



UNIVERSIDAD AUTÓNOMA DE YUCATÁN

FACULTAD DE MATEMÁTICAS

TESIS:

**APLICACIÓN DE LA ESTADÍSTICA
CIRCULAR EN LA INCIDENCIA DE
DENGUE EN YUCATÁN**

que es presentada por el L.M. Miguel Fernando Uicab Perera para
obtener el grado de Maestro en Ciencias Matemáticas

Dirigida por:
Dr. Jorge Armando Argáez Sosa

Mérida, Yucatán, México, noviembre de 2015

Dedicado a todos aquellos que se sienten inmerecidos de las cosas buenas que acontecen en sus vidas. Porque si uno siente que no se merece tal o cual bendición, debe luchar siempre para merecerla. O bien, hacer todo lo que esté en sus manos para aportar con una bendición de igual o mayor magnitud que aquella que le fue otorgada.

Dedicado a todos aquellos que no valoran todo lo bueno que está presente en sus vidas. Al darse cuenta tarde, poco se puede hacer. Pero aún hay tiempo para darse cuenta temprano, porque siempre habrá algo bueno presente, siempre habrá algo que valorar y por lo cual dar gracias a la vida, a Dios.

Dedicado a todos aquellos que luchan por hacer siempre lo mejor, por luchar siempre para ser mejores personas, hijos, hermanos, amigos, parientes,..., porque aquellas personas escasean, y sin embargo aún existen. Aún existen personas buenas, que luchan por serlo en un mundo plagado de maldad, mentira, engaño, infidelidad, superficialidad, perversión, abuso, mediocridad, indiferencia, egoísmo, ingratitud,... Estas personas son las mejores, pues luchan por revertir muchos de estos grandes males que aquejan al mundo, y siempre sin pedir nada más a cambio que ser felices.

Dedicado a todos aquellos “extraños” que hacen cosas “raras”, fuera de lo común, fuera de lo que todos hacen, yendo a contracorriente, para tratar de hacer lo que ellos piensan, creen, sienten, intuyen,..., que es lo mejor. Aunque no los entiendan, aunque los tachen de locos, de complicados, de hipócritas,..., a pesar de tantas etiquetas, ahí están ellos, haciéndolo; haciendo lo que para ellos es lo mejor.

Dedicado a ti lector, que te has tomado el tiempo de leer estas dedicatorias, en las que un servidor pretende dar a entender que no todo está perdido, que siempre se puede rectificar, cambiar o mejorar. Que los milagros existen, pero es necesario, imprescindible, luchar para merecerlos. Gracias.

Agradecimientos

A mis papás, don Valentín y doña Irma, por ser de las personas más bondadosas que pueblan este pequeño gran mundo, y que con su bondad iluminan mi vida todos los días. Muchas gracias por alentarme a dar siempre la mejor versión de mí. Ellos, desde hace mucho, ya se han ganado un lugar en el cielo; lucharé por tener asegurado un “cahito” de cielo, un “lugarcito” junta al de ellos. A ellos, les ofrezco todo de mí, mi amor y mi cariño, todo, y aún así no me alcanzará la vida para devolverles todo lo que han hecho por mí,....., sin embargo, haré el intento.

A mi hermanitos, Jasmín y Luis, por creer que soy todo un “genio”, que lo sé todo, y recurrir a mí sin pensarlo mucho, y confiar en que lo haré bien (sin duda, ellos, lo harán mucho mejor que yo en el futuro). En esencia, son buenas personas, así que les esperan mil y un bendiciones en el camino. Siempre podrán contar conmigo, y, si me lo permiten, siempre estaré junto a ellos.

A mis amigos, los de verdad, a los que estuvieron y ya no están, y los que se quedaron y aún están, por creer en mí. Sin duda, mencionar la aportación de cada uno de ellos me llevaría muchísimas páginas, así que confío en que ellos sabrán a quiénes tengo en mente; confío en que ellos ya sabrán el porqué les tengo tanto afecto.

Al Dr. Jorge Argáez Sosa, por ser mi asesor, por su paciencia, por sus inmerecidos halagos, por sus sabias sugerencias, por creer en mí, y confiar en que puedo alcanzar aún más. De la mano de él, de su guía, este trabajo a llegado al “mejor puerto posible”. Le tendré afecto, admiración y respeto profundos siempre.

A mi familia, mis tíos y tías, primos y primas, por sentirse orgullosos de mí, por no dudar de mí, y creer en que puedo lograr mucho más. A ellos, mi disposición y ayuda siempre. Para ellos, mil bendiciones.

Al Dr. José Luis Batún, por sus consejos y guía durante la realización de este trabajo, por su paciencia, por creer en que podría hacerlo mejor. Mi más profundo respeto, admiración y afecto.

Al MCM Salvador Medina Peralta, por sus valiosas observaciones, comentarios y aportes para con este trabajo. El también tiene mi más profundo respeto y admiración.

A la Facultad de Matemáticas de la UADY, por permitirme crecer académicamente durante mi recorrer por sus pasillos, y por brindarme el apoyo oportuno para poder estar donde estoy ahora. Gracias a los profesores por aprender de ellos más que teoremas y demostraciones. Gracias a mis amigas y amigos pertenecientes al personal administrativo de la Facultad; como buenas personas que son, se han portado muy bien conmigo.

A Dios, por las experiencias vividas durante mi estancia en la Facultad. Por permitirme aprender de aquellas experiencias y crecer. Por conocer a personas inolvidables y valiosas, independientemente de las diversas circunstancias. Por darnos a mí y a mi familia una nueva oportunidad de hacerlo bien. Por llenarme de mil bendiciones inmerecidas. Por permitirme ser como soy. Por ser lo que soy ahora.

¡Gracias! ¡Muchas gracias a todos!

Índice general

Introducción	1
1. Datos circulares	15
1.1. Breve historia del estudio de datos direccionales	15
1.2. Ejemplos de datos direccionales	16
1.3. Datos circulares	17
1.4. Periodicidad de los datos circulares	18
1.5. Representación gráfica de datos circulares	19
1.5.1. Diagrama Circular	19
1.5.2. Diagrama de Rosa	20
1.5.3. Histogramas lineales	21
2. Estadísticos descriptivos circulares	25
2.1. Momentos trigonométricos muestrales	26
2.2. Medidas de Localización	29
2.2.1. Dirección Media Muestral	29
2.2.2. Dirección Mediana Muestral	30
2.3. Medidas de Concentración y Dispersión	31
2.3.1. Longitud Media Resultante Muestral	31
2.3.2. Varianza Muestral Circular y Desviación Estándar	31
2.3.3. Otras Medidas Muestrales de Dispersión	32
2.4. Medidas de Sesgo y Curtosis	33
2.5. Correcciones para datos agrupados	34
3. Distribuciones de probabilidad circulares	37
3.1. Teoría distribucional circular	37
3.1.1. Funciones de distribución circular y de densidad de probabilidad	38
3.1.2. Función característica circular y Momentos Trigonométricos.	40
3.1.3. Medidas poblacionales básicas	42

3.1.4.	Distribuciones simétricas	44
3.2.	Enfoques generales para la generación de distribuciones circulares	45
3.3.	Modelos circulares clásicos	46
3.3.1.	Distribución Circular Uniforme Continua	46
3.3.2.	Distribución Cardioide	47
3.3.3.	Distribución Potencia del Coseno de Cartwright	48
3.3.4.	Distribución Cauchy Envolvente	49
3.3.5.	Distribución Normal Envolvente	50
3.3.6.	Distribución von Mises	51
3.4.	Familia de distribuciones simétricas Jones-Pewsey	55
3.4.1.	Estimación de parámetros vía máxima verosimilitud	57
3.5.	Distribuciones Inversas Batschelet	58
3.5.1.	Familias sesgadas	58
3.5.2.	Familias simétricas más amplias	60
3.5.3.	Familias simétricas-sesgadas	62
3.5.4.	Estimación máxima verosimilitud	62
3.5.5.	Distribución Inversa Batschelet con base von Mises	63
4.	Aproximaciones para muestras grandes, pruebas de clasificación, bondad de ajuste y estimación insesgada	65
4.1.	Aproximaciones para muestras grandes	65
4.1.1.	Distribución conjunta para a_1 y b_1	65
4.2.	Inferencia básica para una muestra simple	67
4.2.1.	Pruebas de uniformidad	67
4.2.2.	Pruebas de bondad de ajuste	69
4.2.3.	Prueba de simetría reflexiva para muestras grandes	70
4.3.	Inferencia para estadísticos descriptivos circulares	73
4.3.1.	Estimación puntual insesgada (bias-corrected)	73
4.3.2.	Intervalos de confianza para los estimadores insesgados	74
5.	Análisis y modelación de la incidencia de dengue utilizando estadística circular	75
5.1.	Los datos	76
5.2.	Establecimiento de los períodos de estudio	76
5.3.	Consideraciones con respecto al tiempo	80
5.4.	Consideraciones con respecto a la base de datos	81
5.5.	La incidencia de dengue como datos circulares	82
5.6.	Uso del diagrama circular en un período de estudio mayo - abril anual	84
5.7.	Modelos circulares elegidos y herramientas utilizadas	85
5.8.	Período 1: Mayo 2009 - Abril 2010	87
5.9.	Período 2: Mayo 2010 - Abril 2011	93

5.10. Período 3: Mayo 2011 - Abril 2012	99
5.11. Período 4: Mayo 2012 - Abril 2013	104
5.12. Período 5: Mayo 2013 - Abril 2014	109
5.13. Período 6: Mayo 2014 - Abril 2015	114
6. Comparación entre los períodos de estudio	121
6.1. Comparación entre gráficos circulares	121
6.2. Comparación muestral	124
6.2.1. Sobre la dirección media muestral $\bar{\theta}$	124
6.2.2. Sobre la longitud media resultante muestral \bar{R}	126
6.2.3. Sobre la medida de sesgo \bar{b}_2	127
6.2.4. Sobre la medida de curtosis \bar{a}_2	129
6.3. Comparación entre los modelos ajustados	130
6.4. Resumen de los casos de dengue en dos períodos de estudio generales	134
6.4.1. Resumen 1: Mayo Año Impar - Abril Año Par	136
6.4.2. Período Resumen 2: Mayo Año Par - Abril Año Impar	141
6.5. Comparación entre los períodos de estudio y los períodos generales .	146
7. Discusión y conclusiones	151
7.1. Sobre la utilización de la estadística circular como herramienta de estudio	151
7.2. Sobre los períodos de estudio	155
7.3. Sobre la temporada de lluvias	164
7.4. Conclusiones generales	170
7.5. Trabajos futuros	172
A. Tablas	175
B. Gráficos de comparación entre los períodos de estudio	183
C. Consideraciones con respecto a la base de datos	193
Referencias	197

Introducción

¿Qué es el dengue?

El dengue es una enfermedad vírica que se transmite a los seres humanos por la picadura de mosquitos hembra infectadas, principalmente, de la especie *Aedes aegypti* y, en menor grado, de *Aedes albopictus*.

La infección causa síntomas gripales y puede evolucionar, en ocasiones, a un cuadro potencialmente mortal llamada dengue grave (aunque aún es muy conocido como dengue hemorrágico).

La enfermedad se presenta en mayor medida en lugares con climas tropicales y subtropicales del planeta, sobre todo en zonas urbanas y suburbanas. Las variaciones locales de riesgo de contraer la enfermedad dependen, en gran medida, de las precipitaciones, la temperatura y la urbanización rápida sin planificar.

Según la Organización Mundial de la Salud, en las últimas décadas ha aumentado enormemente la incidencia de dengue en el mundo. El número real de casos de dengue está insuficientemente notificado y muchos casos están mal clasificados.

Aedes aegypti y transmisión del dengue

El vector principal del dengue es el mosquito hembra *Aedes aegypti*. Las personas infectadas son los portadores y multiplicadores principales del virus, y los mosquitos se infectan al picarlas. Tras la aparición de los primeros síntomas, las personas infectadas con el virus pueden transmitir la infección (durante 4 ó 5 días; 12 días como máximo) a los mosquitos *Aedes*¹. Un mosquito infectado puede transmitir el agente patógeno durante toda la vida, la cual, en condiciones naturales, dura

¹OMS en línea: Dengue y dengue hemorrágico. <http://www.who.int/mediacentre/factsheets/fs117/es/>

en promedio de entre 15 a 30 días ². Su ciclo para poner huevos es de aproximadamente cada tres días.

El mosquito *Aedes Aegypti* vive en hábitats urbanos y se reproduce principalmente en recipientes artificiales. A diferencia de otros mosquitos, este se alimenta durante el día; los períodos en que se intensifican las picaduras son el principio de la mañana y el atardecer, antes de que oscurezca. En cada período de alimentación, el mosquito hembra pica a muchas personas ¹.

Hay una convivencia estrecha del vector con el ser humano, ya que este le propicia sus principales criaderos para su reproducción y posteriormente lo alimenta con su sangre. Debido a lo anterior, y al no existir una cura reconocida a nivel internacional, las acciones de control de la enfermedad se dirigen específicamente a las de erradicación de vector (lucha antivectorial).

Las acciones antivectoriales se efectúan teniendo en cuenta dos fases del desarrollo del vector: la fase aérea y la fase acuática. En la fase aérea el mosquito hembra pica para alimentarse a una persona infectada en estado febril, desarrollando la enfermedad en su organismo; posteriormente, para continuar alimentándose, pica a otros individuos sanos infectándolos, continuando el ciclo de la misma. En su fase acuática el vector deposita sus huevecillos en un depósito con agua, de allí se desarrollan en unos pocos días en un ambiente propicio de humedad y calor, en el que logran aparecer sus larvas, que en 4 ó 5 días se transforman en pupas que en pocas horas se transforman en mosquitos adultos ³.

Características de la enfermedad

Según la OMS, el dengue es una enfermedad de tipo gripal que afecta a bebés, niños pequeños y adultos, pero raras veces resulta ser mortal.

Se debe sospechar que una persona padece dengue cuando, junto con una fiebre alta (40° C), se presentan síntomas tales como: dolor de cabeza muy intenso, dolor detrás de los globos oculares, dolores musculares y articulares, náuseas, vómitos, agrandamiento de ganglios linfáticos o salpullido. Los síntomas se presentan al cabo de un periodo de incubación de 4 a 10 días después de la picadura de un mosquito infectado y por lo común duran entre 2 y 7 días.

El dengue grave es una complicación potencialmente mortal porque cursa con

²<http://www.monografias.com/trabajos29/dengue/dengue.shtml>

³Secretaría de Salud de Yucatán: El Dengue en Yucatán. <http://salud.yucatan.gob.mx/programas/el-dengue/>

extravasación de plasma, acumulación de líquidos, dificultad respiratoria, hemorragias graves o falla orgánica. Los signos que advierten de esta complicación se presentan entre 3 y 7 días después de los primeros síntomas, acompañados de un descenso de la temperatura corporal (menos de 38° C), y son los siguientes: dolor abdominal intenso, vómitos persistentes, respiración acelerada, hemorragias de las encías, fatiga, inquietud y presencia de sangre en el vómito. Las siguientes 24 a 48 horas de la etapa crítica pueden ser letales.

Se conocen cuatro serotipos distintos, pero estrechamente emparentados, del virus: DEN-1, DEN-2, DEN-3 y DEN-4. Cuando una persona se recupera de la infección adquiere inmunidad de por vida contra el serotipo en particular. Sin embargo, la inmunidad cruzada a los otros serotipos es parcial y temporal. Las infecciones posteriores causadas por otros serotipos aumentan el riesgo de padecer el dengue grave ¹.

Prevención, tratamiento y lucha antivectorial

Debido a que la única manera de enfermarse es por la picadura del mosquito hembra infectado, la enfermedad puede prevenirse impidiendo ser picados por estos insectos, cuestión sumamente difícil en áreas tropicales húmedas .

No hay tratamiento específico del dengue ni del dengue grave, pero la detección oportuna y el acceso a la asistencia médica adecuada reducen las tasas de mortalidad por de debajo del 1 % ¹.

La prevención y control del dengue dependen exclusivamente de las medidas eficaces de lucha antivectorial.

La Secretaría de Salud del Estado de Yucatán reconoce diversas maneras de combatir el dengue, a través de la NORMA Oficial Mexicana NOM-032-SSA2-2002, para la vigilancia epidemiológica, prevención y control de enfermedades transmitidas por vector: la nebulización, la eliminación de criaderos, la abatización y la participación.

Sobre la nebulización, existe la creencia generalizada de que es la solución para terminar con los mosquitos, sin embargo solo brinda alivio temporal y pasajero para las molestias causadas por ellos. Es la acción mas cara y menos efectiva en la lucha contra el padecimiento, pero produce satisfacción a la sociedad. La nebulización para combatir el mosco adulto debe realizarse, ante la presencia del vector infectado (casos de dengue confirmados en un área) y en sitios de concentración de población por diversos eventos, pero es mejor evitarla para mantener el medio ambiente

limpio de insecticidas potencialmente nocivos a largo plazo para la naturaleza y el ser humano.

La eliminación de criaderos (llamada antes “descacharrización”) es utilizada en muchos lugares en donde el sistema de recolección de basura no es efectivo o no se cuenta con el mismo. Cuando se realiza, este debe ser selectivo para los criaderos verdaderos: llantas, botes, botellas, latas, corcholatas, etcétera. Lo anterior debe llevarse a cabo junto con otras acciones complementarias para otros criaderos no desechables dentro de los patios de las viviendas o incluso en su interior, que deben taparse, “embrocarse” o lavarse periódicamente (tinacos, pilas, cubetas útiles, bebederos de animales, etcétera).

En lo que se refiere a la abatización (polvo insecticida que dejan los brigadistas en las pilas y otros recipientes) este no es dañino para la salud humana ni de los animales. Es decir, es una de las principales acciones que se realizan en el interior de los domicilios y se efectúa únicamente a aquellos depósitos que contienen agua la mayor parte del tiempo y que no pueden ser eliminados (pilas, bebederos, tinacos, etcétera).

La mejor manera de combatir al vector es logrando que la comunidad haga conciencia acerca de que la enfermedad es un problema que generalmente se origina en el interior de los domicilios, en los sitios donde se acumula agua y que son los sitios preferidos para la ovipostura de las hembras de mosquito. Manteniendo el patio de las casas limpio, no solo se logra combatir el dengue, sino que se evita una gran cantidad de enfermedades que la acumulación de basura propicia ².

Panorama sobre el estudio del dengue

El dengue se ha abordado utilizando múltiples herramientas de modelación matemática, que incluyen tanto ecuaciones diferenciales como estadística. En los trabajos en los que se ha utilizado estadística para el estudio del dengue, con frecuencia se modela su incidencia, es decir, el número de casos nuevos de la enfermedad que aparecen o que se reportan en cierto período de tiempo determinado.

La Dirección General de Epidemiología, en sus reportes semanales sobre la incidencia de dengue, utiliza modelos no probabilísticos de series de tiempo para estimar el número de casos nuevos que se esperarían para una semana epidemiológica próxima.

La OMS, en su portal en línea, presenta un resultado en el cual se estima se producen 390 millones de infecciones por dengue cada año (intervalo creíble del 95 %:

284 a 528 millones), de los cuales 96 millones (67 a 136 millones) se manifiestan clínicamente, cualquiera que sea la gravedad de la enfermedad (Bhatt *et al.*, 2013).

Se ha llegado a la estimación anterior utilizando árboles de regresión impulsados, los cuales difieren fundamentalmente de los métodos tradicionales de regresión porque producen un modelo único (“el mejor”), que es generado mediante la combinación un gran número de modelos de árboles de regresión, y con esto se logra optimizar el desempeño predictivo de los modelos (Elith *et al.*, 2008). Los árboles de regresión impulsados utilizan dos algoritmos: árboles de regresión que son un grupo de modelos de arboles que clasifican y aplican regresión (árboles de decisión), y el “impulso” que construye y combina una colección de modelos (Friedman, 2002).

También se han desarrollado diversos modelos para la distribución global futura de dengue. Hales *et al.* (2002) muestran un estudio que incluye reportes de dengue de 1975 a 1996, en el que se usa regresión logística para modelar la incidencia de dengue con base en varias variables meteorológicas relacionadas con la temperatura, la precipitación y la humedad, así como las interacciones entre las variables como covariables de modelado. Luego, se selecciona un modelo con un único parámetro dependiente (presión media anual de vapor) al comparar varios y utilizando el criterio de parsimonia. Al final, extrapolan este modelo a años futuros usando proyecciones de los cambios en la presión de vapor media anual para mapear el riesgo de dengue mundial proyectado. El estudio incluyó más tarde reportes de 1975 a 2009 y usaron modelos aditivos generalizados para incluir el producto interno bruto (PIB) per capita, además de variables meteorológicas de modelos y proyecciones de referencia, y usando probabilidad de corte para luego mapear la presencia o ausencia de dengue en una ubicación (Åström *et al.*, 2012).

Jaimes Jaimes y Pineda Garavito (2012) utilizan una herramienta llamada *estadística circular* para ofrecer un análisis descriptivo y un modelo de probabilidad circular para la distribución de la incidencia de dengue en el área metropolitana de Bucaramanga (Colombia), en un año en particular. El modelo que utilizan es uno clásico y reflexivamente simétrico, llamado la distribución von Mises. Ellos discuten, a través de gráficos de probabilidad-probabilidad, qué tan razonable es el ajuste del modelo propuesto a la distribución real de los casos de nuevos de dengue.

Historia del dengue en México

La historia del dengue se liga al de la fiebre amarilla en el país. Si bien, los primeros reportes de dengue en México datan la década de los 40's, estos se interrumpieron en 1946 con el inicio de la campaña de erradicación de la fiebre amarilla, que consistió en la eliminación de su mosquito transmisor, el *Aedes egypti*, que es el mismo del dengue.

