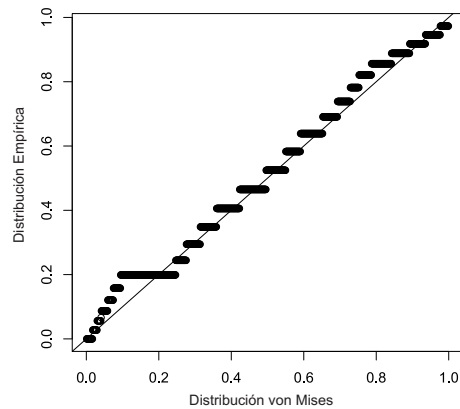


Figura 2.16: Ajuste gráfico de la asistencia a urgencias a una distribución von Mises.

### 2.2.1. Asistencia a urgencias según el género

En la base de datos aportada por el Observatorio de Salud Pública de Santander, Registros Individuales de Prestación de Servicios en Salud (RIPS), 2009; cuenta con las variables: hora y fecha de ingreso, edad y el género de las personas que asistieron a urgencias en el año 2009 a la Clínica Chicamocha S.A. La variable de estudio ahora es el género, la cual está dividida en: femenino (F) y masculino (M).

En la Tabla 2.13 encontramos el resumen estadístico de la entrada a urgencias según el género, claramente podemos observar que las mujeres tuvieron mayor asistencia que los hombres, con un total de 525 casos, donde en promedio la asistencia a urgencias se da a las 7:00 pm, con una dispersión de 0,80. En cambio los hombres asisten en promedio a urgencias a la 6:00 pm con una dispersión de 0,79. La hora más frecuente en donde los pacientes asisten a urgencias a la Clínica Chicamocha S.A es a las 00:00 am para los dos géneros.

En la Figura 2.17, se observa la similitud en el comportamiento de los

Tabla 2.13: Parámetros estadísticos de los casos de urgencias por género

	FEMENINO	MASCULINO
<b>Total</b>	525	433
<b>Dirección Media</b>	284,52	277,52
<b>Moda</b>	0°	0°
<b>Rango</b>	360°	360°
<b>Varianza</b>	0,80	0,79
$k$	0,3893	0,4206

datos para los dos géneros, además, se presentó asistencia a urgencia tanto de hombres como de mujeres durante todo el día, aunque son menos frecuentes los casos entre la 01:00 am y las 06:00 am.

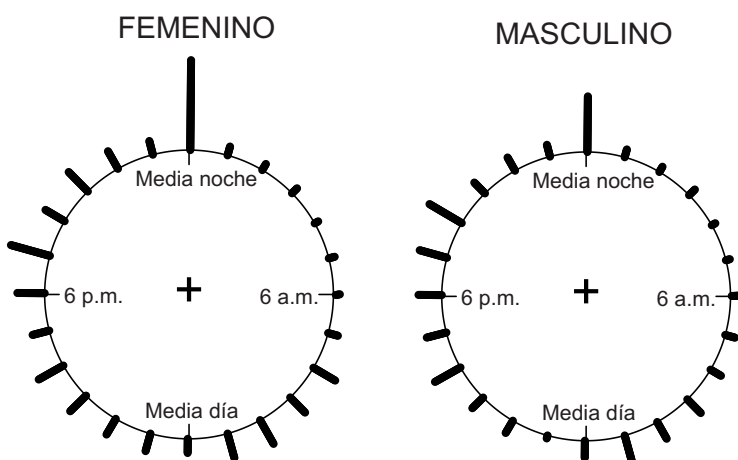


Figura 2.17: Gráfico de datos circulares en ramas para los casos de asistencia a urgencias por género.

Una característica importante la cual hay que resaltar, es que las asistencias a urgencias divididas por género tienen un comportamiento similar al de todos las asistencias a urgencias, de esta manera queremos ver si estos datos también se pueden ajustar a una distribución von Mises.