Los resultados positivos de esta campaña se certificaron en 1963. De hecho, en ese año se tienen los siguientes hechos anecdóticos. El 6 de agosto de 1963, las autoridades mexicanas entregaron a las autoridades sanitarias de Estados Unidos, dentro de un cubo de plástico, la última pareja de *Aedes aegypti* capturados en México. El primero de septiembre de 1963, el Presidente Adolfo López Mateos informó al país, a través de las Cámaras legislativas, que México había quedado libre del vector de la fiebre amarilla urbana (Góngora-Biachi, 2004).

Sin embargo, en 1969 el vector *Aedes Aegypti* reapareció pero esta vez transmitiendo otro virus, el del dengue. Esta enfermedad resurgió más notoriamente en 1979, como consecuencia de la disminución de las acciones para la erradicación del *Aedes aegypti*, a los cambios demográficos y a otros factores de riesgo, como han sido las migraciones de poblaciones del campo a las ciudades, que se concentraron en áreas semiurbanas o conurbadas, con el establecimiento de viviendas precarias, hacinamiento, deficientes servicios públicos de agua y recolección de basuras, así como de importantes brotes de dengue en América del Sur, por virus importados de otros continentes. Entrando así al país, por la circulación habitual de personas, a través de Chiapas y extendiéndose hacia todo el país con predominio en las regiones de las Costas del Atlántico y Pacífico en donde prácticamente ha permanecido hasta la presente fecha en forma endémica ².

El sur del país fue donde se reportaron los primeros casos de dengue asociados con DEN-1 en 1978 que inicialmente se confundieron con un brote de rubeola. Las epidemias se extendieron por todo el territorio asociados al mismo serotipo hasta el aislamiento de DEN-2 en 1981 y del DEN-4 en 1982. Los primeros casos de fiebre hemorrágica por dengue FHD (dengue grave) se reportaron en 1984. DEN-3 empezó a circular en 1995. Desde entonces, la transmisión involucró una combinación de serotipos con diferentes efectos en la severidad de la enfermedad. Todos los serotipos circularon de forma simultánea en el país y desde 1982 se han reportado brotes con más de un serotipo circulante. La transmisión de los cuatro serotipos se detectó por primera vez en 1994 (Gómez y San Martín, 2011).

Entre 2001 y 2002, la enfermedad se empezó a contener mejor debido a cambios en el programa de prevención y control y la vigilancia epidemiológica, en cuanto a la

oportunidad de las acciones, la confirmación o descarte de los casos presentados y al involucramiento de los municipios en términos de tratarse de un problema con grandes rasgos de mal saneamiento ambiental derivado de las carencias o rezagos de servicios de recoja de basuras. A partir del 2000, se ha comenzado a tener otra situación epidemiológica, al encontrar una proporción cada vez mayor de casos con manifestaciones hemorrágicas y otros de mayor severidad y en grupos de edad más jóvenes, con relación al número de casos de dengue clásico, como ocurre en otros países en donde existe la endemidad del padecimiento ².

En cuanto al estado de Yucatán, al inicio del 2015 se ha registrado una disminución realmente importante de los casos registrados de dengue. De hecho es la más importante de los últimos años ⁴. De acuerdo con el registro de la *Dirección General de Epidemiología* (DGE), organismo dependiente de la Secretaría de Salud (SSA) del gobierno mexicano, de enero a abril de 2015 se habían registrado 93 casos de dengue, mientras que en los tres primeros meses de 2014 se registraron 322, y 258 en 2013. Lo anterior representa una reducción del 71.12 % respecto de 2014, y del 63.95 % respecto de 2013.

Importancia del estudio del dengue

Si bien en los últimos años, la distribución de los casos de dengue ha sido ampliamente estudiado con numerosas técnicas de modelación y análisis, no deja de ser un tema relevante debido, entre otras cosas, al aumento en las últimas décadas de la incidencia de dengue en el mundo, a que en muchos países de Asia y Latinoamérica el dengue grave es causa de enfermedad y muerte en los niños, y a la hiperendemicidad de los múltiples serotipos del virus en muchas países. Sin duda, el dengue ha tenido, y sigue teniendo, una repercusión importante en la salud humana en las economías nacionales y mundial ¹.

La precipitación pluvial es uno de los principales factores de incidencia de dengue, más cuando hay un período de tiempo establecido para la temporada de lluvias, además de que este tiene un comportamiento aparentemente regular.

La periodicidad de la temporada de lluvias repercutirá a que los casos nuevos de dengue aumenten o disminuyan en ciertos intervalos de tiempo. Sin embargo, estos intervalos podrían variar debido a los cambios en los niveles de precipitación pluvial, y debido a otros factores.

Saber sobre los tiempos en que se esperaría se presenten un gran número de casos nuevos de dengue, implicaría, a su vez, a considerar los tiempos en que se deberían

⁴<http://sipse.com/milenio/yucatan-baja-cifra-casos-dengue-146535.html>

llevar a cabo las acciones de lucha antivectorial de manera más eficaz y contundente. Además, el destinamiento de los recursos necesarios (tanto económicos como materiales y humanos) para la lucha antivectorial podrían ser optimizado al tener una mejor idea sobre los tiempos con mayor probabilidad de incidencia y los de menor probabilidad, hasta repercutir en un posible ahorro en el presupuesto público.

La incidencia de dengue en Yucatán de 2009 a 2014

La *incidencia* de una enfermedad es el número de casos nuevos que se registran de una enfermedad en cierto lugar y período determinados. En México, la *Dirección General de Epidemiología* (DGE) emite reportes semanales acerca de diversas enfermedades que afectan a la población del territorio nacional, entre ellas, por ejemplo, el dengue. De aquellos reportes, se puede derivar la incidencia semanal de la enfermedad, es decir, el número de casos nuevos que se registran de dicha enfermedad en una *semana epidemiológica* específica de un año determinado.

La Tabla 1 muestra los totales de los casos de dengue registrados anualmente en el rango de 2009 a 2014, según los reportes emitidos por la DGE. La Figura 1 exhibe que las cantidades totales varían. En estos seis años se puede observar que el valor registrado para un año impar es relativamente más grande que el registrado para el año inmediato posterior, el cual es necesariamente par. En la Figura se exhibe un repunte considerable del total de los casos registrados en el 2011, con respecto al 2009 y 2010, para luego dar paso a un decrecimiento en los valores totales que se ha mantenido hasta el 2014. Hasta el año 2014 respecto al 2011, ha habido una reducción de casi el 82% de los casos reportados.

Año	Total
2009	3220
2010	2489
2011	6032
2012	5654
2013	2758
2014	1059

Tabla 1: Total de casos registrados por años, según los reportes de la DGE.

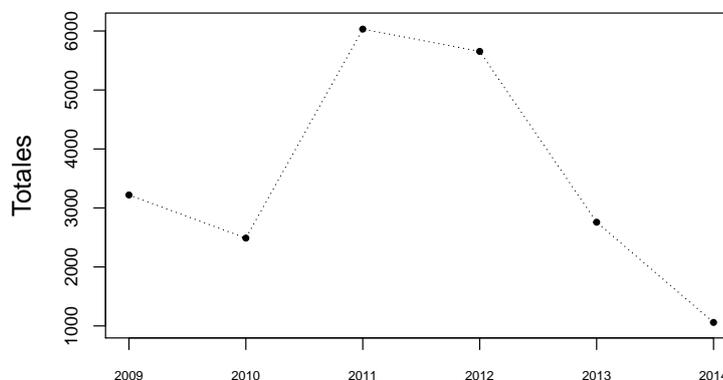


Figura 1: Total de casos registrados por años, según los reportes de la DGE.

Describiendo los aumentos y decrementos de los totales de incidencia, paso a paso, se puede decir que del 2009 a 2010 hubo una reducción de casi el 23 % de los casos registrados; del 2010 al 2011 hubo un aumento de casi el 142 %; del 2011 al 2012 hubo una reducción de casi el 6 %; del 2012 al 2013 hubo una reducción de casi el 52 %; y del 2013 al 2014 de casi el 62 %.

La Figura B.9 del Apéndice B muestra los histogramas lineales de casos registrados semanalmente en cada año, del rango de 2009 a 2014. Al observar individualmente estos histogramas, llama la atención que en cada uno de los años impares 2009, 2011 y 2013 se aprecia lo que pareciera ser un tramo largo de frecuencias relativamente bajas (ver Figuras B.91, B.93) y B.95, respectivamente); en contraste, en cada uno de los años pares 2010, 2012 y 2014 se aprecia un tramo corto de frecuencias relativamente bajas (ver Figuras B.92, B.94 y B.96, respectivamente). A fin de tener cierta idea sobre el comportamiento de la incidencia de dengue en cada uno de estos tramos largos y cortos de frecuencias bajas, uno podría utilizar, en primera instancia, la función de distribución empírica calculada en cada uno de los años del rango de 2009 a 2014. La función de distribución empírica ofrece la frecuencia relativa acumulada de una muestra.

En el 2009, el tramo largo de frecuencias relativamente bajas abarca de la semana 1 a la 32 (1^a de agosto), en el 2011 abarca de la semana 1 a la 36 (transición entre agosto - septiembre), y en el 2013 abarca de la semana 1 a la 25 (2^a de julio). En el 2009, el tramo largo correspondiente abarca aproximadamente 7 meses, y el porcentaje acumulado de mayo a julio fue de casi el 8 % del total de casos registrados en este año. Para el año 2011, el tramo largo de frecuencias bajas abarca casi 8 meses, de enero a agosto, y aquí el porcentaje acumulado aproximado fue de tan sólo el 4 % del total de los casos de dengue. Para el 2013, el tramo largo abarca casi 6 meses, de enero a junio, y el porcentaje acumulado en esos meses fue

de casi el 15 % del total de incidencia.

En el 2010, el tramo corto de frecuencias relativamente bajas abarca de la semana 12 a la 25 (3^a de marzo - 3^a de junio), en el 2012 abarca de la semana 12 a la 26 (3^a de marzo - transición junio-julio), y en el 2014 abarca de la semana 13 a la 25 (transición marzo-abril - 3^a de junio). Estos tramos abarcan aproximadamente 3 meses, principalmente de abril a junio. Para el año 2010, el porcentaje acumulado en esos meses es de casi el 5 %, en el 2012 es de casi el 7 %, y en el 2014 es de casi el 9 %, del total de los casos en esos años.

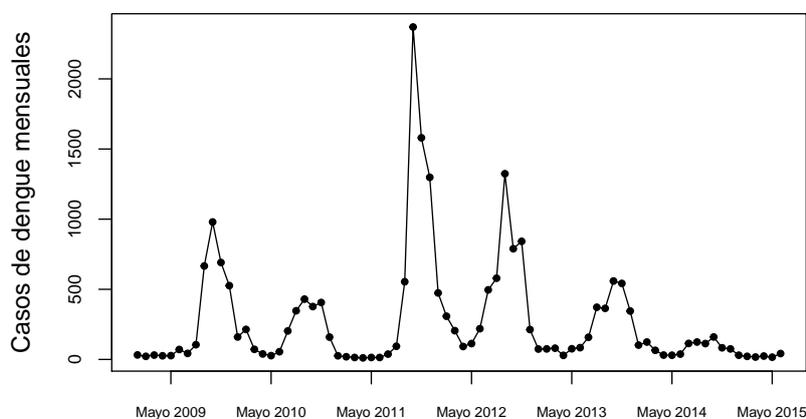


Figura 2: Valores de incidencia mensuales que se dan entre enero de 2009 y junio de 2015.

La Figura 2 muestra un panorama gráfico de los casos de dengue que se han registrado de manera mensual, durante el 2009 hasta los primeros 6 meses del 2015. Se puede ver que los tramos de frecuencias relativamente bajas alternan entre uno largo (en los años impares) y uno corto (en los años pares). Se aprecia que el mes de mayo funge como un mes intermedio, es decir, un mes próximo a las semanas en las que se alcanza la parte mitad de estos tramos de bajas frecuencias. También, se observa que el lapso que abarca del mes de mayo de un año al mes de abril del año siguiente, encasilla un grupo de frecuencias relativamente grandes. Se puede notar que estos grupos de grandes frecuencias no tiene la misma forma ni contemplan la misma cantidad de casos de dengue, entre sí.

Estudio del dengue con el enfoque de la estadística circular

La estadística circular es una herramienta que toma en cuenta la naturaleza periódica de los llamados datos circulares, los cuales incluyen tanto mediciones angulares, como datos tomados de un reloj de 24 o 12 horas, o que pertenecen a fenómenos periódicos a través del tiempo.

La incidencia de dengue en Yucatán es un fenómeno periódico, pues su ocurrencia depende, en gran medida, de la temporada de lluvias presente cada año en el estado. De hecho, si se observan los valores de precipitación mensual (medidos en mm) que se tiene para el estado de Yucatán de enero de 2009 a junio de 2015 (ver Figura 3) ⁵, y teniendo en cuenta que el mes de mayo es el inicio de la temporada de lluvias a nivel nacional ⁶, se puede notar la presencia de ciclos, con periodicidad de 12 meses, en el comportamiento de la precipitación.

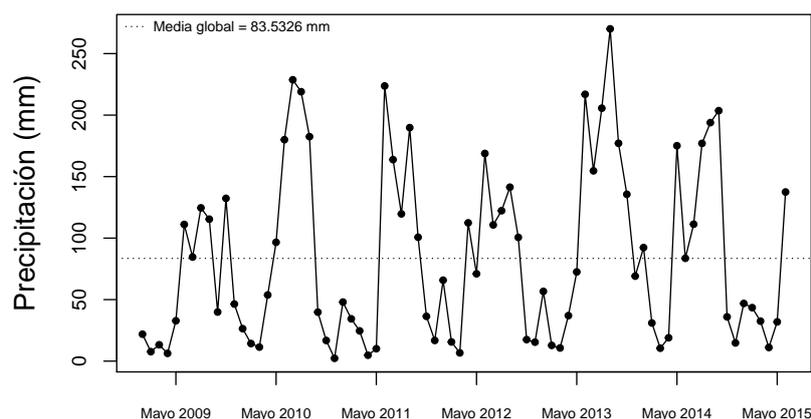


Figura 3: Valores de precipitación pluvial (mm) mensual que se presentan entre enero de 2009 y junio de 2015. La media global que se obtiene en este lapso es de 83.5326 mm.

De la Figura 3, en cada ciclo se presenta un lapso de grandes valores de precipitación (aquellos valores que están arriba de la media histórica), y estas grandes precipitaciones son las que influyen en el aumento de los casos nuevos de dengue en un tiempo determinado; de este modo, se puede decir que se tendrán picos de incidencia cada año. Por ese comportamiento cíclico en la incidencia de dengue en

⁵CONAGUA. Temperatura y lluvia: http://smn.cna.gob.mx/index.php?option=com_content&view=article&id=12&Itemid=77

⁶CONAGUA. ¿Qué es un ciclón?: http://smn.cna.gob.mx/index.php?option=com_content&view=article&id=39&Itemid=47.

Yucatán, los casos nuevos de dengue pueden ser estudiados como datos circulares. Siendo así, el presente trabajo tiene como objetivo modelar la incidencia de dengue en el estado de Yucatán utilizando estadística circular.

Para tal fin, este trabajo se ha organizado de la manera siguiente. Los primeros cinco capítulos proporcionan las herramientas gráficas y teóricas que necesitaremos para el desarrollo de este estudio. En el Capítulo 1 se presenta la definición de datos circulares, una breve historia sobre su estudio y ejemplos que surgen en diferentes ramas de la ciencia. En el Capítulo 2 se presentan las herramientas utilizadas para el análisis gráfico de datos circulares. En el Capítulo 3 se introducen los estadísticos descriptivos circulares más comúnmente utilizados para resumir las principales características de este tipo de datos. En el Capítulo 4 se ofrece una introducción a la teoría distribucional que sustenta la estadística circular; se presentan modelos circulares clásicos y recientes reportados en la literatura de estadística circular, entre las que se encuentran la distribución von Mises (o Normal Circular), la familia de distribuciones simétricas Jones-Pewsey y la familia de distribuciones flexibles Inversa Batschelet. En el Capítulo 5 se ofrece la herramienta básica para el análisis inferencial de los datos circulares, así como métodos de estimación puntal y por intervalos para los principales estadísticos descriptivos circulares.

En el Capítulo 6 se presentan los resultados que se han obtenido al aplicar estadística circular en el análisis descriptivo y modelado de los casos nuevos de dengue en cierto período de tiempo (incidencia). Para tal fin, se han considerado seis períodos anuales de estudio que se establecen entre los años 2009 y 2015. En el capítulo 7 se da un comparación entre los resultados obtenidos en los períodos de estudio considerados, y en el Capítulo 8 se discuten los resultados más importantes que se han hallado. Finalmente, en el Capítulo 9 se ofrecen las conclusiones generales de este trabajo de investigación.

Cabe señalar que para el contenido teórico de este trabajo se han tomado como base los libros *Topics in Circular Statistics* (Jammalamadaka y Sengupta, 2001), *Directional Statistics* (Mardia y Jupp, 1999), pero sobre todo se ha tenido muy en cuenta el contenido de *Circular Statistics in R* (Pewsey *et al.*, 2013) el cual ofrece un magnífico resumen acerca de los contenidos sobre datos circulares de los dos primeros libros, además de que ofrece las herramientas computacionales utilizadas aquí; Pewsey *et al.* (2013) brindan una lectura muy amena y presentan diversos ejemplos para ilustrar el uso de diferentes técnicas estadísticas circulares, que van acompañados del código necesario para implementarlas de manera computacional.

El programa informático aquí utilizado para realizar las figuras y gráficas presentadas, y los cálculos necesarios para el desarrollo de este trabajo, es el software libre R Core Team (2015), version 3.2.1, a través del entorno de desarrollo integrado RStudio Team (2015), versión 0.99.467 de escritorio.

1

Datos circulares

La *estadística direccional* se encarga del estudio de las observaciones que pueden ser representadas como vectores unitarios en el plano o en el espacio tridimensional, o equivalentemente, como puntos de la circunferencia unitaria o de la esfera unitaria. Como el espacio muestral es un círculo o una esfera, los métodos estándar para datos con soporte en la recta real no pueden ser utilizados. Los métodos direccionales que se desarrollen requieren que tomen en cuenta la estructura de estos espacios.

En particular, los datos en la esfera unitaria reciben el nombre de *datos esféricos*, y la *estadística esférica* se encarga de ellos. Por ende, la *estadística circular* se encarga de aquellos datos en la circunferencia del círculo unitario, es decir, los *datos circulares*.

En este capítulo se muestra una breve historia del estudio de datos direccionales y algunos ejemplos que se dan en distintas ramas de la ciencia. Además, se presentan las características de los datos circulares (entre ellas la periodicidad), así como también algunas formas de representación gráfica para este tipo de datos; estos gráficos tienen como fin describir el comportamiento de los datos circulares.

1.1. Breve historia del estudio de datos direccionales

Mena (2004) ofrece una breve recopilación sobre el estudio de los datos direccionales. Las referencias más remotas de análisis de datos circulares datan de finales del siglo XVII, con los trabajo John Mitchell, padre de la sismología, quien empleó es-

tos análisis para estudiar las separaciones angulares entre estrellas con el objetivo de probar la hipótesis de que sus direcciones estaban uniformemente distribuidas. Sin embargo, muchos de los primeros trabajos en estadística circular se originaron a través del interés en comprender la habilidad de las palomas mensajeras en retornar a sus hogares. Estos análisis luego se extendieron a los estudios del empleo que hacen las aves migratorias de su reloj interno y de la dirección del sol para orientarse.

La estadística direccional cobró gran impulso cuando a principios de la década de los 50's investigadores en Geofísica y Geodesia de la Universidad de Cambridge llevaban a cabo estudios direccionales de magnetización remanente en rocas volcánicas de Islandia, y encontraron que estas direcciones formaban dos grupos apartados 180° uno del otro. Este hecho era importante en la demostración de la ocurrencia de inversiones de polaridad del campo terrestre (tema aún controvertido para la época) y, por consiguiente, surgió la necesidad de contar con un método estadístico para el tratamiento de los datos y poder obtener así conclusiones sostenibles. Ronald Fisher, en ese entonces profesor de aquella Universidad, se interesó en este tema, diseñó rápidamente su estadística en la esfera y junto con el grupo de geofísicos aplicó por primera vez estos métodos. En 1953 Fisher publicó su trabajo clásico de estadística direccional *Dispersion on a sphere*.

Durante los últimos 40 años la estadística direccional se ha desarrollado ampliamente. La bibliografía existente sobre el tema se enriqueció con los libros *Statistics on Spheres* (Watson, 1980), *Circular Statistics in Biology* (Batschelet, 1981), *Statistical Analysis of Spherical Data* (Fisher et al., 1987), *Statistical Analysis of Circular Data* (Fisher, 1993), y los más recientes *Directional Statistics* (Mardia y Jupp, 1999), *Topics in Circular Statistics* (Jammalamadaka y Sengupta, 2001) y *Circular Statistics in R* (Pewsey et al., 2013).

1.2. Ejemplos de datos direccionales

Los datos direccionales surgen con bastante frecuencia en muchas ciencias naturales y físicas como Biología, Medicina, Ecología, Geología, Meteorología, Física, Astronomía, etcétera.

Jammalamadaka et al. (1986) discuten una aplicación médica interesante donde los ángulos de flexión de la rodilla fueron medidos para evaluar la recuperación de paciente ortopédicos.

Los ecologistas consideran que la dirección predominante del viento es un factor importante en muchos estudios, incluidos los que implican el transporte de conta-

minantes.

Los geólogos estudian paleocorrientes para hacer inferencia acerca de la dirección del flujo de los ríos en el pasado (Sengupta y Rao, 1966) y analizar las direcciones paleomagnéticas del polo magnético de la Tierra, para investigar el fenómeno de la inversión de los polos, así como en apoyo de la hipótesis de la deriva continental.

En Física, un conjunto de datos el cual llevó a la introducción de una de las distribuciones básicas en estadística circular consiste de las partes fraccionarias (convertidas en ángulos) de los pesos atómicos de los elementos conocidos hasta 1918. Antes del descubrimiento de los isótopos, se había especulado que las medidas de los pesos atómicos eran enteros sujetos a error. Von Mises (1918) propone verificar esto al probar si la correspondiente distribución en el círculo tenía o no moda en 0° .

En Biología, los estudios sobre la orientación de los animales generan datos circulares. Las preguntas típicas de interés son (i) si las direcciones de los animales se distribuyen uniformemente en el círculo y (ii) si los animales tienden a ir en una dirección específica. La respuesta a esta última pregunta es útil en los intentos de determinar si los animales utilizan signos de navegación tales como la dirección del sol o el campo magnético de la tierra. Se dan discusiones de diversas investigaciones en el campo de la navegación de las aves en Schmidt-Koenig (1965) y Batschelet (1981).

Datos circulares surgen también del estudio de los ritmos biológicos (o circadianos). El círculo puede ser usado para representar un ciclo, y el interés puede estar en el tiempo de un evento dentro del ciclo, digamos por ejemplo cuando la temperatura corporal o la presión de la sangre alcanzan su punto máximo durante el día.

En Psicología, los datos direccionales aparecen en la percepción de la dirección bajo varias condiciones, tales como la simulación de la gravedad cero (Ross *et al.*, 1969). Datos circulares se producen también en los estudios de los mapas mentales que las personas usan para representar su entorno (Gordon *et al.*, 1989).

Ejemplos adicionales incluyen los ángulos de unión de moléculas, las veces durante el día en que suceden ciberataques a un centro de inteligencia, y la incidencia a través del año del sarampión.

1.3. Datos circulares

Las observaciones que se pueden representar como vectores unitarios en \mathbb{R}^2 reciben el nombre de *datos circulares*. Equivalentemente, los datos circulares son aquellos

que pueden ser representados como puntos de una circunferencia de un círculo unitario, de aquí que reciben su peculiar nombre.

Los datos circulares incluyen direcciones medidas usando instrumentos tales como la brújula, el transportador, la veleta, el sextante y el teodolito. Pero aunque las direcciones, registradas en ángulos, son un tipo de datos circulares, no todos los datos circulares son necesariamente inicialmente medidos o registrados como ángulos. La clave para entender este punto es la *periodicidad* de los datos (Pewsey *et al.*, 2013). De hecho, cualquier fenómeno periódico con período conocido digamos un día, un mes o un año, puede ser representado en el círculo (donde la circunferencia corresponde a este período), mediante la agregación, según sea necesario, de datos sobre varios individuos o subperíodos (Jammalamadaka y Sengupta, 2001).

Por ejemplo, consideremos el tiempo de un día medido en un reloj de 24 horas. Los tiempos 0:00 y 24:00, ambos corresponden a la media noche, y 1:00 y 23:00 ambos definen tiempos de una hora después y antes de la media noche. Cada tiempo corresponde a una posición de la aguja horaria en un reloj de 24 horas, y cada una de estas posiciones pueden ser convertidas en un ángulo medido en grados al multiplicar el tiempo en horas por $360/24$. Estos ángulos pueden ser usados para definir puntos en alrededor de la circunferencia de un círculo unitario. De este modo, los datos así convertidos pueden representar los veces durante el día de la llegada de pacientes a la sala de urgencias de un hospital. Otro ejemplo, es que se podría estar interesado en la ocurrencia a lo largo del año de los sismos en Japón con magnitud de 5 o más en la escala de Richter. Entonces, el gran terremoto ocurrido en el 11 de Marzo de 2011 se podría considerar como muy cerca, en términos del tiempo del año, de uno de magnitud 6.9 ocurrido el 12 de Marzo de 2012. Los datos de este tipo pueden ser convertidos a ángulos medidos en grados al multiplicar el tiempo del año en días, con 0° representando el 1 de Enero, por $360/365$ (no tomando en cuenta años bisiestos), y, posteriormente, representados como puntos en la circunferencia de un círculo unitario (Pewsey *et al.*, 2013).

1.4. Periodicidad de los datos circulares

Dado un punto inicial de referencia, la *dirección cero*, y un sentido de rotación positivo (horario o antihorario), una dato circular puede ser registrado como un ángulo medido desde la dirección cero al punto en la circunferencia correspondiente a la observación. Es usual medirlos en grados o radianes.

Estos requerimientos de especificar una posición de origen y un sentido positivo de rotación no se dan para datos en la línea recta, pues ahí, el origen es 0, los valores a

la derecha del cero son positivos, y los valores a la izquierda del cero son negativos. Para datos circulares, cada ángulo define un punto en la circunferencia de un círculo unitario, al igual que cada valor de una variable lineal define un punto en la línea recta. A medida que el valor absoluto de una variable lineal aumenta, más nos moveremos lejos del origen. Así, en la recta real, el valor de 360 está relativamente cerca de 355 pero relativamente lejos del origen. Esta situación es diferente para variables circulares. Mientras que un ángulo de 355° corresponde a un punto en la circunferencia del círculo unitario que está cerca de la que corresponde a 360° , los ángulos 0° y 360° definen exactamente el mismo punto, por lo que este punto y el primero están relativamente cercanos. Es esta periodicidad natural de los datos circulares la que obliga a abandonar las técnicas estadísticas estándar diseñadas para datos lineales en favor de otras que respeten la periodicidad de los datos circulares (Pewsey *et al.*, 2013).

Un ejemplo clásico de lo que podría pasar si tratamos los datos circulares con herramientas lineales se da, suponiendo que se miden ángulos de la manera clásica (dirección inicial al este y sentido de rotación positivo), cuando se tiene interés sobre las direcciones de vuelo de cuatro aves que se liberan desde un mismo punto. Si consideramos que los ángulos registrados son 355° , 350° , 5° y 10° , el sentido común nos diría que, en general, estas aves volaron aproximadamente a una dirección de 0° , es decir, al este. Sin embargo, la media aritmética de estos datos es 180° , y apunta a una dirección opuesta al origen, o sea al oeste, lo cual no es para nada razonable si se quiere tomar este valor como un representante del comportamiento de las direcciones de vuelo.

El punto de referencia, o “cero”, tales como el 1 de enero, medianoche o norte, es una construcción arbitraria que generalmente no se relaciona bien con los factores subyacentes del sistema bajo estudio. Podemos, por convención, etiquetar 1 de Enero como día uno y 31 de Diciembre como día 365, pero en términos de la incidencia de sarampión en un área metropolitana grande, o la dirección predominante del viento en una estación meteorológica, podría esperarse que los valores en estos días presenten cierto comportamiento similar.

1.5. Representación gráfica de datos circulares

1.5.1. Diagrama Circular

La representación más simple un dato circular es un *raw circular data plot*. Es el gráfico más natural que describe este tipo de datos. En él, un dato circular es

representado por un punto, o cualquier otro símbolo, localizado sobre la circunferencia de un círculo unitario. Cuando un dato se repite, los nuevos puntos se colocan fuera del círculo en la dirección correspondiente.

En la literatura, estos gráficos son citados en español como “gráficos de datos circulares en ramas” (Jaimes Jaimes y Pineda Garavito, 2012) o simplemente como “diagramas circulares” (Nuñez Antonio, 2010). Utilizaremos esta última denominación.

Un ejemplo del uso de este gráfico es el de las posiciones de parada de una ruleta presentado por Mardia y Jupp (1999). Se exhibe a continuación.

Ejemplo 1.1. *Una ruleta fue girada y las posiciones en las que se detuvo fueron medidas. Las posiciones de parada en 9 intentos fueron 43° , 45° , 52° , 61° , 75° , 88° , 88° , 279° , 357° .*

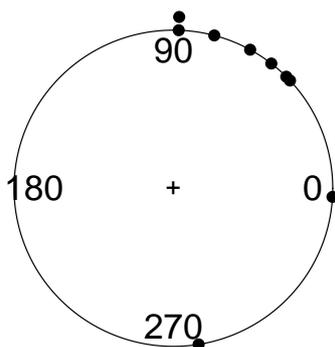


Figura 1.1: Diagrama circular de los datos de la ruleta del Ejemplo 1.1.

1.5.2. Diagrama de Rosa

Análogo al histograma lineal para datos circulares es el *diagrama de rosa*, en el cual las frecuencias son representadas por áreas de sectores en lugar de barras.

La Figura 1.2 muestra un ejemplo sobre el uso del diagrama de rosa que es presentado originalmente por (Pewsey *et al.*, 2013). En este ejemplo, se presentan 310 direcciones del viento, medidos en sentido horario desde el norte, en radianes, registrados en una estación meteorológica en los Alpes italianos cada 15 minutos desde las 3:00 am a 4:00 am, entre el 29 de Enero de 2001 al 31 de Marzo de 2001. Estos datos fueron introducidos a la literatura de la estadística circular por Agostinelli (2007).

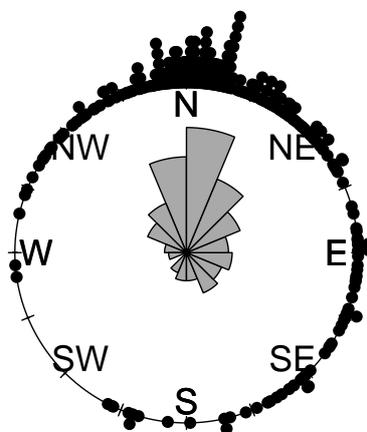


Figura 1.2: Diagrama circular y, en su centro, un diagrama de rosa para 310 direcciones del viento. Las áreas de los sectores en el diagrama de rosa representan las frecuencias relativas en los 16 intervalos de clase.

Más formalmente, en un diagrama de rosa, el área de cada sector es proporcional a la frecuencia en el grupo correspondiente. Para lograr esto en grupos de igual amplitud, el radio de cada sector debe ser proporcional a la raíz cuadrada de la frecuencia relativa. No todos los autores en la literatura de la estadística circular siguen esta convención, sin embargo, en el presente trabajo sí se utilizará.

Al igual que los histogramas, se puede elegir cuántos sectores del diagrama de rosa dividirán los datos. Esta elección puede influir fuertemente en cómo la gráfica es interpretada, así que se recomienda probar un rango de valores para los números de sectores y comparar los resultados. Como regla empírica, la raíz del tamaño de muestra es usada frecuentemente como una primera aproximación razonable a un apropiado número de sectores.

1.5.3. Histogramas lineales

Ya que los estadísticos han adquirido experiencia interpretando histogramas en la línea recta, o sea *histogramas estándares o lineales*, una primera opción gráfica podría ser el uso adecuado de ellos en datos circulares.

Para producir un histograma lineal de datos circulares, cortamos la circunferencia del círculo en un punto específico y asociamos los dos puntos finales del círculo cortado con los dos puntos finales del histograma lineal. En consecuencia, su

interpretación puede estar potencialmente obstaculizada debido a que la periodicidad de los datos se pierde. Implícitamente se asume que uno será consciente de que ambos extremos del histograma corresponden al mismo punto. Se requiere el uso de un poco de esfuerzo mental para envolver el histograma lineal de nuevo en la circunferencia del círculo unitario. Además, la cuestión de dónde cortar la circunferencia es especialmente delicada.

La primera impresión visual dada por un histograma lineal puede ser sensible al punto donde la circunferencia fue cortada Mardia y Jupp (1999). Algunos autores, como ?, abogan por dibujar dos veces el mismo histograma, esto es, una repetición seguida inmediatamente de la otra, con el fin de evitar efectos visuales erróneos derivados de la naturaleza arbitraria del punto de corte. Sin embargo, para Pewsey *et al.* (2013), esto podría dar una interpretación considerablemente errónea si un observador casual pierde de vista el hecho de que los datos han sido graficados dos veces (y por tanto, digamos, podría interpretar equivocadamente un conjunto de datos unimodales como bimodales). En vez de esto, si se requiere usar histogramas lineales, ellos recomiendan graficar solo una vez los datos, pero analizando primero acerca del punto de corte a usar más apropiado.

Mardia y Jupp (1999) consideran que si los datos tienen una sola moda, o esa una dirección preferida, es buena idea usar un punto de corte muy cercano a la dirección opuesta a la moda. Un punto muy cercano a la moda podría dar la impresión errónea de que los datos son bimodales.

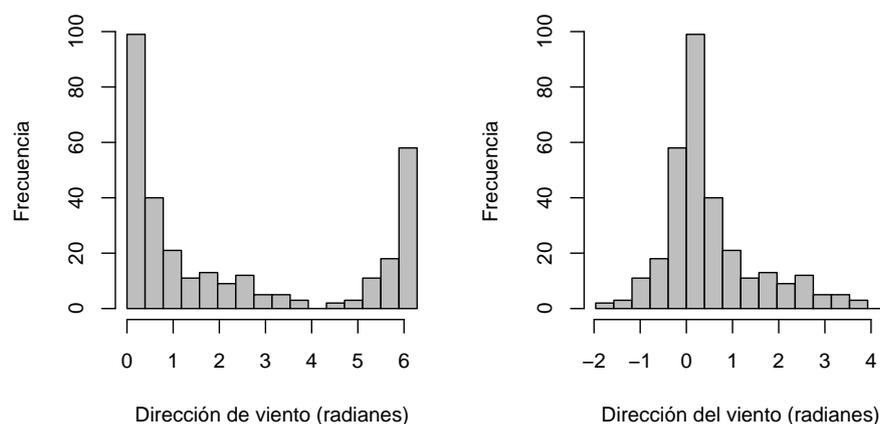


Figura 1.3: Histogramas lineales de 310 direcciones del viento (izquierda) con punto de corte en 0 (norte) y en un rango de $[0, 2\pi)$, y (derecha) con punto de corte en $2\pi - 5\pi/8$ y en un rango de $[-5\pi/8, 2\pi - 5\pi/8)$

Para los datos presentados por Agostinelli (2007), al ver la Figura 1.3 sin tener

en cuenta las observaciones ya planteadas al principio, un error de interpretación inicial del primer histograma podría ser que la distribución de los datos sea bimodal. Sin embargo, esta impresión es consecuencia de que la circunferencia ha sido cortada en un lugar inapropiado, a saber 0, que corresponde al norte, alrededor del cual los datos se distribuyen muy cercanamente. Con el fin de interpretar los datos correctamente, se requiere tomar en cuenta que los valores 0 y 2π se refieren al mismo punto. Claramente, un mejor lugar para cortar el círculo sería $2\pi - 5\pi/8$, pues allí se concentran menos datos.

En el segundo histograma, los valores negativos a través del eje horizontal simplemente representan direcciones medidas de manera antihoraria desde el norte y son obviamente equivalentes a sus valores originales positivos (mod 2π). En este histograma, se refleja claramente las características de unimodalidad y de sesgo de la distribución de los datos en sentido horario de norte a sur. Sin embargo, si bien es mucho más fácil interpretar este histograma que el primero, todavía se requiere equiparar los dos extremos como correspondientes a un mismo punto en la circunferencia unitaria.

Como en la construcción de cualquier histograma, el mejor número de intervalos de clase a usar generalmente requiere de algo de exploración.

2

Estadísticos descriptivos circulares

El camino más natural para representar gráficamente una observación circular es a través de un punto en la circunferencia unitaria; el círculo representa, por ejemplo, la cara de una brújula, un reloj de 12 horas (o 24), o un medio en el cual identificamos la época del año. El radio del círculo, $r > 0$, es claramente irrelevante, y la elección $r = 1$ simplifica las cosas matemáticamente. Por esta elección, cada observación define un vector unitario \mathbf{x} , dirigido desde el centro del círculo unitario al punto en la circunferencia de dicho círculo.

Equivalentemente, una vez que el origen ha sido elegido (por ejemplo el norte, o el utilizado en matemáticas que corresponde al eje horizontal positivo), junto con una orientación (por ejemplo, en sentido horario desde el norte o antihorario desde el origen matemático), una observación circular puede ser representada por el ángulo θ , subtendido por el arco alrededor del círculo unitario, en la orientación escogida, desde el origen a la observación. Los valores negativos de θ corresponden a los ángulos medidos en la dirección opuesta a la orientación elegida.

Con fines de estandarización, al trabajar analíticamente con ángulos se referirá a ellos medidos en radianes y utilizando la representación matemática (eje horizontal positivo hacia el este y orientación positiva antihoraria). Claramente, el ángulo usado para identificar un punto en la circunferencia del círculo unitario no es único, ya que θ y $\theta + 2\pi p$, $p = \pm 1, \pm 2, \dots$, corresponden al mismo punto en la circunferencia. Por lo tanto, cuando se hace referencia a un ángulo implícitamente significará que se ha tomado su valor módulo 2π . Por esta elección de unidades, origen y orientación, el vector unitario \mathbf{x} y el ángulo θ están relacionados a través de $\mathbf{x} = (\cos \theta, \text{sen } \theta)^T$, donde el superíndice T significa la transpuesta del vector.

Una mayor simplificación matemática se logra mediante la representación del vector \mathbf{x} en el plano complejo, en lugar del plano real, al tomar el eje horizontal como su componente real y el eje vertical como el imaginario. Usando esta representación alternativa, una observación circular con vector unitario \mathbf{x} puede ser representado por el número complejo $z = e^{i\theta} = \cos \theta + i \operatorname{sen} \theta$, donde $i = (0, 1)$. Tal representación de una observación circular está descrita en la Figura 2.1.

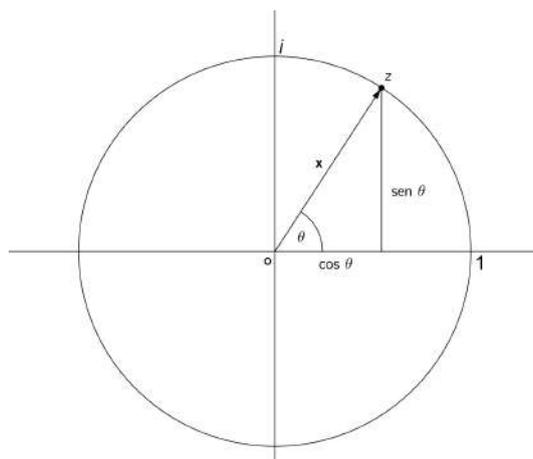


Figura 2.1: Una observación circular con vector unitario \mathbf{x} representada en el plano complejo mediante el número $z = \cos \theta + i \operatorname{sen} \theta$. La flecha identifica al vector unitario \mathbf{x} con el correspondiente ángulo θ medido en sentido antihorario desde el origen $z = (1, 0)$.

En este capítulo se presentan los estadísticos descriptivos circulares comúnmente utilizados para resumir una muestra de datos circulares. Primero se definen los momentos trigonométricos muestrales, ya que estos proporcionan la base para la mayoría de las medidas de localización, concentración, dispersión, sesgo y curtosis. Se termina con las correcciones que se deben aplicar en el análisis de los datos agrupados.

2.1. Momentos trigonométricos muestrales

Considere una muestra aleatoria de observaciones circulares de tamaño n , con vectores unitarios asociados $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, ángulos $\theta_1, \dots, \theta_n$ y números complejos z_1, \dots, z_n .

Definición 2.1. Para $p = 0, \pm 1, \pm 2, \dots$, el p -ésimo momento trigonométrico muestral alrededor de la dirección cero está dado por

$$\begin{aligned}
t_{p,0} &= \frac{1}{n} \sum_{j=1}^n z_j^p = \frac{1}{n} \sum_{j=1}^n (e^{i\theta_j})^p = \frac{1}{n} \sum_{j=1}^n e^{ip\theta_j} = \frac{1}{n} \sum_{j=1}^n (\cos p\theta_j + i \operatorname{sen} p\theta_j) \\
&= a_p + ib_p,
\end{aligned} \tag{2.1}$$

donde

$$a_p = \frac{1}{n} \sum_{i=1}^n \cos p\theta_j, \quad b_p = \frac{1}{n} \sum_{i=1}^n \operatorname{sen} p\theta_j. \tag{2.2}$$

De la ecuación 2.2, se llega a que $a_{-p} = a_p$ y $b_{-p} = -b_p$. Claramente, $t_{0,0} = 1$.

El número complejo $t_{p,0}$ define un **vector medio resultante** en el plano complejo, de longitud

$$\bar{R}_p = (a_p^2 + b_p^2)^{1/2} \in [0, 1], \tag{2.3}$$

con dirección

$$\bar{\theta}_p = \operatorname{arctan}^*(b_p, a_p), \tag{2.4}$$

donde

$$\operatorname{arctan}^*(b_p/a_p) = \begin{cases} \operatorname{arctan}(b_p/a_p), & \text{si } a_p > 0, \\ \operatorname{arctan}(b_p/a_p) + \pi & \text{si } b_p \geq 0, a_p < 0, \\ \operatorname{arctan}(b_p/a_p) - \pi & \text{si } b_p < 0, a_p < 0, \\ \pi/2 & \text{si } b_p > 0, a_p = 0, \\ -\pi/2 & \text{si } b_p < 0, a_p = 0, \\ \text{indefinida} & \text{si } b_p = 0, a_p = 0. \end{cases} \tag{2.5}$$

La función tangente inversa, arctan devuelve valores en $(-\pi/2, \pi/2)$. La función arctan^* regresa valores en $(-\pi, \pi]$ los cuales pueden ser mapeados a $(0, 2\pi]$ al sumar 2π a cada valor negativo.