Haciendo uso del software R y su librería *CircStats* podemos observar en la Figura 2.18 que efectivamente las asistencias a urgencias a la Clínica

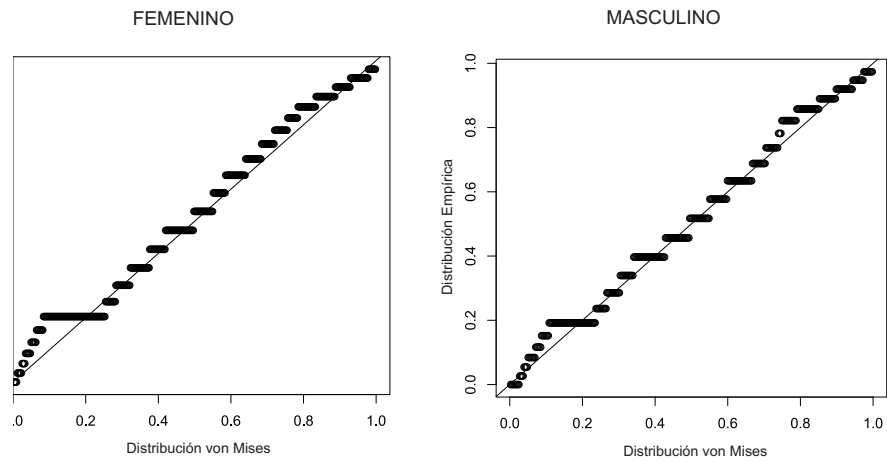


Figura 2.18: Ajuste gráfico de la asistencia a urgencias por género a una distribución von Mises.

Chicamocha S.A en el año 2009 tanto para hombres como para mujeres con  $(k = 0,4206$  y  $\mu = 277,52)$  y  $(k = 0,3893$  y  $\mu = 284,52)$  respectivamente, se ajustan razonablemente cada una a la distribución von Mises.

# Capítulo 3

## Conclusiones

Las técnicas de análisis de los datos circulares aplicadas al estudio de los casos de dengue en el AMB entre el período 2003 al 2008 demuestran que, la incidencia de esta enfermedad no presenta una dirección estable (es decir, no hay rango donde la enfermedad tenga una alta prevalencia), debido a factores climáticos como la temperatura y la precipitación las cuales fueron diferentes para cada año.

De acuerdo con lo anterior son muchas y muy variadas las herramientas utilizadas para el análisis de la incidencia del dengue, en esta ocasión se hizo el ajuste a una distribución von Mises, el cual se repitió para cada año y luego para todos los casos de dengue. Este último ajuste se considero razonable, lo cual quiere decir que los datos siguen una distribución von Mises con media circular  $\mu = 179,68^\circ$  y parámetro de concentración  $k = 0,4944$ . De esta manera podemos decir que los casos de dengue presentan una alta prevalencia entre finales de junio y principios de julio.

Teniendo en cuenta el período de tiempo en donde el mosquito *Aedes Aegypti* se desarrolla y el ajuste de los datos realizado, proponemos que las

campaññas y los programas de prevención para la enfermedad comiencen a principios del mes de junio, con el objetivo de disminuir el número de casos de dengue, ya que hoy en día esta enfermedad se considera como un problema de salud pública.

En el segundo ejemplo al realizar un análisis descriptivo circular y hacer un ajuste de las asistencias a urgencias a la Clínica Chicamocha S.A a una distribución von Mises, podemos afirmar que las personas llegan a urgencias a partir de las 7:00 pm. Dicha asistencia se puede considerar independiente del género (esto se pudo deducir ya que las asistencias tanto para hombres como para mujeres tuvieron el mismo comportamiento que todas las asistencias a urgencias), aunque el género femenino tiende en proporción a asistir más en ese período de tiempo, debido a la concentración que se presentó de 0.8, pero dicha proporción no es significativa ya que el género masculino también alcanzó una concentración de 0.79.

Acorde con las anteriores deducciones podemos afirmar que las técnicas de análisis de los datos circulares son de gran importancia en diferentes ciencias como se había nombrado en el primer capítulo, así mismo, se pudo comprobar que el modelamiento de este tipo de datos no depende del origen y el sentido de rotación, como se pudo ver con los datos de dengue, donde la dirección cero era en el Oriente y el sentido de rotación era contrario a las manecillas de reloj, en cambio, en las asistencias a urgencias la dirección cero era el Norte y el sentido de rotación era en forma horaria, es decir, en sentido a las manecillas del reloj.

Para terminar, podemos decir que las técnicas de análisis de datos circulares se pueden considerar trascendentales en la línea estándar de análisis estadísticos, aunque este tipo de datos han sido poco explorados (hay pocos

trabajos e investigaciones donde se aplican estas técnicas: Brunsdon y Corcoran, 2006; Gill y Hangartner, 2010; Muñoz, 2008; Polo, 2007; Oliveira, 2011). En esta trabajo encontramos una breve introducción y algunas aplicaciones de las técnicas más usadas de los datos circulares; para las personas que estén interesadas o quieran profundizar en este tema pueden ver la bibliografía, en donde hay artículos y libros que les pueden ayudar.