Cuando $\bar{R}_p > 0$, la representación polar de $t_{p,0}$ es

$$t_{p,0} = \bar{R}_p e^{i\bar{\theta}_p} = \bar{R}_p (\cos \bar{\theta}_p + i \operatorname{sen} \bar{\theta}_p) \tag{2.6}$$

con \bar{R}_p y $\bar{\theta}_p$ denominados **longitud media resultante** y **dirección media**, respectivamente, del vector medio resultante de $p\theta_1, \dots, p\theta_n$. Se sigue de 2.1 y 2.6 que, si $\bar{R}_p > 0$,

$$a_p = \bar{R}_p \cos \bar{\theta}_p, \quad b_p = \bar{R}_p \sen \bar{\theta}_p. \quad (2.7)$$

La dirección media muestral y la longitud media muestral resultante de $\theta_1, \dots, \theta_n$, $\bar{\theta}_1$ y \bar{R}_1 , usualmente son denotados simplemente por $\bar{\theta}$ y \bar{R} , y son denominados inequívocamente como **la dirección media muestral** y **la longitud media resultante muestral**, respectivamente. Se sigue de 2.2 y 2.7 que, cuando $\bar{R} > 0$,

$$\frac{1}{n} \sum_{j=1}^n \cos(\theta_j - \bar{\theta}) = \bar{R}, \quad \frac{1}{n} \sum_{j=1}^n \sen(\theta_j - \bar{\theta}) = 0.$$

Teorema 2.1. Si $\bar{\theta}$ es la dirección del vector resultante de $(\theta_1, \theta_2, \dots, \theta_n)$ entonces

$$\frac{1}{n} \sum_{i=1}^n \sen(\theta_i - \bar{\theta}) = 0 \quad (2.8)$$

y

$$\frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = \bar{R} \quad (2.9)$$

Demostración: Se sigue de que:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sen(\theta_i - \bar{\theta}) &= \frac{1}{n} \sum_{i=1}^n (\sen \theta_i \cos \bar{\theta} - \cos \theta_i \sen \bar{\theta}) \\ &= \frac{1}{n} \sum_{i=1}^n \sen \theta_i \cos \bar{\theta} - \frac{1}{n} \sum_{i=1}^n \cos \theta_i \sen \bar{\theta} \\ &= b_1 \cos \bar{\theta} - a_1 \sen \bar{\theta} \\ &= \bar{R} \sen \bar{\theta} \cos \bar{\theta} - \bar{R} \cos \bar{\theta} \sen \bar{\theta} \\ &= 0 \end{aligned}$$

Similarmente,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) &= \frac{1}{n} \sum_{i=1}^n (\cos \theta_i \cos \bar{\theta} + \sen \theta_i \sen \bar{\theta}) \\ &= \frac{1}{n} \sum_{i=1}^n \cos \theta_i \cos \bar{\theta} + \frac{1}{n} \sum_{i=1}^n \sen \theta_i \sen \bar{\theta} \\ &= a_1 \cos \bar{\theta} + b_1 \sen \bar{\theta} \\ &= \bar{R} \cos \bar{\theta} \cos \bar{\theta} + \bar{R} \sen \bar{\theta} \sen \bar{\theta} \\ &= \bar{R} \cos^2 \bar{\theta} + \bar{R} \sen^2 \bar{\theta} \\ &= \bar{R} \end{aligned}$$



La ecuación 2.8 es análoga a

$$\sum_{j=1}^n (x_j - \bar{x}) = 0,$$

para las observaciones x_1, \dots, x_n en la línea recta con media muestral \bar{x} .

Los momentos trigonométricos muestrales alrededor de la dirección media $\bar{\theta}$ pueden ser definidos y resultan ser útiles.

Definición 2.2. *El p -ésimo momento trigonométrico muestral alrededor de la dirección media está dado por*

$$\begin{aligned} t_{p,\bar{\theta}} &= \frac{1}{n} \sum_{j=1}^n e^{ip(\theta_j - \bar{\theta})} \\ &= \frac{1}{n} \sum_{j=1}^n \{ \cos p(\theta_j - \bar{\theta}) + i \operatorname{sen} p(\theta_j - \bar{\theta}) \} \\ &= \bar{a}_p + i \bar{b}_p \end{aligned} \quad (2.10)$$

donde

$$\bar{a}_p = \frac{1}{n} \sum_{i=1}^n \cos p(\theta_j - \bar{\theta}), \quad \bar{b}_p = \frac{1}{n} \sum_{i=1}^n \operatorname{sen} p(\theta_j - \bar{\theta}). \quad (2.11)$$

De (2.8)

$$\bar{a}_1 = \bar{R}, \quad \bar{b}_1 = 0, \quad (2.12)$$

y por lo tanto $t_{1,\bar{\theta}} = \bar{R}$. En consecuencia, para la muestra centrada, $\theta_1 - \bar{\theta}, \dots, \theta_n - \bar{\theta}$, la dirección media es 0 pero la longitud media resultante no cambia.

Como a continuación se tratará, la mayoría de las medidas utilizadas para resumir datos circulares están relacionadas con los momentos trigonométricos muestrales. Los análogos poblacionales de los momentos trigonométricos muestrales serán introducidos posteriormente.

2.2. Medidas de Localización

2.2.1. Dirección Media Muestral

La *dirección media muestral* $\bar{\theta} \equiv \bar{\theta}_1$ definida en la sección 2.1 es comúnmente la medida de localización para datos circulares más utilizada. Cuando ésta existe,

corresponde a la dirección del vector medio resultante. No existirá, por ejemplo, para una muestra de la forma $\theta_1, \dots, \theta_{n/2}, \theta_1 + \pi, \dots, \theta_{n/2} + \pi$; los pares obvios de los vectores unitarios se cancelan entre sí y esto conduce a que $\bar{R} = 0$. Otras formas de muestras cíclicas, para triples, cuádruples, etcétera, de los vectores unitarios se anulan entre sí, y también se llega a que $\bar{R} = 0$. Para muestras unimodales que están cerca de ser simétricas, $\bar{\theta}$ ofrece una buena medida de localización central.

Se puede demostrar que la dirección media es equivariante bajo rotación; es decir, si una muestra de datos circulares con dirección media $\bar{\theta}$ se rota en sentido horario a través de un ángulo ψ , la dirección media de la muestra rotada será $\bar{\theta} + \psi \pmod{2\pi}$.

2.2.2. Dirección Mediana Muestral

La ***dirección mediana muestral***, denotada por $\tilde{\theta}$, ofrece una alternativa robusta a la dirección media muestral, $\bar{\theta}$, como una medida de localización. Es particularmente útil cuando la muestra es sesgada. Una dirección mediana muestral es definida como cualquier ángulo ψ para la cual la mitad de los datos pertenecen al arco $[\psi, \psi + \pi)$ y la mayoría de los datos está cerca de ψ que de $\psi + \pi$. De esta definición, es claro que la dirección mediana no necesariamente es única. Cuando n es impar, una dirección mediana corresponderá a uno de los puntos de los datos. Cuando n es par, es usual tomar el promedio de dos puntos adyacentes.

Formalmente, una dirección mediana puede ser identificada por minimizar la medida de dispersión (ver Sección 2.3.3)

$$d_2(\psi) = \frac{1}{n} \sum_{j=1}^n \{\pi - |\pi - |\theta_j - \psi||\}. \quad (2.13)$$

Un ejemplo sobre el uso de los estadísticos circulares dirección media y la dirección mediana muestrales es ofrecido a continuación, y se utilizan para los datos circulares del ejemplo 1.1 del capítulo 1.

Ejemplo 2.1. *Para el ejemplo de la ruleta en el Ejemplo 1.1, las observaciones (en grados) son $43^\circ, 45^\circ, 52^\circ, 61^\circ, 75^\circ, 88^\circ, 88^\circ, 279^\circ, 357^\circ$. Para estos datos, la dirección mediana $\tilde{\theta} = 52^\circ$ y la dirección media $\bar{\theta} = 51.0501^\circ$ (ver Figura 2.2).*

El Ejemplo 2.1 refleja la razonabilidad de los valores de la media y la mediana muestrales obtenidos para los datos de ruleta del ejemplo 1.1. Si éstos se hubiesen tratado como datos en la línea recta, entonces $\hat{x} = 75^\circ$ y $\bar{x} = 120.8889^\circ$ no son medidas representativas coherentes con la distribución de los datos de ruleta en la circunferencia.

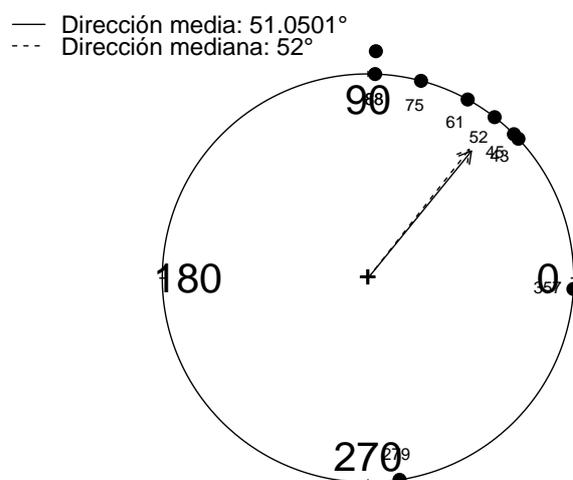


Figura 2.2: Direcciones media y mediana muestrales para los datos del Ejemplo 1.1.

2.3. Medidas de Concentración y Dispersión

2.3.1. Longitud Media Resultante Muestral

La *longitud media resultante muestral*, $\bar{R} = \bar{R}_1 \in [0, 1]$, definida en la sección 2.1, es la medida de *concentración* más comúnmente utilizada para datos circulares unimodales. La medida \bar{R} es igual a 1 solamente cuando los datos están localizados en el mismo punto en la circunferencia unitaria (o, equivalentemente, todos los vectores unitarios son idénticos). Un valor cercano a 1 indica que los datos están estrechamente cercanos alrededor de la dirección media. Cuando los datos están distribuidos uniformemente alrededor de la circunferencia, \bar{R} tomará el valor cercano a 0. Sin embargo, el valor de 0 para \bar{R} no debería ser interpretado necesariamente como un indicador de que los datos estén uniformemente distribuidos alrededor del círculo ya que, como se ha visto en la Sección 2.2.1, $\bar{R} = 0$ para cualquier muestra con estructura cíclica.

2.3.2. Varianza Muestral Circular y Desviación Estándar

La *varianza circular muestral* está definida como

$$V = 1 - \bar{R}. \quad (2.14)$$

Ya que $\bar{R} \in [0, 1]$, entonces $V \in [0, 1]$. Batschelet (1981) usa el término “varianza angular” para referirse a $2(1 - \bar{R}) \in [0, 2]$ como una medida de variabilidad circular.

La *desviación estándar circular muestral* está dada por

$$\hat{\sigma} = \{-2 \log(1 - V)\}^{1/2} = \{-2 \log \bar{R}\}^{1/2} \in [0, \infty]. \quad (2.15)$$

Para muestras concentradas, con V pequeña,

$$\hat{\sigma} \simeq (2V)^{1/2} = \{2(1 - \bar{R})\}^{1/2}, \quad (2.16)$$

lo cual explica la definición de la varianza angular como una alternativa a la varianza circular (2.14). Batschelet (1981) se refiere a $\{2(1 - \bar{R})\}^{1/2}$ como la *desviación angular (media)*.

De las dos medidas, hay una preferencia por V , ya que la escala $[0, 1]$ parece más natural que $[0, 2]$.

2.3.3. Otras Medidas Muestrales de Dispersión

La llamada *dispersión circular muestral* está definida como

$$\hat{\delta} = \frac{1 - \bar{R}_2}{2\bar{R}^2}, \quad (2.17)$$

donde \bar{R}_2 es la longitud media resultante del doble de los ángulos, $2\theta_1, \dots, 2\theta_n$, como fue definido en (2.3).

Comúnmente se usan dos *medidas de distancia* entre dos ángulos, ψ y ω , las cuales conducen a dos medidas alternativas de dispersión de los datos circulares. La primera medida de distancia es

$$1 - \cos(\psi - \omega) \quad (2.18)$$

con la medida de dispersión asociada de $\theta_1, \dots, \theta_n$ alrededor de ψ ,

$$d_1(\psi) = \frac{1}{n} \sum_{j=1}^n \{1 - \cos(\theta_j - \psi)\}. \quad (2.19)$$

La medida de dispersión $d_1(\psi)$ es minimizada cuando $\psi = \bar{\theta}$, con $d_1(\bar{\theta}) = 1 - a_1 = 1 - \bar{R} = V$. Usando (2.18) como la medida de distancia, la distancia media entre dos datos es

$$\bar{d}_1 = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \{1 - \cos(\theta_j - \theta_k)\} = 1 - \bar{R}^2.$$

La segunda medida de distancia comúnmente usada es

$$\min(\psi - \omega, 2\pi - (\psi - \omega)) = \pi - |\pi - |\psi - \omega||, \quad (2.20)$$

con su medida de dispersión asociada (2.13). Como se explicó en la Sección 2.2.2, $d_2(\psi)$ es minimizada cuando $\psi = \tilde{\theta}$ y $d_2(\tilde{\theta})$ es denominada como la **desviación mediana circular**. Usando (2.20) como la medida de distancia, la distancia media entre dos puntos está dada por:

$$\bar{d}_2 = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \{\pi - |\pi - |\theta_j - \theta_k||\} \in [0, \pi/2].$$

Finalmente, el **rango circular**, W , está definido como la longitud del arco más pequeño que contiene a todas las observaciones.

Se recomienda usar la longitud media resultante \bar{R} o la varianza circular V para describir la distribución de muestras unimodales.

2.4. Medidas de Sesgo y Curtosis

Como vimos en (2.12), los elementos del primer momento trigonométrico muestral alrededor de la dirección media, $t_{1,\bar{\theta}}$, son $\bar{a}_1 = \bar{R}$ y $\bar{b}_1 = 0$. Por tanto \bar{a}_1 es una medida de concentración pero \bar{b}_1 no es útil como estadístico descriptivo. Por ello, los elementos del segundo momento trigonométrico alrededor de la dirección media, $t_{2,\bar{\theta}}$, ofrecen las medidas básicas de sesgo y curtosis para datos circulares.

El “seno” del segundo momento trigonométrico central

$$\bar{b}_2 = \frac{1}{n} \sum_{j=1}^n \sin 2(\theta_j - \bar{\theta}) = \bar{R}_2 \sin(\bar{\theta}_2 - 2\bar{\theta}) \quad (2.21)$$

es el “seno” de los momentos trigonométricos centrales que es no cero y así ofrece una medida del **sesgo** de datos circulares alrededor de la dirección media. Para un conjunto de datos unimodales, \bar{b}_2 será cercano a cero si la distribución de los datos es casi simétrica alrededor del $\bar{\theta}$, y relativamente grande y negativa (positiva) cuando los datos están sesgados en sentido antihorario (horario) lejos de la dirección media.

El “coseno” del segundo momento central

$$\bar{a}_2 = \frac{1}{n} \sum_{j=1}^n \cos 2(\theta_j - \bar{\theta}) = \bar{R}_2 \cos(\bar{\theta}_2 - 2\bar{\theta}) \quad (2.22)$$

es una medida de la curtosis de los datos circulares. Toma el valor de 1 cuando los datos son todos idénticos, que corresponde a la máxima picudez, y los valores cercanos a cero se atribuyen a datos casi uniformemente distribuidos alrededor del círculo.

Las propiedades de \bar{b}_2 y \bar{a}_2 para la concentración de datos circulares llevan a Mardia (1972) a proponer las *medidas estandarizadas para el sesgo y curtosis*

$$\hat{s} = \frac{\bar{b}_2}{(1 - \bar{R})^{3/2}} \quad (2.23)$$

y

$$\hat{k} = \frac{\bar{a}_2 - \bar{R}^4}{(1 - \bar{R})^2}. \quad (2.24)$$

La interpretación de \hat{s} es similar a la de \bar{b}_2 . Para datos de una *distribución normal envolvente* (presentada en la Sección 3.3.5), $\bar{a}_2 - \bar{R}^4$ y \hat{k} estarán cercanos a cero. Como $\bar{R} \in [0, 1]$, las magnitudes de \hat{s} y \hat{k} generalmente serán más grandes que sus contrapartes no estandarizadas, \bar{b}_2 y $\bar{a}_2 - \bar{R}^4$.

2.5. Correcciones para datos agrupados

En muchos contextos, en los cuales datos circulares son registrados, se pueden presentar dificultades para medir el ángulo de manera precisa. Bajo tales circunstancias, la circunferencia del círculo unitario es generalmente dividida en un número par de sectores, con arcos de igual longitud, y las frecuencias de los datos que caen en los diferentes sectores son lo que se registra. Los datos resultantes están claramente agrupados. La agrupación es más o menos “gruesa” en función de la longitud de arco empleada.

Cuando se trata de analizar tales frecuencias de datos, los estadísticos descriptivos usualmente son calculados con los puntos que caen en un arco, cuando todos a ellos se les ubica en el centro del arco. En general, el valor de todas las direcciones medias resultantes $\bar{\theta}_p$ calculados por medio de este camino solo diferirán ligeramente de los valores que se pudiesen haber obtenido si los datos se hubiesen medido precisamente (esto por cálculos hechos en Stuart y Ord (1987)). Sin embargo, los valores de todas las longitudes medias resultantes \bar{R}_p generalmente serán un tanto más chicas de lo que hubiesen sido con más precisión. Esto es importante ya que, que se ha visto, \bar{R} y \bar{R}_2 aparecen en las definiciones de algunos de los estadísticos más comúnmente usados para describir datos circulares.

La corrección defendida por Mardia (1972) es calcular

$$\bar{R}_p^* = c(p\psi) \bar{R}_p, \quad (2.25)$$

donde el término de corrección está dado por

$$c(p\psi) = \frac{p\psi/2}{\text{sen}(p\psi/2)},$$

y ψ es el ángulo, medido en radianes, subtendido por cada arco. Por ejemplo, para 16 arcos de igual longitud, $\psi = \pi/8$. La magnitud del factor de corrección $c(p\psi)$ incrementa conforme aumentan p y el “grosor” de la agrupación representada por ψ .

Siguiendo a Jammalamadaka y Sengupta (2001), se recomienda aplicar esta corrección solo si los datos se dividen en menos de ocho segmentos.

3

Distribuciones de probabilidad circulares

Las propiedades de las funciones de distribución circulares y las densidades de probabilidad tienen mucho en común con sus contrapartes lineales, pero también se presentan importantes diferencias como consecuencia del hecho de procurar estén bien definidas, y de que reflejen la periodicidad.

En este capítulo se presentan las definiciones de función de distribución y de densidad de probabilidad circulares, y de la función característica circular. Se presentan los momentos trigonométricos poblacionales y las medidas poblacionales de localización, dispersión, sesgo y curtosis. Se muestran algunos enfoques para la generación de modelos circulares. Se presentan los modelos circulares clásicos y algunos propuestos recientemente en la literatura; estos últimos se caracterizan por ser modelos flexibles, al tomar en cuenta no sólo la localización y la concentración, sino también la curtosis (familia de distribuciones simétricas Jones-Pewsey), o la curtosis y el sesgo (distribuciones Inversa Batschelet).

El contenido de este capítulo se ha basado considerando como principal soporte el trabajo de resumen hecho por Pewsey *et al.* (2013).

3.1. Teoría distribucional circular

Al especificar una distribución circular se hace a través de una variable aleatoria Θ , la cual toma un valor angular que corresponde a un punto en la circunferencia unitaria. Cuando definimos una variable aleatoria angular es necesario seleccionar

una dirección inicial, un sentido de rotación y las unidades de medida. En el presente trabajo, se sigue utilizando la representación “matemática” de un ángulo, es decir, la dirección cero se toma a través del eje horizontal positivo, y los ángulos son medidos en sentido antihorario en radianes. El rango de Θ puede ser tomado en $[0, 2\pi)$ o $[-\pi, \pi)$.

Las distribuciones circulares son esencialmente de dos tipos: pueden ser discretas, asignando masas de probabilidad solamente a un número contable de direcciones, o pueden ser absolutamente continuas (con respecto a la medida de Lebesgue en la circunferencia). La mayoría de los modelos para distribuciones circulares asumen como absolutamente continua en Θ y pueden ser especificadas mediante una función de densidad de probabilidad.

3.1.1. Funciones de distribución circular y de densidad de probabilidad

Un camino para especificar una distribución en el círculo unitario es por medio de su función de distribución. Supongamos que una dirección inicial y una orientación han sido elegidos, y se especifica un ángulo aleatorio Θ . La *función de distribución* F es definida como la función de toda la recta real dada por

$$F(\theta) = P(0 < \Theta \leq \theta), \quad 0 \leq \theta \leq 2\pi \quad (3.1)$$

y

$$F(\theta + 2\pi) - F(\theta) = 1, \quad -\infty < \theta < \infty. \quad (3.2)$$

La ecuación 3.1 es un análogo circular de la definición usual de una función de distribución para una variable aleatoria observada en la recta real. La ecuación 3.2 es una condición extra impuesta con el fin de reflejar la periodicidad de una distribución circular. Esta propiedad está relacionada con la propiedad de que la probabilidad total asociada con una distribución lineal debe ser 1. Esta puede ser interpretada como la condición de que la probabilidad de obtener un punto en el círculo dentro de cualquier arco de longitud 2π radianes sea 1. Esta importante diferencia con respecto a la función de distribución de una variable aleatoria lineal implica que

$$\lim_{\theta \rightarrow -\infty} F(\theta) = -\infty, \quad \lim_{\theta \rightarrow \infty} F(\theta) = \infty.$$

Así, en general, los valores tomados por F claramente no son probabilidades. Sin embargo, para $\alpha \leq \beta \leq \alpha + 2\pi$,

$$P(\alpha < \Theta \leq \beta) = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} dF(\theta), \quad (3.3)$$

donde la integral es una integral de Lebesgue-Stieltjes. Esta función es continua por la derecha y por definición $F(0) = 0, F(2\pi) = 1$.

Hay que notar que, aunque la función F depende de la elección de la dirección cero, (3.3) muestra que $F(\beta) - F(\alpha)$ es independiente de esta elección.

La *función cuantil* se define en $0 \leq u \leq 1$ como

$$Q(u) = \inf \{ \theta : F(\theta) \geq u \}, \quad (3.4)$$

es decir, el valor mínimo de θ para el cual $F(\theta) = P(0 < \Theta \leq \theta)$ es al menos u .

Si la función de distribución circular F es absolutamente continua entonces tiene una función de densidad circular de probabilidad f tal que

$$\int_{\phi}^{\psi} f(\theta) d\theta = F(\psi) - F(\phi), \quad -\infty < \phi \leq \psi < \infty.$$

Claramente, el valor tomado por esta integral no necesariamente es una probabilidad (por ejemplo, si $\psi = \phi + 4\pi$). Una función f es una función de densidad de probabilidad de una distribución circular absolutamente continua si y solo si

1. $f(\theta) \geq 0$ casi dondequiera en $(-\infty, \infty)$,
2. $f(\theta + 2\pi) = f(\theta)$ casi dondequiera en $(-\infty, \infty)$,
3. $\int_0^{2\pi} f(\theta) d\theta = 1$.

La primera propiedad también se sostiene para la densidad de una variable aleatoria observada en la recta real. Las otras dos son consecuencias de la periodicidad de una distribución circular.

Pewsey *et al.* (2013) ofrecen el ejemplo de la Figura 3.1 para representar las funciones F y f . La Figura 3.11 muestra las representaciones lineales de la función de distribución $F(\theta)$ (línea discontinua), y de la densidad $f(\theta)$ (línea sólida) para una distribución von Mises con media en $\mu = \pi$ y parámetro de concentración $\kappa = 2$, para $\theta \in [-8\pi$ y $8\pi]$ radianes. La Figura 3.12 muestra la representación polar de $f(\theta)$ de esta distribución y la Figura 3.2 muestra la representación lineal de la misma graficada para $\theta \in [0, 2\pi)$.

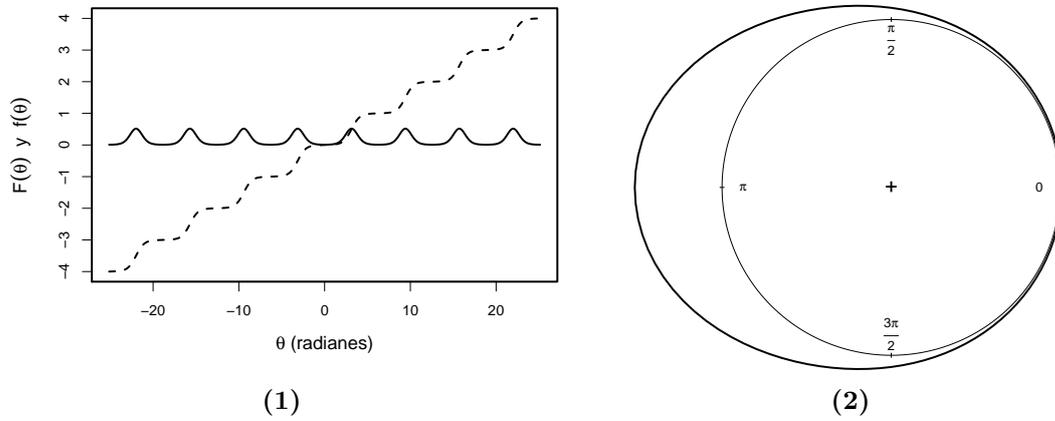


Figura 3.1: Representaciones de las funciones F (línea discontinua) y f (línea continua) de una distribución von Mises con dirección media $\mu = \pi$ y parámetro de concentración $\kappa = 2$.

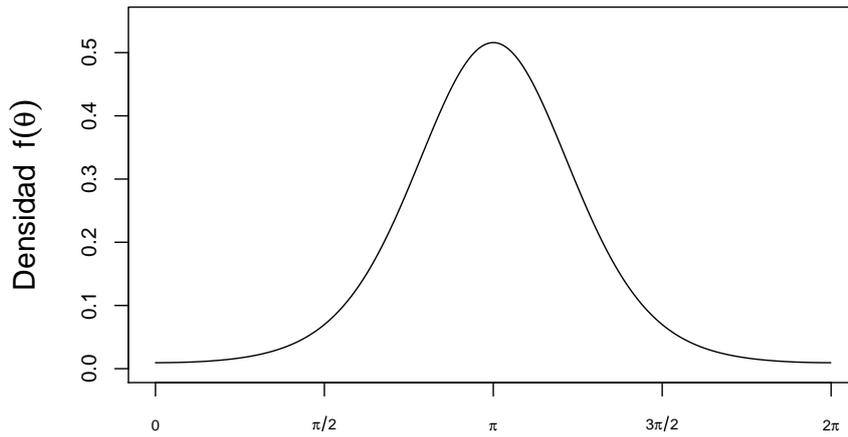


Figura 3.2: Representación lineal de una distribución circular continua.

3.1.2. Función característica circular y Momentos Trigonométricos.

Una herramienta útil para manejar la distribución del ángulo aleatorio θ podría ser la función

$$t \mapsto E[e^{it\Theta}].$$

Sin embargo, ya que Θ es una variable aleatoria periódica, tiene la misma distribución que $(\Theta + 2\pi)$, por lo tanto,

$$E(e^{it\Theta}) = E(e^{it(\Theta+2\pi)}) = e^{it2\pi} \cdot E(e^{it\Theta}).$$

Para que esta última ecuación se cumpla, tiene que pasar que $E(e^{it\Theta}) = 0$ o $e^{it2\pi} = 1$. Asumimos que lo último se sostiene, lo cual significa que t tiene que ser un entero. Por lo tanto, para variables aleatorias circulares, la función característica necesita definirse solamente para valores enteros.

Definición 3.1. La **función característica** de un ángulo aleatorio Θ es la sucesión doblemente infinita de números complejos

$$\{\tau_{p,0} : p = 0, \pm 1, \pm 2, \dots\}$$

dada por

$$\tau_{p,0} = E[e^{ip\Theta}] = \int_0^{2\pi} e^{ip\theta} dF(\theta), \quad p = 0, \pm 1, \pm 2, \dots \quad (3.5)$$

Entonces

$$\tau_{p,0} = 1, \quad \bar{\tau}_{p,0} = \tau_{-p,0}, \quad |\tau_{p,0}| \leq 1. \quad (3.6)$$

Los $\tau_{p,0}$ son los coeficientes de Fourier de F . El número complejo $\tau_{p,0}$ es llamado el *p -ésimo momento trigonométrico de Θ alrededor de la dirección cero* y pueden ser representados como

$$\tau_{p,0} = \alpha_p + i\beta_p, \quad (3.7)$$

donde

$$\alpha_p = E[\cos p\Theta] = \int_0^{2\pi} \cos p\theta dF(\theta) \quad (3.8)$$

y

$$\beta_p = E[\sin p\Theta] = \int_0^{2\pi} \sin p\theta dF(\theta). \quad (3.9)$$

Entonces

$$\alpha_{-p} = \alpha_p, \quad \beta_{-p} = -\beta_p, \quad |\alpha_p| \leq 1, \quad |\beta_p| \leq 1, \quad (3.10)$$

Note que $\tau_{p,0}$, α_p y β_p son las versiones poblacionales de $t_{p,0}$, a_p y b_p .

Cada $\tau_{p,0}$ define un vector medio resultante en el plano complejo de longitud

$$\rho_p = |\tau_{p,0}|(\alpha_p^2 + \beta_p^2)^{1/2} \in [0, 1], \quad (3.11)$$

y con dirección

$$\mu_p = \arctan^*(\beta_p, \alpha_p). \quad (3.12)$$

Las cantidades ρ_p y μ_p son los análogos poblacionales de \bar{R}_p y $\bar{\theta}_p$. Los casos especiales ρ_1 y μ_1 son las medidas fundamentales de concentración y localización, respectivamente, y con frecuencia simplemente son denotados como ρ y μ y se refieren inequívocamente como la *longitud media resultante poblacional* y la *dirección media poblacional*.

Para la representación polar,

$$\tau_{p,0} = \rho_p e^{i\mu_p} = \rho_p (\cos \mu_p + i \operatorname{sen} \mu_p), \quad (3.13)$$

y aquí

$$\alpha_p = \rho_p \cos \mu_p, \quad \beta_p = \rho_p \operatorname{sen} \mu_p \quad (3.14)$$

El *p-ésimo momento trigonométrico poblacional alrededor de la dirección media* se define como

$$\tau_{p,\mu} = E [e^{ip(\Theta - \mu)}] = \bar{\alpha}_p + i\bar{\beta}_p, \quad (3.15)$$

donde

$$\bar{\alpha}_p = E [\cos p(\Theta - \mu)], \quad \bar{\beta}_p = E [\operatorname{sen} p(\Theta - \mu)], \quad (3.16)$$

son los análogos poblacionales de \bar{a}_p y \bar{b}_p definidos en la ecuación (3.16) y son referidos como los *momentos coseno y seno p-ésimos poblacionales centrales*. De su definición,

$$\tau_{p,\mu} = \tau_{p,0} e^{-ip\mu} = \rho_p [\cos(\mu_p - p\mu) + i \operatorname{sen}(\mu_p - p\mu)]. \quad (3.17)$$

3.1.3. Medidas poblacionales básicas

En esta sección se considerarán los análogos poblacionales de las medidas circulares muestrales introducidas en las secciones 2.2 - 2.4.

Como hemos visto en la Sección 3.1.2, la *dirección media poblacional* $\mu = \mu_1$, definida en la Ecuación (3.12), es la medida circular básica de *localización*. Es importante destacar que μ no está definida cuando la longitud media resultante ρ toma el valor 0; una distribución fundamental circular en la que $\rho = 0$ es la distribución circular continua que será presentada en la Sección 3.3.1. Cuando μ existe para una distribución dada, el efecto de la rotación de la distribución a través de un ángulo ϕ resulta en la dirección media dada por $\mu + \psi \pmod{2\pi}$, es decir, μ es equivariante bajo rotación.

Una medida circular de localización alternativa es la *dirección mediana poblacional* $\tilde{\mu}$. Formalmente, $\tilde{\mu}$ está definida como cualquier ángulo ψ que minimiza

$$E[\pi - |\pi - |\Theta - \psi||]. \quad (3.18)$$

La dirección mediana no necesariamente es única, pero lo será para cualquier distribución circular cuya representación polar sea unimodal. Cualquier dirección mediana satisface

$$P[\Theta \in [\tilde{\mu}, \tilde{\mu} + \pi)] \geq 1/2, \quad P[\Theta \in (\tilde{\mu} - \pi, \tilde{\mu}]] \geq 1/2. \quad (3.19)$$

Una tercera medida circular de localización es la ***dirección modal poblacional*** $\check{\mu}$. Si Θ es discreta entonces $\check{\mu}$ es la dirección con la mayor probabilidad (en la representación polar de la distribución), mientras que si Θ tiene una densidad f entonces $\check{\mu}$ es la dirección para la cual la representación polar de f alcanza un máximo. Dada esta definición, $\check{\mu}$ no necesariamente es única.

La medida circular fundamental de *concentración*, la ***longitud media resultante***, $\rho = \rho_1 \in [0, 1]$ definida en la Ecuación (3.11), siempre existe. La longitud media resultante se dice que es una medida de concentración invariante bajo rotación así como en reflexión, ya que su valor no cambia bajo la rotación de una distribución a través de un ángulo ψ o bien, bajo la reflexión alrededor de cualquier eje. Cuando $\rho > 0$ (y por lo tanto μ existe),

$$\begin{aligned} \bar{\beta}_1 &= E[\text{sen}(\Theta - \mu)] = E[\text{sen } \Theta \cos \mu - \cos \Theta \text{sen } \mu] = E[\text{sen } \Theta] \frac{\alpha_1}{\rho} - E[\cos \Theta] \frac{\beta_1}{\rho} \\ &= \beta_1 \frac{\alpha_1}{\rho} - \alpha_1 \frac{\beta_1}{\rho} = 0. \end{aligned} \quad (3.20)$$

Con un cálculo similar obtenemos

$$\bar{\alpha}_1 = E[\cos(\Theta - \mu)] = \rho. \quad (3.21)$$

Estas dos importantes identidades son los análogos poblacionales de sus contrapartes muestrales en la Ecuación (2.12).

El análogo de V , definida en la Ecuación (2.14), es la ***varianza circular poblacional***,

$$v = 1 - \rho \in [0, 1]. \quad (3.22)$$

Claramente, $v = 0$ cuando $\rho = 1$, es decir, cuando la distribución tiene un punto de distribución concentrada en $\Theta = \mu$. Similarmente, cuando $v = 1$, $\rho = 0$. Es verdad que $\rho = 0$ para una distribución circular uniforme continua (ver Sección 3.3.1), así como también $\rho = 0$ para cualquier distribución cíclicamente simétrica (ver Pewsey *et al.* (2013)), así que $v = 1$ no puede interpretarse como que necesariamente la distribución tenga alta dispersión.

Las análogos poblacionales de $\hat{\sigma}$ y $\hat{\delta}$, definidas en las Ecuaciones (2.15) y (2.17), son la ***desviación circular estándar poblacional*** y la ***dispersión circular poblacional***

$$\sigma = \{-2 \log(1 - v)\}^{1/2} = \{-2 \log \rho\}^{1/2} \in [0, \infty), \quad \delta = \frac{1 - \rho^2}{2\rho^2}. \quad (3.23)$$

Las medidas de población básica de *sesgo* y *curtosis* son los análogos poblacionales de \bar{b}_2 y \bar{a}_2 discutidos en la Sección 2.4, llamadas las partes seno $\bar{\beta}_2$ y coseno $\bar{\alpha}_2$ del segundo momento central definido en la Ecuación (3.16). Los análogos poblacionales de las medidas estandarizadas \hat{s} y \hat{k} , definidas en las ecuaciones (2.23) y (2.24), son

$$s = \frac{\bar{\beta}_2}{(1 - \rho)^{3/2}}, \quad k = \frac{\bar{\alpha}_2 - \rho^4}{(1 - \rho)^2}. \quad (3.24)$$

3.1.4. Distribuciones simétricas

La distribución de una variable aleatoria lineal se dice que es “simétrica” si hay un único punto, digamos ξ , en la recta real alrededor del cual la reflexión de la distribución es idéntica a la distribución original. Más precisamente, una distribución de este tipo es *reflexivamente simétrica* (alrededor de ξ).

Los modelos clásicos de estadística circular son en su mayoría “reflexivamente simétricos”, pero alrededor de un único *eje*. Más generalmente, si una distribución es reflexivamente simétrica alrededor de $\theta = \psi$ entonces también es reflexivamente simétrica alrededor de $\theta = \psi + \pi$. Las distribuciones multimodales circulares pueden ser reflexivamente simétricas alrededor de más de un eje, y la distribución circular uniforme continua (Sección 3.3.1) es reflexivamente simétrica alrededor de cualquier eje. Este es un punto importante ya que la distribución circular uniforme continua es una distribución límite de muchos modelos circulares. Si Θ tiene una distribución que es reflexivamente simétrica alrededor de ψ , y tiene densidad $f(\theta)$, entonces

$$f(\theta - \psi) = f(\psi - \theta) \quad (3.25)$$

Para distribuciones reflexivamente simétricas unimodales, la dirección media μ y la dirección mediana $\tilde{\mu}$ son idénticas, y son iguales a la dirección modal $\check{\mu}$ cuando esta última es única. Más generalmente, para distribuciones reflexivamente simétricas con $\rho > 0$, todas las partes seno de los momentos centrales, las $\bar{\beta}_p$, son iguales a 0 y se sigue que $\tau_{p,\mu} = \bar{\alpha}_p = \rho_p$. Más aún, si $\mu = 0$ entonces $\bar{\alpha}_p = \alpha_p$ y $\mu_p = 0$.

3.2. Enfoques generales para la generación de distribuciones circulares

Hay varios métodos generales que pueden ser usados para obtener distribuciones circulares. Quizás el enfoque más simple es la *perturbación*. En este enfoque, la función de una densidad circular existente se multiplica por una función apropiada para garantizar que su producto sea también una densidad circular genuina. La distribución Cardiode es un ejemplo de este tipo de construcción.

Un segundo enfoque es el *envolvimiento (wrapping)*. Se considera una variable aleatoria X , definida en la recta real. Envolver la distribución X alrededor de la circunferencia de un círculo unitario produce una variable aleatoria circular

$$\Theta = X \pmod{2\pi}. \quad (3.26)$$

Una implicación práctica importante de esta relación entre X y Θ es que si es posible simular variables de la distribución de X , entonces la simulación de variables de la distribución envolvente de Θ es trivial. Denotando la función de distribución de X por $F_X(x)$, la función de distribución circular de Θ está dada por

$$F_{\Theta}(\theta) = \sum_{-\infty}^{\infty} \{F_X(\theta + 2\pi k) - F_X(2\pi k)\}, \quad 0 \leq \theta \leq 2\pi. \quad (3.27)$$

Por otra parte, si X tiene una función de densidad de probabilidad $f_X(x)$, entonces Θ tiene función de densidad de probabilidad

$$f_{\Theta}(\theta) = \sum_{-\infty}^{\infty} f_X(\theta + 2\pi k). \quad (3.28)$$

Esta última resulta ser una propiedad poco atractiva porque, para todas las distribuciones envolventes circulares continuas conocidas así obtenidas, la suma infinita en 3.28 no se simplifica a una forma cerrada. En consecuencia, las densidades de las distribuciones envolventes son generalmente complicadas de tratar. Las distribuciones Normal y Cauchy envolventes son ejemplos de la utilización de este enfoque.

Mardia y Jupp (1999) en *Directional Statistics* exhiben una serie de características interesantes que distinguen a las distribuciones envolventes, entre ellas el hecho de que la distribución envolvente originada de una suma $X + Y$, es igual a la suma de la envolvente originada por X más la originada por Y , o bien, que si conocemos

la función característica de X entonces obtener los momentos trigonométricos de Θ sería trivial.

Sea $g(x)$ una función densidad con soporte en S_X . Jones (2010) describe un enfoque menos estudiado que consiste en aplicar una transformación a la abscisa x ; esto es lo que se conoce como **transformación de escala**. Es decir, consideramos densidades de la forma

$$f(x) \propto g(t(x)), \quad x \in S_X.$$

Note que $g(t(x))$ no es necesariamente una función de densidad en sí misma: podría no ser integrable. Incluso si lo es, en general, se introduce una nueva constante de normalización y su cálculo explícito puede ser nada fácil. Las distribuciones flexibles Inversa Batschelet para datos circulares son obtenidas utilizando este método.

Finalmente, las distribuciones **proyectadas** pueden ser conseguidas primero obteniendo la representación polar $f(r, \theta)$ de una densidad bivariada lineal $f(x, y)$, y luego integrando sobre r para obtener la densidad marginal de Θ , $f(\theta)$. La distribución normal proyectada es representativa de este enfoque.

3.3. Modelos circulares clásicos

3.3.1. Distribución Circular Uniforme Continua

La *distribución circular uniforme continua*, con densidad

$$f(\theta) = \frac{1}{2\pi}, \quad (3.29)$$

es el más fundamental de los modelos circulares y es apropiado cuando no hay una dirección más probable que cualquier otra. Es la única distribución circular que es invariante bajo rotación y reflexión. Como todas las direcciones son igualmente probables, esta distribución caracteriza la isotropía o aleatoriedad circular.

Integrando (3.29), la distribución de Θ es

$$F(\theta) = \frac{\theta}{2\pi}. \quad (3.30)$$

Para $\phi \leq \psi \leq \phi + 2\pi$, se tiene que

$$P(\phi < \Theta \leq \psi) = F(\psi) - F(\phi) = \frac{\psi - \phi}{2\pi}. \quad (3.31)$$

Se puede demostrar que $\rho = 0$ y por lo tanto μ , y todas las medidas basadas en ella, están indefinidas. Debido a que la representación polar de (3.29) es de forma similar a un anillo, $\tilde{\mu}$ y $\check{\mu}$ son claramente no únicos.

Mardia y Jupp (1999) muestran que la distribución de la suma de n variables aleatorias independientes e idénticamente distribuidas de una distribución circular tiende a una distribución circular uniforme cuando $n \rightarrow \infty$.

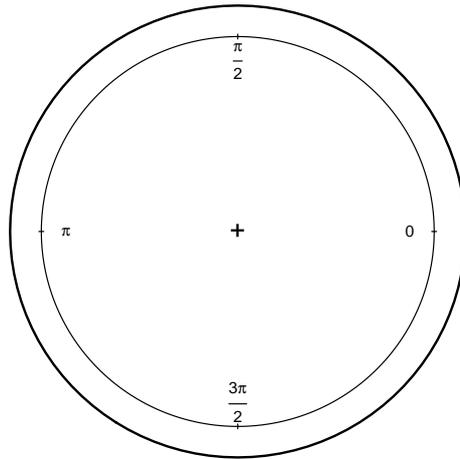


Figura 3.3: Distribución Circular Uniforme Continua.