## Capítulo 4

# Ilustración de un ejemplo usando Software R

Todos los gráficos anteriores y los datos presentados en la tabla fueron realizados en el software estadístico R con la librería *CircStats*. El siguiente es un ejemplo en el cual nos dan los parámetros estadísticos y las figuras de los casos de dengue presentados en el año 2003.

```
> z<-c(100.38, 107.30, 114.23, 121.15, 121.15, 121.15, 121.15,
128.07, 128.07, 135, 135, 135, 135, 135, 135, 141.92, 141.92,
141.92, 141.92, 141.92, 148.84, 155.76, 162.69, 162.69, 162.69,
162.69, 162.69, 169.61, 169.61, 169.61, 176.53, 176.53, 183.46,
183.46, 183.46, 183.46, 183.46, 183.46, 183.46, 183.46, 183.46,
190.38, 190.38, 190.38, 197.30, 197.30, 197.30, 197.30, 197.30,
197.30, 197.30, 204.23, 204.23, 204.23, 204.23, 211.15, 211.15,
211.15, 218.07, 218.07, 218.07, 218.07, 218.07, 218.07, 225,
225,225, 225, 225, 225, 231.92, 231.92, 231.92, 231.92, 231.92,
238.84, 238.84, 238.84, 238.84, 238.84, 245.76, 245.76, 245.76,
245.76, 252.69, 252.69, 259.61, 259.61, 259.61, 259.61, 259.61,
259.61, 259.61, 266.53, 266.53, 273.46, 273.46, 273.46, 273.46,
273.46, 280.38, 280.38, 280.38, 280.38, 280.38, 287.3, 287.3,
```

```

287.3, 294.23, 294.23, 301.15, 301.15, 321.92, 321.92, 328.84,
335.76, 342.69)
> circ.plot(rad(z), stack=TRUE, bins=52, shrink=2, dotsep=10,
cex=0.7)
> rose.diag(rad(z), bins=52, prop=2.5, pts=FALSE, shrink=2,
cex=0.7)
> circ.mean(rad(z))
[1] -2.573481
> circ.disp(rad(z))
      n      r      rbar      var
1 117 69.6403 0.5952163 0.4047837
> circ.range(rad(z))
      range
1 4.229107
> deg(-2.573481)
[1] -147.4496
> pp.plot(rad(z), ref.line=TRUE)
      mu      kappa
1 3.709704 1.489646

```

# Capítulo 5

## Bibliografía

Brunsdon C. y Corcoran J., *Using circular statistics to analyse time patterns in crime incidence*, 2006.

Fisher N.I., *Statistics Analysis of Circular Data*. Cambridge University Press: Cambridge, 1995.

Gill J. y Hangartner D., *Circular Data in Political Science and How to Handle It*, 2010.

Gonzales S.P. y Velásquez E. M., *Modelo matemático de simulación para el control del Aedes Aegypti en epidemias de dengue clásico y dengue hemorrágico: bajo un enfoque sistémico*. Universidad Industrial de Santander, Tesis de pregrado, 1995.

Jammalamadaka S. R. y SenGupta A., *Topics in Circular Statistics*. World Scientific, Singapore, 2001.

Mardia K. V. y Jupp P. E., *Directional Statistics*. Academic Press, 2000.  
Ministerio de Protección Social. *Guía de atención clínica integral del paciente con dengue*, Bogotá, 2010.

Muñoz M. D., *Aportaciones al Estudio de la Anisotropía y Modelado Espacial de la Información*, Departamento de Informática y Automática, Universidad de Salamanca, Julio 2008.

Oliveira M., *Análisis Estadístico en Geografía Física*, Universidad de Santiago de Compostela, Julio 2011.

Polo M. E., *Gestión Integral del Error de Posición en las Bases de Datos Espaciales*, Universidad Extremadura, 2007.

Rodríguez M. L., *Análisis Estadístico de Datos Direccionales. Aplicaciones Medioambientales*, Departamento de Estadística e Investigación Operativa, Universidad de Santiago de Compostela, Julio 2011.