3.3.2. Distribución Cardioide

La distribución *cardioide* o *coseno*, introducido por Jeffreys (1948), tiene una densidad que surge de la *perturbación coseno* de la densidad circular uniforme continua. Su densidad es

$$f(\theta) = \frac{1}{2\pi} \{1 + 2\rho \cos(\theta - \mu)\}, \quad \rho \in [0, 1/2], \quad (3.32)$$

donde el parámetro ρ representa la longitud media resultante de la distribución. La distribución uniforme es obtenida cuando $\rho = 0$. De otra manera, cuando $0 < \rho \leq 1/2$, la distribución es unimodal y reflexivamente simétrica respecto a $\mu = \tilde{\mu} = \check{\mu}$. Para $0 < \rho < 0.1$, la distribución corresponde a una desviación suave de la uniformidad con un poco más de densidad en la vecindad de μ que alrededor de la antimoda $\mu + \pi$. Integrando (3.32),

$$F(\theta) = \frac{1}{2\pi} \{\theta + 2\rho[\text{sen } \mu + \text{sen}(\theta - \mu)]\}, \quad (3.33)$$

y por lo tanto para $\phi \leq \psi \leq \phi + 2\pi$,

$$P(\phi < \Theta \leq \psi) = F(\psi) - F(\phi) = \frac{1}{2\pi} \{(\psi - \phi) + 2\rho[\text{sen}(\psi - \mu) - \text{sen}(\phi - \mu)]\}.$$

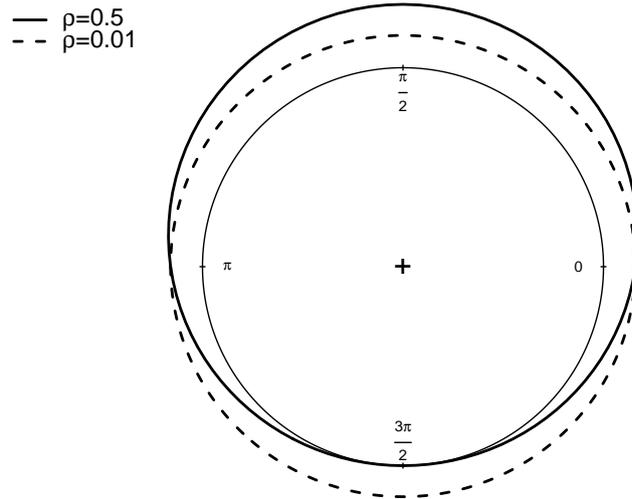


Figura 3.4: Densidades Cardioides con $\mu = \pi/2$ y $\rho = 0.01, 0.5$.

3.3.3. Distribución Potencia del Coseno de Cartwright

Un modelo relacionado con la distribución cardioide es la distribución *potencia del coseno* y su densidad está dada por

$$f(\theta) = \frac{2^{-1+1/\zeta} \Gamma^2(1 + 1/\zeta)}{\pi \Gamma(1 + 2/\zeta)} (1 + \cos(\theta - \mu))^{1/\zeta}, \quad (3.34)$$

donde Γ denota la función gamma y $\zeta > 0$ es un parámetro que controla la concentración.

La distribución Cardiode, con $\rho = 1/2$, se obtiene cuando $\zeta = 1$. Cuando $\zeta \rightarrow 0$ la densidad llega a estar más concentrada alrededor de la dirección media μ , mientras que cuando $\zeta \rightarrow \infty$ la densidad tiende a la distribución circular continua uniforme.

La distribución potencia del coseno fue propuesta por primera vez por Cartwright (1963) como un modelo para los espectros direccionales de las olas oceánicas.

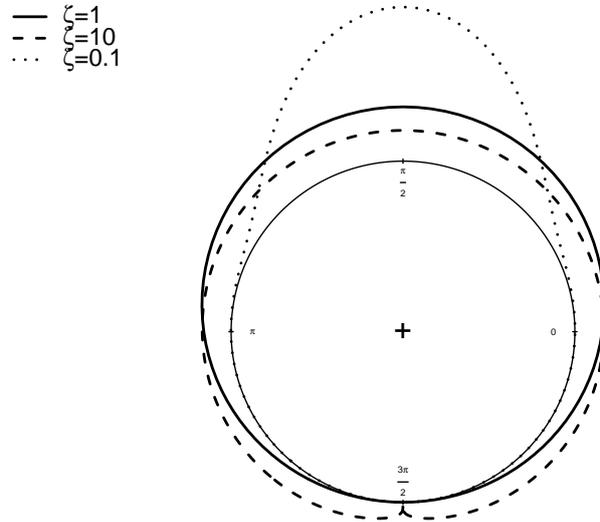


Figura 3.5: Representación polar de densidades Potencia del Coseno con $\mu = \pi/2$ y $\zeta = 1, 10, 0.1$. Cuando $\zeta = 1$, se tiene una distribución cardioide con $\rho = 1/2$.

3.3.4. Distribución Cauchy Envolvente

Sea X una variable aleatoria Cauchy definida en la recta real, entonces tiene densidad

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - \xi)^2}, \quad (3.35)$$

donde $-\infty < \xi < \infty$ es la mediana y la moda de la distribución, pero no tiene media, y $\gamma > 0$ es un parámetro de escala.

Por las propiedades de las distribuciones envolventes (ver Mardia y Jupp (1999)), se tiene que la *distribución Cauchy envolvente*, de la variable aleatoria $\Theta = X \pmod{2\pi}$, tiene densidad

$$f(\theta) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^p \cos p(\theta - \mu) \right\}, \quad (3.36)$$

donde $\rho = e^{-\gamma}$ es la longitud media resultante y $\mu = \xi \pmod{2\pi}$ es la dirección media.

La densidad (3.36) puede demostrarse que puede ser simplificada a

$$f(\theta) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}. \quad (3.37)$$

La distribución circular uniforme continua se obtiene cuando $\rho = 0$. Más generalmente, para $\rho > 0$ la representación polar de la densidad es unimodal y reflexivamente simétrica alrededor de μ .

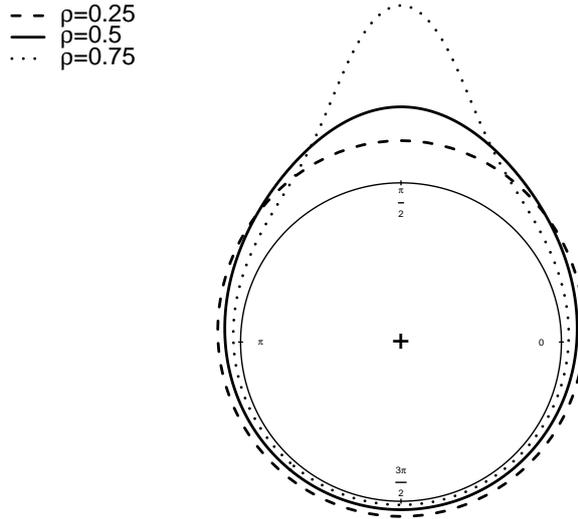


Figura 3.6: Representación polar de densidades Cauchy envolventes con $\mu = \pi/2$ y $\rho = 0.25, 0.5, 0.75$.

3.3.5. Distribución Normal Envolvente

Envolver una variable aleatoria normal con media $\xi \in (-\infty, \infty)$ y varianza $\sigma^2 > 0$, $X \sim N(\xi, \sigma^2)$, alrededor de la circunferencia de un círculo unitario produce una variable aleatoria circular Θ con una *distribución normal envolvente*, cuya densidad está dada por

$$f(\theta) = \frac{1}{(-4\pi \log \rho)^{1/2}} \sum_{k=-\infty}^{\infty} \exp \left\{ \frac{(\theta - \mu + 2\pi k)^2}{4 \log \rho} \right\}, \quad (3.38)$$

donde $\rho = e^{-\sigma^2/2} \in (0, 1)$ es la longitud media resultante y $\mu = \xi \pmod{2\pi}$ es la dirección media. Se puede demostrar que una representación alternativa de la densidad de Θ está dada por

$$f(\theta) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu) \right\}. \quad (3.39)$$

La representación polar de la distribución normal envolvente es unimodal y reflexivamente simétrica alrededor de μ , obteniendo la distribución circular continua uniforme cuando $\rho \rightarrow 0$ y un punto de distribución concentrado en μ cuando $\rho \rightarrow 1$. Como una comparación, en la Figura 3.7 se verifica que dada una longitud media resultante, las densidades normal envolventes son generalmente menos concentradas alrededor de la dirección media, y tiene menos densidad alrededor de la antimoda, que sus contrapartes Cauchy envolventes (ver Figura 3.6).

-- $\rho=0.25$
 — $\rho=0.5$
 ... $\rho=0.75$

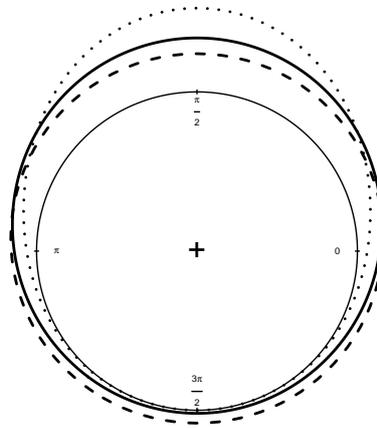


Figura 3.7: Representación polar de densidades Normal envolventes con $\mu = \pi/2$ y $\rho = 0.25, 0.5, 0.75$.

3.3.6. Distribución von Mises

La *distribución von Mises* tiene densidad

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \quad (3.40)$$

donde $k \geq 0$ es llamado el parámetro de concentración,

$$I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos p\theta e^{\kappa \cos \theta} d\theta \quad (3.41)$$

es la función de Bessel modificada de primera clase y orden p , y, cuando $\kappa > 0$, μ denota la dirección media. La aparente simplicidad de (3.40), especialmente cuando es comparada con una densidad envolvente como (3.38), es bastante engañosa. En

lugar de usar una suma infinita, (3.40) requiere el uso de integración numérica para evaluar $I_0(\kappa)$.

La distribución circular continua uniforme se obtiene cuando $\kappa = 0$, y la distribución tiende a un punto de distribución centrado en μ cuando $\kappa \rightarrow \infty$. Cuando $\kappa > 0$, la representación polar de la densidad es unimodal y reflexivamente simétrica con respecto a $\mu = \tilde{\mu} = \check{\mu}$. Por esta simetría, para $k > 0$, se tiene que $\bar{\beta}_p = 0$ y

$$\bar{\alpha}_p = \frac{1}{2\pi I_0(\kappa)} \int_0^{2\pi} \cos p(\theta - \mu) e^{\kappa \cos(\theta - \mu)} d\theta = \frac{I_p(\kappa)}{I_0(\kappa)} = A_p(\kappa). \quad (3.42)$$

Así que cuando $\kappa > 0$, $\tau_{p,\mu} = A_p(\kappa)$ y, usando (3.17), se sigue que

$$\tau_{p,0} = \tau_{p\mu} e^{ip\mu} = A_p(\kappa) e^{ip\mu}. \quad (3.43)$$

Por lo tanto, cuando $\kappa > 0$,

$$\alpha_p = A_p(\kappa) \cos p\mu, \quad \beta_p = A_p(\kappa) \sin p\mu, \quad \rho_p = A_p(\kappa), \quad \mu_p = p\mu. \quad (3.44)$$

En particular, la longitud media resultante es $\rho = \bar{\alpha}_1 = A_1(\kappa)$, y se puede mostrar que $\bar{\alpha}_2 = A_2 = 1 - 2A_1(\kappa)/\kappa$.

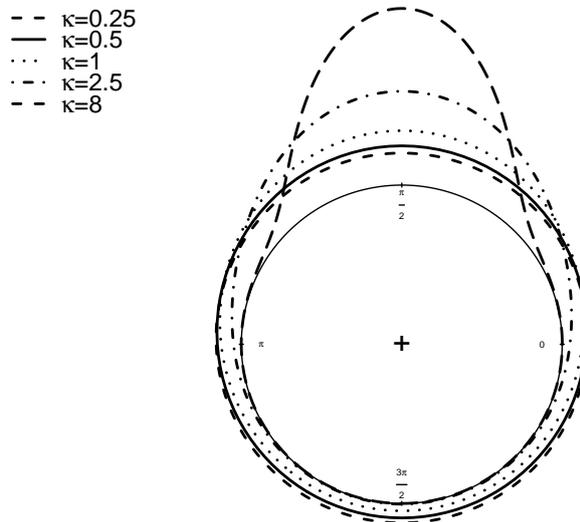


Figura 3.8: Representación polar de densidades von Mises con $\mu = \pi/2$ y $\kappa = 0.25, 0.5, 1, 2.5, 8$.

Pewsey *et al.* (2013) ofrecen la siguiente comparación de modelos clásicos circulares con respecto a la distribución von Mises: Considerando las representaciones polares

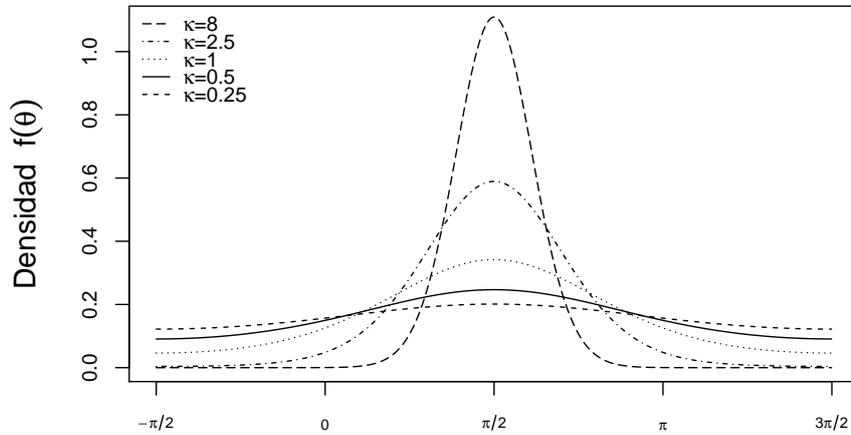


Figura 3.9: Representación lineal de densidades von Mises con $\mu = \pi/2$ y $\kappa = 0.25, 0.5, 1, 2.5, 8$.

de las densidades retratadas en la Figura 3.10, es evidente que las densidades Cardioide, Cauchy envolvente, Normal envolvente y von Mises, con los mismos valores para dirección media y longitud media resultante, son todas muy similares cuando la longitud media resultante ρ es pequeña. Cuando ρ crece, la disparidad entre la Cauchy envolvente y las otras tres densidades incrementa. De entre las otras tres, la densidad Normal envolvente ofrece la aproximación más cercana a la densidad von Mises, a pesar de que generalmente tiene levemente menor densidad alrededor de la moda y ligeramente más densidad de alrededor de los “hombros” que la von Mises. Claramente, para valores grandes de ρ , la Cauchy envolvente ofrece una muy mala aproximación a la von Mises.

Para $0 \leq \theta \leq 2\pi$, los valores de la función de distribución de la distribución von Mises está dada por

$$F(\theta) = \frac{1}{2\pi I_0(k)} \int_0^\theta e^{k \cos(\phi - \mu)} d\phi, \quad (3.45)$$

el cálculo de la función de Bessel y la integral requiere utilizar el método de cuadratura.

La distribución von Mises fue introducida por von Mises (1918) cuando estudiaba la desviaciones de los pesos atómicos medidos con respecto a valores enteros.

En muchos sentidos, la distribución von Mises juega un rol análogo a la de una distribución normal para datos lineales. De hecho, en la literatura es frecuentemente referida como la distribución *normal circular*. Debido a sus propiedades matemáticas atractivas, las técnicas que asumen los datos circulares distribuidos de acuerdo a ella están mucho más desarrolladas que la de cualquier otro tipo de distribución circular. Como consecuencia de ello, y en paralelo con el abuso de la distribución

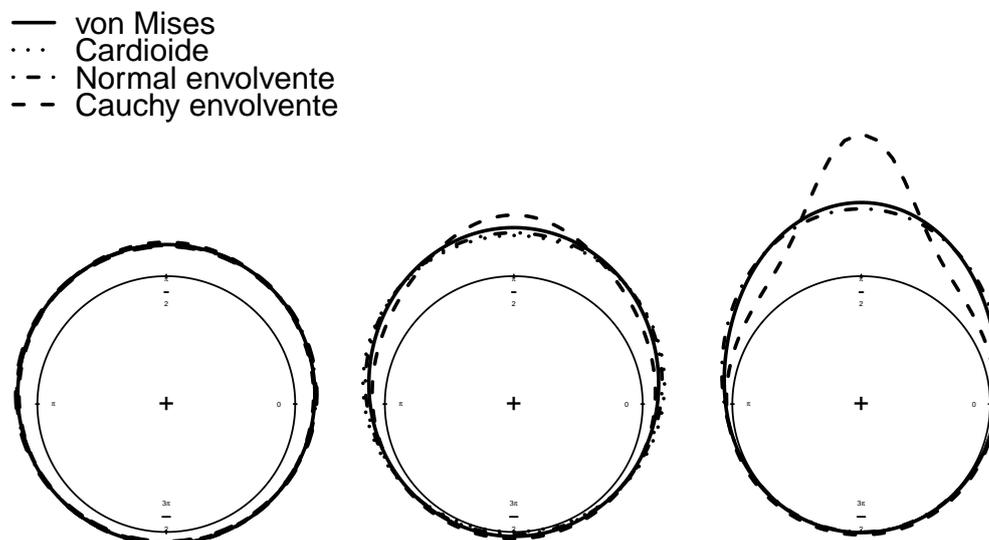


Figura 3.10: Representación polar de las densidades de las distribuciones von Mises, Cardioide, Normal envolvente y Cauchy envolvente con $\mu = \pi/2$ y longitud media resultante $\rho = 0.25$ (izquierda), $\rho = 0.5$ (centro) y $\rho = 0.75$ (derecha).

normal, estas técnicas han sido a menudo aplicadas indiscriminadamente sin la debida consideración de los supuestos que los sustentan.

3.3.6.1 Estimación de parámetros vía máxima verosimilitud

Sea $\theta_1, \dots, \theta_n$ un conjunto de observaciones de una distribución von Mises con parámetros μ y κ . La función de verosimilitud está dada por

$$L(\mu, \kappa) = \frac{1}{[2\pi I_0(\kappa)]^n} e^{\sum_{i=1}^n \kappa \cos(\theta_i - \mu)},$$

y la función log-verosimilitud es,

$$l = \log L = -n \log(2\pi I_0(\kappa)) + \kappa \sum_{i=1}^n \cos(\theta_i - \mu).$$

Al diferenciar esta función con respecto a μ y κ , respectivamente, e igualar a cero, obtenemos las ecuaciones de verosimilitud

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^n \sin(\theta_i - \mu) = 0, \quad (3.46)$$

y,

$$\frac{\partial l}{\partial \kappa} = -n \frac{I_1(\kappa)}{I_0(\kappa)} + \sum_{i=1}^n \cos(\theta_i - \mu) = 0, \quad (3.47)$$

usando el hecho de que $\frac{dI_0(\kappa)}{d\kappa} = I_1(\kappa)$. De la ecuación (3.46), tenemos

$$\sum_{i=1}^n (\sin \theta_i \cos \mu - \cos \theta_i \sin \mu) = 0.$$

$$\text{o, } b_1 \cos \mu - a_1 \sin \mu = 0,$$

así que,

$$\hat{\mu} = \arctan^* \left(\frac{b_1}{a_1} \right) = \bar{\theta}, \quad (3.48)$$

donde \arctan^* está definida en la ecuación (2.5). Sustituyendo esta $\hat{\mu}$ en la otra ecuación (3.47) y notando que $\sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = R$ del Teorema 2.1, obtenemos

$$\begin{aligned} -n \frac{I_1(\kappa)}{I_0(\kappa)} + R &= 0 \\ A_1(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)} &= \frac{R}{n} = \bar{R} \end{aligned} \quad (3.49)$$

entonces,

$$\hat{\kappa} = A_1^{-1}(\bar{R}). \quad (3.50)$$

3.4. Familia de distribuciones simétricas Jones-Pewsey

Jones y Pewsey (2005) proponen una familia de distribuciones reflexivamente simétricas unimodales de tres parámetros, en el círculo, que posee como submodelos a las distribuciones Uniforme, Cardioide, Potencia del Coseno, Cauchy envolvente, y von Mises.

Su densidad cumple que

$$f_\psi(\theta) \propto \{1 + \tanh(\kappa\psi) \cosh(\theta - \mu)\}^{1/\psi}, \quad \mu - \pi < \theta \leq \mu + \pi. \quad (3.51)$$

Aquí, μ es un parámetro de localización, $\kappa \geq 0$ es un parámetro de concentración equivalente al parámetro usual de la distribución von Mises, y el índice $\psi \in \mathbb{R}$ es un parámetro de forma.

El hecho reflejado en (3.51) es utilizado para obtener una constante de normalización para la densidad f_ψ . Para ello, se supone que $\theta \in [-\pi, \pi]$ y, sin pérdida de generalidad, que $\mu = 0$.

De (3.51), se llega a que

$$f_\psi(\theta) \propto \{\cosh(\kappa\psi) + \sinh(\kappa\psi) \cos \theta\}^{1/\psi}. \quad (3.52)$$

Ahora bien, $\cosh(x) \in [1, \infty]$ y $\sinh(x) \in (-\infty, \infty)$. Como $\kappa \geq 0$ y $\psi \in \mathbb{R}$, entonces el signo de $\sinh(\kappa\psi)$ depende de ψ . Sea $z = \cosh(\kappa\psi)$, entonces $\sinh(\kappa\psi) = \pm \sqrt{\cosh(\kappa\psi)^2 - 1} = \text{sgn}(\psi) \sqrt{z^2 - 1}$, entonces de (3.52), llegamos a que $f_\psi(\theta) \propto \{z + \text{sgn}(\psi) \sqrt{z^2 - 1} \cos \theta\}^{1/\psi}$.

Si $\psi > 0$, entonces al considerar la ecuación 3.664.1 de Gradshteyn y Ryzhik (2007) se llega a que

$$\begin{aligned} \int_{-\pi}^{\pi} (z + \sqrt{z^2 - 1} \cos x)^{1/\psi} dx &= 2 \int_0^{\pi} (z + \sqrt{z^2 - 1} \cos x)^{1/\psi} dx \\ &= 2\pi P_{1/\psi}(z), \end{aligned}$$

donde $P_{1/\psi}(z)$ es la función de Legendre asociada de primera clase, de grado $1/\psi$ y orden cero (Gradshteyn y Ryzhik, 2007).

Si $\psi < 0$, entonces al considerar las ecuaciones 3.664.2 y 8.731.5 de Gradshteyn y Ryzhik, se llega a que

$$\begin{aligned} \int_{-\pi}^{\pi} (z - \sqrt{z^2 - 1} \cos x)^{1/\psi} dx &= \int_{-\pi}^{\pi} \frac{1}{(z - \sqrt{z^2 - 1} \cos x)^{-1/\psi}} dx \\ &= 2\pi P_{-(1/\psi)-1}(z) \\ &= 2\pi P_{1/\psi}(z). \end{aligned}$$

Se sigue que la función de densidad de probabilidad de esta familia es

$$f_\psi(\theta) = \frac{\{\cosh(\kappa\psi) + \sinh(\kappa\psi) \cos(\theta)\}^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa\psi))}.$$

En general se tiene que

$$f(\theta) = \frac{\{\cosh(\kappa\psi) + \sinh(\kappa\psi) \cos(\theta - \mu)\}^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa\psi))}. \quad (3.53)$$

Se puede mostrar de manera fácil que la densidad dada en 3.53 es unimodal con moda en $\theta = \mu$ y antimoda en $\mu + \pi$ para toda $\psi \in \mathbb{R}$ y $\kappa > 0$.

Jones y Pewsey (2005) explican a detalle cuando de (3.53) se obtienen los cinco modelos clásicos antes mencionados, al considerar: $\kappa = 0$ o $\psi \rightarrow \pm\infty$ y κ finito (circular uniforme continua); $\psi = 1$ (cardioide); $\psi > 0$ y $\kappa \rightarrow \infty$ (potencia del coseno); $\psi = -1$ (Cauchy envolvente); $\psi \rightarrow 0$ (von Mises).

Las representaciones polares de las densidades mostradas en la Figura 3.11 exhiben muchas de las características principales de las densidades Jones-Pewsey. Para todas, excepto el caso circular uniforme continua (para la cual la dirección media no existe), la representación polar de la densidad es unimodal alrededor de dirección media μ . Cuando κ decrece, el efecto de ψ disminuye y, para un valor fijo de κ , una densidad con $\psi \leq 0$ está más concentrada alrededor de μ y tiene “hombros” más ligeros que una con $\psi \geq 0$. Cuando $\psi > 0$, hay una diferencia relativamente pequeña entre las densidades de distribuciones con diferentes valores de κ .

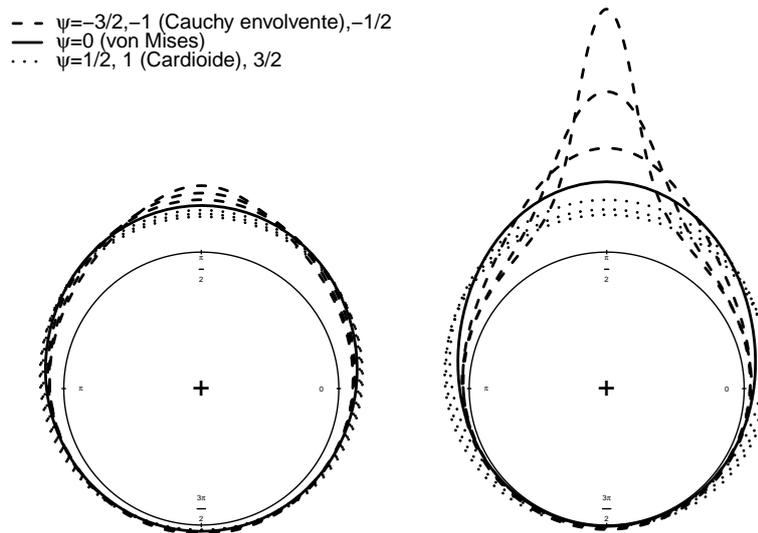


Figura 3.11: Representaciones polares de densidades Jones-Pewsey con $\mu = \pi/2$, $\kappa = 1$ (izquierda) y $\kappa = 2$ (derecha) y, en orden decreciente de la altura de la moda, $\psi = -3/2, -1, -1/2$ (línea discontinua; $\psi = -1$ es una Cauchy envolvente), $\psi = 0$ (línea sólida; von Mises) y $\psi = 1/2, 1, 3/2$ (línea punteada; $\psi = 1$ es cardioide).

3.4.1. Estimación de parámetros vía máxima verosimilitud

Sea $\theta_1, \dots, \theta_n$ una muestra independiente e idénticamente distribuida. La función verosimilitud es

$$l(\mu, \kappa, \psi) = \frac{1}{\psi} \sum_{i=1}^n \log(\cosh(\kappa\psi) + \sinh(\kappa\psi) \cos(\theta_i - \mu)) - n \log \{2Q_{1/\psi}(\kappa\psi)\}, \tag{3.54}$$

donde

$$Q_\nu(t) = \int_0^\pi (\cosh(t) + \sinh(t) \cos x)^\nu dx. \quad (3.55)$$

(Jones y Pewsey, 2005) mencionan que, generalmente, no hay formas cerradas de los estimadores máxima verosimilitud de los tres parámetros, así que métodos numéricos de optimización deben ser usados para obtenerlos vía maximización de la función log-verosimilitud 3.54; o equivalentemente, minimizando la función menos log-verosimilitud.

3.5. Distribuciones Inversas Batschelet

Jones y Pewsey (2012) utilizan el enfoque de la transformación de escala para generar distribuciones que consideran amplios rangos de sesgo y picudez que ya se tienen disponibles. La clave es emplear las inversas de las transformaciones tipo Batschelet, las cuales exhiben ventajas considerables sobre las transformaciones Batschelet directas.

Para generar estas distribuciones, se comienza dando $f(\theta)$, $-\pi < \theta \leq \pi$, una densidad de una distribución circular simétrica unimodal. Entonces, sin pérdida de generalidad, se toma f teniendo moda y centro de simetría en 0, así que $f(\theta) = f(-\theta)$, y tiene antimoda en π .

3.5.1. Familias sesgadas

Utilizando el enfoque de la transformación de escala, se quieren considerar distribuciones circulares unimodales sesgadas con densidades de la forma

$$g_\tau(\theta) \propto f\{\tau(\theta)\}, \quad (3.56)$$

para alguna “función de sesgo” τ monótona apropiada, que depende de ν .

Batschelet (1981) utiliza la transformación

$$\tau = \tau_\nu = \theta + \nu \cos \theta \quad (3.57)$$

en (3.56) en el caso especial cuando f es la densidad de una distribución von Mises. La elección de (3.57) asegura que g_ν sea unimodal si el parámetro de sesgo ν satisface $-1 \leq \nu \leq 1$ (una condición que siempre se empleará para ν de aquí en adelante). Sin embargo, la constante de normalización asociada a g_τ no está

inmediatamente disponible en general y se tiene que recurrir a procedimientos numéricos.

Jones y Pewsey (2012) proponen mejorar este enfoque al cambiar la transformación τ_ν añadiendo ν , cambiando el signo de ν , y entonces, de manera crucial, emplear la inversa de la transformación. Esto es, definen

$$t_\nu(\theta) = t_{1,\nu}^{-1}(\theta) \quad \text{donde} \quad t_{1,\nu}(\theta) = \theta - \nu - \nu \cos \theta, \quad -1 \leq \nu \leq 1. \quad (3.58)$$

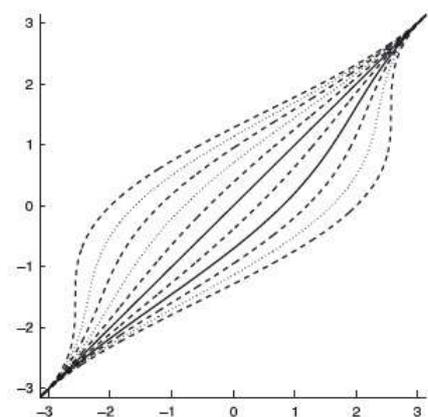


Figura 3.12: Función $t_\nu(\theta) = t_{1,\nu}^{-1}(\theta)$, $\theta \in (-\pi, \pi]$. En orden creciente de la altura en $\theta = 0$, las gráficas corresponden a los valores $\nu = -1, -0.8, \dots, 0.8, 1$, respectivamente. La línea sólida corresponde a $\nu = 0$.

Tanto t_ν como $t_{1,\nu}$ son biyecciones de \mathbb{S}^1 a \mathbb{S}^1 .

Una ventaja de este enfoque es que no hay cambio en la constante de normalización de la de f , es decir, obtenemos densidades del tipo

$$g_\nu(\theta) = f\{t_\nu(\theta)\}. \quad (3.59)$$

Probemos lo anterior. Se usa la sustitución $\omega = t_\nu(\theta)$ para mostrar que

$$\begin{aligned} \int_{-\pi}^{\pi} f\{t_\nu(\theta)\} d\theta &= \int_{-\pi}^{\pi} t'_{1,\nu}(\omega) f(\omega) d\omega \\ &= \int_{-\pi}^{\pi} (1 + \nu \sin \omega) f(\omega) d\omega \\ &= 1 \end{aligned}$$

por la simetría de f . La transformación de escala t_ν permite que esta notable propiedad se mantenga.

La periodicidad de g_ν se debe al siguiente argumento. Primero, notamos que $t_{1,\nu}(\theta + 2k\pi) = t_{1,\nu}(\theta) + 2k\pi$, $k = 0, \pm 1, \dots$. Por lo tanto, $\theta + 2k\pi = t_\nu \{t_{1,\nu}(\theta) + 2k\pi\}$, pero también es igual a $t_\nu \{t_{1,\nu}(\theta)\} + 2k\pi$ por definición. El ajuste $\phi = t_{1,\nu}(\theta)$, con $-\pi \leq \phi \leq \pi$ mediante la biyección natural de $t_{1,\nu}$, muestra que $t_\nu(\phi + 2k\pi) = t_\nu(\phi) + 2k\pi$ y de aquí que

$$g_\nu(\phi + 2k\pi) = f \{t_\nu(\phi + 2k\pi)\} = f \{t_\nu(\phi) + 2k\pi\} = f \{t_\nu(\phi)\} = g_\nu(\phi)$$

como es requerido.

La unimodalidad de g definida mediante (3.56), y por lo tanto en (3.59), es asegurada por cualquier transformación τ monótona creciente. Esto es debido a que los valores de g crecen y luego decrecen, asumiendo los valores de f , pero a una velocidad diferente gobernada por la transformación de escala. La moda de (3.56) está, por lo tanto, en $\tau_{-1}(0)$. Este valor no está disponible explícitamente para distribuciones Batschelet existentes pero está explícitamente disponible en la versión de Jones y Pewsey (2012) ya que $\tau^{-1}(0) = t_{1,\nu}(0)$. Esto es, la moda de (3.59) está en -2ν . La antimoda permanece en π . También es el caso de la densidad asociada con $-\nu$, la cual es la reflexión en cero de la densidad asociada con ν . Para ver esto, notar que $t_{1,-\nu}(\theta) = -t_{1,\nu}(-\theta)$ así que $t_{-\nu}(\theta) = -t_\nu(-\theta)$ y por tanto

$$g_{\theta;-\nu} = f \{t_{-\nu}(\theta)\} = f \{-t_\nu(-\theta)\} = f \{t_\nu(-\theta)\} = g_\nu(-\theta; \nu). \quad (3.60)$$

3.5.2. Familias simétricas más amplias

Si, en lugar de τ_ν dada en 3.57, τ está dada por

$$\tau(\theta) = \tau_\lambda(\theta) = \theta + \lambda \sin \theta, \quad (3.61)$$

donde $-1 \leq \lambda \leq 1$ es un parámetro que controla la forma, particularmente la picudez, entonces se obtiene una familia de distribuciones simétricas circulares que incluye ejemplos de distribuciones de cimas planas conocidos. Sin embargo, usando la inversa de la transformación de escala Batschelet, resultan numerosas densidades fuertemente puntiagudas pero sin los casos de cimas planas más conocidos.

Un rango más amplio de densidades con formas simétricas está disponible a través de una transformación diferente de escala, $t_\lambda(\theta)$, la cual incluye la transformación directa más “cima-aplanada” $\theta - \sin \theta$ cuando $\lambda = -1$ y la transformación directa inversa más “agudamente-puntiaguda” $(\theta - \sin \theta)^{-1}$ cuando $\lambda = 1$ (o bien θ cuando $\lambda = 0$). Esta transformación está dada por

$$t_\lambda(\theta) = \frac{1 - \lambda}{1 + \lambda} \theta + \frac{2\lambda}{1 + \lambda} s_\lambda^{-1}(\theta), \quad -1 < \lambda \leq 1, \quad (3.62)$$

donde

$$s_\lambda(t) = t - \frac{1}{2}(1 + \lambda) \operatorname{sen} t \quad (3.63)$$

((3.63) es un subconjunto de (3.61)). La definición se complementa al permitir que $t_{-1}(\theta) = \lim_{\lambda \rightarrow -1} t_\lambda(\theta)$ es cual se muestra fácilmente que es igual a $\theta - \operatorname{sen} \theta$.

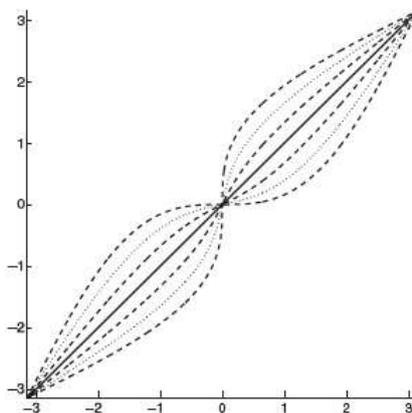


Figura 3.13: Función $t_\lambda(\theta)$, $\theta \in (-\pi, \pi]$. En orden creciente de la altura en $\theta = \pi/2$, las gráficas corresponden a los valores $\lambda = -1, -\frac{2}{3}, \dots, \frac{2}{3}, 1$, respectivamente. La línea sólida corresponde a $\lambda = 0$.

La transformación $t_\lambda : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ es inyectiva creciente tal que $t_\lambda(0) = 0$ y $t_\lambda(\pm\pi) = \pi$ y que $t_\lambda(\theta + 2k\pi) = t_\lambda(\theta) + 2k\pi$. También se cumple que $t_\lambda^{-1}(\theta) = t_{-\lambda}(\theta)$. La clave para probar esto es que, como es fácil ver,

$$t_\lambda \{s_\lambda(\theta)\} = s_{-\lambda}(\theta). \quad (3.64)$$

Entonces, poniendo $-\lambda$ a λ de (3.64) tenemos que $s_{-\lambda}(\theta) = t_{-\lambda}^{-1} \{s_\lambda(\theta)\}$ la cual, cuando comparamos con (3.64), produce el resultado.

Supongamos ahora que f involucra un parámetro de concentración $\kappa > 0$. Entonces, la nueva familia simétrica tiene densidad

$$g_\lambda(\theta) = K_{\kappa, \lambda}^{-1} f \{t_\lambda(\theta)\}. \quad (3.65)$$

Esta familia es, de hecho, unimodal con moda en 0 y antimoda en π . Después de las sustituciones pertinentes, Jones y Pewsey (2012) muestran que

$$\begin{aligned} K_{\kappa, \lambda} &= \int_{-\pi}^{\pi} \left\{ 1 - \frac{1}{2}(1 + \lambda) \cos \theta \right\} f \left\{ \theta - \frac{1}{2}(1 + \lambda) \operatorname{sen} \theta \right\} d\theta \\ &= \frac{1 + \lambda}{1 - \lambda} - \frac{2\lambda}{(1 - \lambda)} \int_{-\pi}^{\pi} f \left\{ \theta - \frac{1}{2}(1 - \lambda) \operatorname{sen} \theta \right\} d\theta \end{aligned} \quad (3.66)$$

para $\lambda < 1$, con $K_{\kappa, 1} = 1 - E_f(\cos \Theta)$. Esta constante normalizadora requiere tratamiento numérico.

3.5.3. Familias simétricas-sesgadas

El mecanismo de sesgo en (3.58) puede ser aplicado a cualquier distribución circular simétrica, pero es especialmente atractivo aplicarlo a la familia (3.65) para intentar ofrecer el más amplio rango de valores posibles que den paso a diversas formas de densidad unimodales simétricas o sesgadas. Las densidades resultantes tienen la forma

$$g_{\nu,\lambda}(\theta) = K_{\kappa,\lambda}^{-1} f [t_\lambda \{t_\nu(\theta)\}], \quad (3.67)$$

donde $K_{\kappa,\lambda}$ está dada por (3.66). En el caso de que f sea una densidad von Mises, (3.67) queda de la forma

$$g_{VM;\nu,\lambda}(\theta) = \frac{1}{2\pi I_0(\kappa) K_{\kappa,\lambda}} \exp \left(\kappa \cos \left[\frac{1-\lambda}{1+\lambda} t_\nu(\theta) + \frac{2\lambda}{1+\lambda} s_\lambda^{-1} \{t_\nu(\theta)\} \right] \right). \quad (3.68)$$

Aquí, s_λ está dada por (3.63), t_ν por (3.58) y

$$K_{\kappa,\lambda} = \frac{1+\lambda}{1-\lambda} - \frac{2\lambda}{(1-\lambda)2\pi I_0(\kappa)} \int_{-\pi}^{\pi} \exp \left[\kappa \cos \left\{ \theta - \frac{1}{2}(1-\lambda) \sin \theta \right\} \right] d\theta \quad (3.69)$$

para $\lambda < 1$, con $K_{\kappa,1} = 1 - \{I_1(\kappa)/I_0(\kappa)\} = 1 - A_1(\kappa)$.

Las propiedades de las densidades (3.67) se siguen de las densidades (3.58) y (3.65). Las densidades de la forma (3.67) permanecen periódicas, son unimodales con moda en $-\nu$ y satisfacen la propiedad reflexiva (3.60).

3.5.4. Estimación máxima verosimilitud

Para un conjunto de datos i.i.d $\theta_1, \dots, \theta_n$, la función log-verosimilitud de la familia completa dada por (3.68) tiene la forma

$$l(\mu, \nu, \kappa, \lambda) = c - n \log \{I_0(\kappa) K_{\kappa,\lambda}\} + \sum_{i=1}^n \kappa \cos \left[\frac{1-\lambda}{1+\lambda} t_\nu(\theta_i - \mu) + \frac{2\lambda}{1+\lambda} s_\lambda^{-1} \{t_\nu(\theta_i - \mu)\} \right], \quad (3.70)$$

donde s_λ, t_ν y $K_{\kappa,\lambda}$ están dadas por (3.63), (3.58) y (3.69), respectivamente, y c es una constante. La maximización de (3.70) en general debe llevarse a cabo de forma numérica. Jones y Pewsey (2012) encontraron que el algoritmo simplex de búsqueda directa de Nelder y Mead (1965) llevan a cabo este proceso con éxito. La optimización puede ser relativamente lenta debido a las dos operaciones inversas contenidos en (3.70).

3.5.5. Distribución Inversa Batschelet con base von Mises

A manera de resumen, Pewsey *et al.* (2013) recalcan que distribuciones Inversa Batschelet, partiendo de una base reflexivamente simétrica von Mises, se generan usando el enfoque de la transformación de escala, el cual involucra la combinación de las inversas de las funciones

$$t_{1,\nu}(\theta) = \theta - \nu(1 + \cos \theta),$$

$$t_\lambda(\theta) = \begin{cases} \frac{1-\lambda}{1+\lambda}\theta + \frac{2\lambda}{1+\lambda}s_\lambda^{-1}(\theta), & -1 < \lambda \leq 1, \\ \theta - \text{sen } \theta, & \lambda = -1, \end{cases}$$

donde

$$s_\lambda(\theta) = t - \frac{1}{2}(1 + \lambda) \text{sen } \theta,$$

$-1 \leq \nu \leq 1$ es un parámetro de sesgo y $-1 \leq \lambda \leq 1$ es un parámetro que regula la picudez de la distribución.

El resultado es una familia reflexivamente simétrica unimodal de distribuciones, que consideran amplios rangos para los valores de sesgo y picudez, con densidad

$$f(\theta) = \frac{1}{2\pi I_0(\kappa) K_{\kappa,\lambda}} \exp \left(\kappa \cos \left[\frac{1-\lambda}{1+\lambda} t_\nu(\theta - \xi) + \frac{2\lambda}{1+\lambda} s_\lambda^{-1} \{t_\nu(\theta - \xi)\} \right] \right), \quad (3.71)$$

donde $t_\nu(\theta) = t_{1,\nu}^{-1}(\theta)$, ξ es un parámetro de localización, $\kappa \geq 0$ es el parámetro de concentración,

$$K_{\kappa,\lambda} = \begin{cases} \frac{1+\lambda}{1-\lambda} - \frac{2\lambda}{(1-\lambda)2\pi I_0(\kappa)} \int_{-\pi}^{\pi} \exp \left[\kappa \cos \left\{ \theta - \frac{1}{2}(1-\lambda) \text{sen } \theta \right\} \right] d\theta, & -1 \leq \lambda < 1, \\ 1 - A_1(\kappa), & \lambda = 1, \end{cases}$$

e $I_0(\kappa)$ y $A_1(\kappa)$ es definidas en 3.41 y 3.42, respectivamente. La inversión de las funciones $t_{1,\nu}$ y s_λ , así como el calculo de la constante $K_{\kappa,\lambda}$, generalmente, tienen que ser desarrollados numéricamente. Cuando $\nu = 0$ y $\kappa > 0$, la dirección media es $\xi \pmod{2\pi}$. Más generalmente, cuando $\kappa > 0$, la moda es $\xi - 2\nu \pmod{2\pi}$.

Los roles de ν y λ son ilustrados en las densidades retratadas en la Figura 3.14. La densidad para un valor negativo de ν es simplemente la reflexión alrededor de ξ de la densidad con el correspondiente valor positivo de ν .

4

Aproximaciones para muestras grandes, pruebas de clasificación, bondad de ajuste y estimación insesgada

En este capítulo se presentan algunos resultados distribucionales para muestras grandes, que sirven como base para obtener pruebas de uniformidad y simetría reflexiva alrededor de la dirección media. Dichas pruebas sirven para clasificar modelos y elegir uno que se ajuste razonablemente a una muestra simple de datos circulares. La bondad de ajuste de un modelo es evaluada utilizando el análogo circular a la transformación integral de probabilidad, cuya aplicación da lugar a las pruebas correspondientes. Por último, se ofrecen métodos de distribución libre de inferencia para los principales estadísticos descriptivos circulares.

4.1. Aproximaciones para muestras grandes

4.1.1. Distribución conjunta para a_1 y b_1

Dado una muestra $\theta_1, \dots, \theta_n$, muchos de los estadísticos circulares que se han presentado pueden ser puestos en términos de (a_1, b_1) . Ya que todos los momentos trigonométricos poblacionales están acotados, entonces el Teorema del Límite Central se puede aplicar para dar una distribución conjunta de (a_1, b_1) (Jammalamadaka y Sengupta, 2001). Supongamos entonces que $\theta_1, \dots, \theta_n$ son ángulos aleatorios i.i.d

en $[0, 2\pi)$. Hay que tener en cuenta los elementos seno y coseno de los primeros dos momentos trigonométricos centrales, es decir,

$$\tau_{1,0} = \alpha_1 + i\beta_1, \quad \tau_{1,0} = \alpha_2 + i\beta_2. \quad (4.1)$$

Sea $\mathbf{X}_i = (\cos \theta_i, \text{sen } \theta_i)$. Al aplicar el Teorema del Límite Central multivariado a los i.i.d vectores unitarios \mathbf{X}_j , $j = 1, \dots, n$, tenemos que

$$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim N_2(\mathbf{0}, \boldsymbol{\sigma}).$$

Aquí,

$$\bar{\mathbf{X}}^T = \sum_{j=1}^n \frac{\mathbf{X}_j^T}{n} = (a_1, b_1)^T,$$

y

$$\boldsymbol{\mu}^T = E(\mathbf{X}^T) = (E(\cos \theta), E(\text{sen } \theta))^T = (\alpha_1, \beta_1)^T,$$

entonces

$$\sqrt{n} [(a_1, b_1)^T - (\alpha_1, \beta_1)^T] \sim N_2(\mathbf{0}, \boldsymbol{\sigma}). \quad (4.2)$$

Los elementos de la matriz de covarianzas $\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ pueden ser identificados vía los primeros dos momentos trigonométricos dados en 4.1. Usando las identidades

$$\cos 2\theta = 2 \cos^2 \theta - 1 = 1 - 2 \text{sen}^2 \theta,$$

y

$$\text{sen } 2\alpha = 2 \text{sen } \theta \cos \theta,$$

llegamos a que

$$\sigma_{11} = \text{var}(\cos \theta) = \frac{1}{2}(1 + \alpha_2 - 2\alpha_1^2),$$

$$\sigma_{12} = \sigma_{21} = \text{cov}(\cos \theta, \text{sen } \theta) = \frac{1}{2}(\beta_2 - 2\alpha_1\beta_1) \quad (4.3)$$

y

$$\sigma_{22} = \text{var}(\text{sen } \theta) = \frac{1}{2}(1 - \alpha_2 - 2\beta_1^2).$$

Así, la distribución conjunta de a_1 y b_1 es asintóticamente normal y, su vector de medias, y su matriz de varianzas-covarianzas están dadas por

$$\boldsymbol{\mu}^T = E [(a_1, b_1)^T] = (\alpha_1, \beta_1)^T \quad (4.4)$$

y $\Sigma = \frac{1}{n}\sigma$, o sea,

$$n\Sigma = n \begin{pmatrix} \text{var}(a_1) & \text{cov}(a_1, b_1) \\ \text{cov}(b_1, a_1) & \text{var}(b_1) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 + \alpha_2 - 2\alpha_1^2 & \beta_2 - 2\alpha_1\beta_1 \\ \beta_2 - 2\alpha_1\beta_1 & 1 - \alpha_2 - 2\beta_1^2 \end{pmatrix}. \quad (4.5)$$

En particular, las distribuciones marginales asintóticas de a_1 y b_1 son normales.

Para encontrar la distribución asintótica de estadísticos que pueden ser expresados en términos de (a_1, b_1) , se puede aplicar el llamado *método delta*, el cual puede ser visto a detalle, en su forma multivariante (y por lo tanto, la más general) en Rao (1973). El método delta es un método que permite derivar un aproximado de la distribución de probabilidad para una función de un estimador estadístico asintóticamente normal; la forma del método delta cuando $g : \mathbb{R}^k \rightarrow \mathbb{R}$ se presenta a continuación.

Lema 4.1. *Suponer que $\mathbf{Y}_n = (Y_{n1}, \dots, Y_{nk})$ es una sucesión de vectores aleatorios tales que*

$$\sqrt{n}(\mathbf{Y}_n - \boldsymbol{\mu}) \sim N_k(\mathbf{0}, \Sigma).$$

Sea $g : \mathbb{R}^k \rightarrow \mathbb{R}$ y sea

$$\nabla g(\mathbf{Y}) = \left(\frac{\partial g}{\partial Y_1}, \dots, \frac{\partial g}{\partial Y_k} \right)^T.$$

Sea $\nabla \boldsymbol{\mu}$ que denota $\nabla g(\mathbf{Y})$ evaluada en $\mathbf{Y} = \boldsymbol{\mu}$ y se asume que los elementos de $\nabla \boldsymbol{\mu}$ son distintos de cero, entonces $\sqrt{n}(g(\mathbf{Y}_n) - g(\boldsymbol{\mu}))$ tiene una distribución asintóticamente normal con media cero y varianza $\nabla \boldsymbol{\mu}^T \Sigma \nabla \boldsymbol{\mu}$.

4.2. Inferencia básica para una muestra simple

En el análisis exploratorio de los datos, dos supuestos resultan de interés para elegir el mejor modelo que se ajuste a ellos. Las hipótesis de uniformidad y simetría reflexiva se utilizan para discriminar modelos, exhibiendo aquellos que son más razonables de considerar en primera instancia.

4.2.1. Pruebas de uniformidad

La uniformidad es el más básico de los supuestos para clasificar modelos. Si no se rechaza, entonces no habría necesidad de considerar modelos complicados, pues la población, de la cual se ha extraído una muestra, puede razonablemente suponerse que está uniformemente distribuida alrededor del círculo.

Prueba de Rayleigh

Considera una alternativa unimodal y se basa en la idea intuitiva de rechazar uniformidad cuando la longitud media muestral resultante \bar{R} es grande. Si \bar{R} es más grande que un valor límite, los datos son interpretados como demasiado concentrados para ser consistentes con la uniformidad y la prueba de Rayleigh rechaza esta hipótesis nula.

La prueba de Rayleigh es conocida por ser la prueba de razón de verosimilitud de uniformidad contra alternativas von Mises. Siendo así, se sigue del lema de Neyman-Pearson que la prueba de Rayleigh es la prueba más potente contra estas alternativas. También, ya que el estadístico \bar{R} es invariante bajo rotación y reflexión de los datos, la prueba es invariante bajo estas operaciones. Por lo tanto, la prueba de Rayleigh es la prueba más potente invariante contra alternativas von Mises (Mardia y Jupp, 1999).

Se pueden establecer estadísticos de prueba para el test de Rayleigh cuando, del Teorema del Límite Central y de los elementos coseno a_1 y seno b_1 bajo uniformidad, se puede demostrar que $2n\bar{R}^2 \sim \chi_2^2$.

Afirmación 4.1. *Dada una muestra $\theta_1, \theta_2, \dots, \theta_n$ de una distribución circular uniforme, es decir, $(\theta_1, \dots, \theta_n) \sim U(0, 2\pi)$, si n es suficientemente grande, entonces $2n\bar{R}^2 \sim \chi_2^2$.*

Demostración: De la sección 4.1.1, se sabe que $(a_1, b_1)^T \sim N_2(\boldsymbol{\mu}^T, \boldsymbol{\Sigma})$, con media y varianzas dadas por (4.4) y (4.5).

Ahora, bajo uniformidad, se tiene que

$$\alpha_1 = E(\cos \theta) = \int_0^{2\pi} \cos \theta dF(\theta) = \int_0^{2\pi} \frac{1}{2\pi} \cos \theta d\theta = 0.$$

Similiaramente, se puede verificar que $\beta_1 = \alpha_2 = \beta_2 = 0$. De (4.4) tenemos que $\boldsymbol{\mu}^T = (\alpha_1, \beta_1)^T = (0, 0)^T$. De (4.5) se tiene que

$$n\boldsymbol{\Sigma} = \frac{1}{2} \begin{pmatrix} 1 + \alpha_2 - 2\alpha_1^2 & \beta_2 - 2\alpha_1\beta_1 \\ \beta_2 - 2\alpha_1\beta_1 & 1 - \alpha_2 - 2\beta_1^2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

es decir,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{2n} & 0 \\ 0 & \frac{1}{2n} \end{pmatrix}.$$

Es interesante notar que a_1 y b_1 llegan a ser asintóticamente independientes. Entonces, ambos $a_1, b_1 \sim N(0, \frac{1}{2n})$, es decir, $\sqrt{2na_1}, \sqrt{2nb_1} \sim N(0, 1)$, y entonces

$2na_1^2, 2nb_1^2 \sim \chi_1^2$. Como $2na_1^2 + 2nb_1^2 \sim \chi_2^2$ y $2na_1^2 + 2nb_1^2 = 2n(a_1^2 + b_1^2) = 2n\bar{R}^2$, entonces $2n\bar{R}^2 \sim \chi_2^2$ ■.

Sin embargo, usando la afirmación 4.1, Jammalamadaka y Sengupta (2001) exhiben una distribución muestral de $R = n\bar{R}$ para n grande. De la afirmación anterior, sea $Z \equiv 2n\bar{R}^2 = \frac{2R^2}{n}$, entonces

$$f(r) = g(z) \left| \frac{dz}{dr} \right| = \frac{1}{2} e^{-\frac{r^2}{n}} \frac{4r}{n} = \frac{2r}{n} e^{-\frac{r^2}{n}}, \quad (4.6)$$

con su correspondiente función de distribución

$$F_n(r) = 1 - e^{-\frac{r^2}{n}}, \quad 0 \leq r \leq n. \quad (4.7)$$

La anterior es conocida como la aproximación de Rayleigh para la longitud muestral resultante y es válida para n grande. Con estas consideraciones, podemos obtener estadísticos de prueba para test de Rayleigh siempre que n sea grande.

Omnibus tests

Los *omnibus test* son aquellos que consideran todas las alternativas, es decir, consideran desviaciones multimodales de la uniformidad. Éstos incluyen las pruebas A_n de Anje, V_n de Kuiper, U^2 de Watson y la de espaciamento de Rao. Los omnibus tests para uniformidad más comúnmente usados son más poderosos que la prueba de Rayleigh cuando se trata de detectar desviaciones multimodales de la uniformidad. Se sugiere usar omnibus tests cuando no haya una clara razón *a priori* para sospechar de desviaciones unimodales de la uniformidad más que salidas multimodales, o bien, cuando se esté más interesado en considerar solo desviaciones unimodales (Pewsey *et al.*, 2013). Mardia y Jupp (1999) describen con detalle algunas de estas pruebas.

4.2.2. Pruebas de bondad de ajuste

Transformación integral de probabilidad. La consideración de la transformación integral de probabilidad en la recta sugiere el siguiente análogo en el círculo. Sea F la función de distribución acumulada de una distribución circular y supongamos que han sido seleccionados una dirección y una orientación iniciales. Entonces la *transformación integral de probabilidad* de la distribución es la transformación del

círculo en el cual se envía θ a $2\pi F(\theta)$. Si F es continua entonces la variable aleatoria transformada

$$U = 2\pi F(\theta) \text{ mod } 2\pi \quad (4.8)$$

está distribuida uniforme en el círculo. Por medio de la transformación integral de probabilidad 4.8, cualquier prueba de uniformidad en el círculo da lugar a una correspondiente prueba de bondad de ajuste (Mardia y Jupp, 1999).

La bondad de ajuste de una distribución propuesta con función de distribución $F(\theta)$ puede ser probada al calcular los valores $2\pi F(\theta_1), \dots, 2\pi F(\theta_n)$ y aplicarles cualquier prueba de uniformidad circular. Si los datos provienen de la distribución propuesta, entonces esperaríamos que la uniformidad circular no sea rechazada.

4.2.3. Prueba de simetría reflexiva para muestras grandes

A continuación se ofrece el siguiente resultado debido a Pewsey (2004).

Teorema 4.1. *Sea $\theta_1, \dots, \theta_n$ una muestra aleatoria de n observaciones de una distribución circular con $\rho \in (0, 1)$. Para una muestra grande, la distribución conjunta asintótica de $\vec{\zeta} = (\bar{\theta}, \bar{R}, \bar{b}_2, \bar{a}_2)^T$ es normal multivariada con vector de medias $\vec{\xi}$ y matriz de covarianza $\vec{\Sigma}$, donde, a orden $O(n^{-3/2})$,*

$$\vec{\xi} = \left[\mu - \frac{\bar{\beta}_2}{2n\rho^2}, \quad \rho + \frac{(1 - \bar{\alpha}_2)}{4n\rho}, \quad \bar{\beta}_2 + \frac{1}{n\rho} \left(-\bar{\beta}_3 - \frac{\bar{\beta}_2}{\rho} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^3} \right), \right. \\ \left. \bar{\alpha}_2 + \frac{1}{n} \left\{ 1 - \frac{\bar{\alpha}_3}{\rho} - \frac{\bar{\alpha}_2(1 - \bar{\alpha}_2) + \bar{\beta}_2^2}{\rho^2} \right\} \right]^T, \quad (4.9)$$

y $2n\rho^2\vec{\Sigma} = (v_{ij} : i, j = 1, \dots, 4)$, donde

$$\begin{aligned} v_{11} &= 1 - \bar{\alpha}_2, \\ v_{12} &= v_{21} = \rho\bar{\beta}_2, \\ v_{13} &= v_{31} = \rho^2 - \rho\bar{\alpha}_3 - 2\bar{\alpha}_2(1 - \bar{\alpha}_2), \\ v_{14} &= v_{41} = \rho\bar{\beta}_3 + 2\bar{\beta}_2(1 - \bar{\alpha}_2), \\ v_{22} &= \rho^2(1 - 2\rho^2 + \bar{\alpha}_2), \\ v_{23} &= v_{32} = -2\rho^3\bar{\beta}_2 + \rho^2\bar{\beta}_3 - 2\rho\bar{\alpha}_2\bar{\beta}_2, \\ v_{24} &= v_{42} = \rho^3(1 - 2\bar{\alpha}_2) + \rho^2\bar{\alpha}_3 + 2\rho\bar{\beta}_2^2, \\ v_{33} &= \rho^2(1 - 4\bar{\alpha}_2 - 2\bar{\beta}_2^2 - \bar{\alpha}_4) + 4\rho\bar{\alpha}_2\bar{\alpha}_3 + 4\bar{\alpha}_2^2(1 - \bar{\alpha}_2), \\ v_{34} &= \rho^2 \{ 2(1 - \bar{\alpha}_2)\bar{\beta}_2 + \bar{\beta}_4 \} - 2\rho(\bar{\alpha}_2\bar{\beta}_3 + \bar{\beta}_2\bar{\alpha}_3) - 4\bar{\alpha}_2\bar{\beta}_2(1 - \bar{\alpha}_2), \end{aligned}$$

$$\begin{aligned} v_{43} &= v_{34}, \\ v_{44} &= \rho^2(1 - 2\bar{\alpha}_2^2 + \bar{\alpha}_4) + 4\rho\bar{\beta}_2\bar{\beta}_3 + 4\bar{\beta}_2^2(1 - \bar{\alpha}_2). \end{aligned} \quad (4.10)$$

Demostración: Pewsey (2004) procede con una idea análoga a Mardia y Jupp (1999) cuando éstos consiguen una distribución conjunta de $(\bar{\theta}, \bar{R})^T$ utilizando la normalidad conjunta de $(a_1, b_1)^T$ (ver Sección 4.8 de Jammalamadaka y Sengupta (2001)).

Aplicando el Teorema del Límite Central, se puede demostrar que $(a_1, b_1, a_2, b_2)^T$ tiene como distribución asintótica una $N_4(\vec{\zeta}, \vec{\Psi})$ donde $\vec{\zeta} = (\alpha_1, \beta_1, \alpha_2, \beta_2)$ y $2n\vec{\Psi} = (\tau_{ij} : i, j = 1, \dots, 4)$ donde

$$\begin{aligned} \tau_{11} &= 1 + \alpha_2 - 2\alpha_1^2, \\ \tau_{12} &= \tau_{21} = \beta_2 - 2\alpha_1\beta_1, \\ \tau_{13} &= \tau_{31} = \alpha_1 + \alpha_3 - 2\alpha_1\alpha_2, \\ \tau_{14} &= \tau_{41} = \beta_1 + \beta_3 - 2\alpha_1\beta_2, \\ \tau_{22} &= 1 - \alpha_2 - 2\beta_1^2, \\ \tau_{23} &= \tau_{32} = \beta_3 - \beta_1 - 2\beta_1\alpha_2, \\ \tau_{24} &= \tau_{42} = \alpha_1 - \alpha_3 - 2\beta_1\beta_2, \\ \tau_{33} &= 1 + \alpha_4 - 2\alpha_2^2, \\ \tau_{34} &= \tau_{43} = \beta_4 - 2\alpha_2\beta_2, \\ \tau_{44} &= 1 - \alpha_4 - 2\beta_2^2. \end{aligned}$$

Ahora, como $p \neq 0$, podemos asumir, después de una rotación adecuada, que $\alpha_1 > 0$, así que, con probabilidad 1, $\bar{\theta} = \arctan(b_1/a_1)$. Además, los otros tres estadísticos,

$$\begin{aligned} \bar{R} &= (a_1^2 + b_1^2)^{1/2}, \\ \bar{b}_2 &= \frac{(\alpha_1^2 - b_1^2)b_2 - 2a_1b_1a_2}{a_1^2 + b_1^2} \end{aligned}$$

y

$$\bar{a}_2 = \frac{(a_1^2 - b_1^2)a_2 + 2a_1b_1b_2}{a_1^2 + b_1^2},$$

son también funciones diferenciables de a_1, b_1, a_2 y b_2 .

Considerando el método delta general, por medio de la expansión de Taylor, uno obtiene, después de una álgebra sencilla pero tediosa, los resultados para $\vec{\xi}$ y $\vec{\Sigma}$

en 4.9 y 4.10. La normalidad asintótica se sigue de la normalidad asintótica de $(a_1, b_1, a_2, b_2)^T$ ■.

El teorema 4.1 ofrece una base teórica para la prueba de simetría circular exhibida, también, por Pewsey (2002), la cual tiene como soporte el siguiente teorema,

Teorema 4.2. *Sea $\theta_1, \dots, \theta_n$ n variables aleatorias independientes e idénticamente distribuidas de una distribución angular con longitud media resultante $\rho \in (0, 1)$. Ya que $\rho \neq 0$, la dirección media, μ , existe y es única. Entonces*

$$E(\bar{b}_2) = \bar{\beta}_2 + \frac{1}{n} \left(-\frac{\bar{\beta}_3}{\rho} - \frac{\bar{\beta}_2}{\rho^2} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4} \right) + O(n^{-3/2}) \quad (4.11)$$

y

$$\text{var}(\bar{b}_2) = \frac{1}{n} \left[\frac{1 - \bar{\alpha}_4}{2} - 2\bar{\alpha}_2 - \bar{\beta}_2^2 + \frac{2\bar{\alpha}_2}{\rho} \left\{ \bar{\alpha}_3 + \frac{\bar{\alpha}_2(1 - \bar{\alpha}_2)}{\rho} \right\} \right] + O(n^{-3/2}), \quad (4.12)$$

donde $\bar{\alpha}_p = E[\cos p(\theta - \mu)]$ y $\bar{\beta}_p = E[\sin p(\theta - \mu)]$ son los elementos coseno y seno del p -ésimo momento alrededor de μ . Cuando $n \rightarrow \infty$,

$$\frac{\bar{b}_2 - E(\bar{b}_2)}{\sqrt{\text{var}(\bar{b}_2)}} \xrightarrow{D} N(0, 1)$$

si $\text{var}(\bar{b}_2) > 0$, con $E(\bar{b}_2)$ y $\text{var}(\bar{b}_2)$ dados por 4.11 y 4.12, respectivamente.

Para una distribución con densidad $f(\theta)$ que es reflexivamente simétrica alrededor de una dirección “central” ϕ , se tiene que $f(\theta - \phi) = f(\phi - \theta)$. Si además $\rho \neq 0$, entonces $\mu = \phi$ y $E(\bar{b}_2) = \bar{\beta}_2 = 0$. Un estimador de la varianza de \bar{b}_2 en este contexto puede ser obtenido usando 4.12 con $\bar{\beta}_2 = 0$, y ρ y $\bar{\alpha}_p$ siendo reemplazados por sus estimadores \bar{R} y

$$\bar{\alpha}_p = \frac{1}{n} \sum_{i=0}^n \cos p(\theta_i - \bar{\theta}),$$

respectivamente. Por lo tanto, una prueba de distribución libre asintótica para H_0 puede ser basada en el estadístico estudentizado

$$z = \frac{\bar{b}_2}{\sqrt{\text{var}(\bar{b}_2)}}, \quad (4.13)$$

donde

$$\text{var}(\bar{b}_2) = \frac{1}{n} \left[\frac{1 - \bar{a}_4}{2} - 2\bar{a}_2 + \frac{2\bar{a}_2}{\bar{R}} \left\{ \bar{a}_3 + \frac{\bar{a}_2(1 - \bar{a}_2)}{\bar{R}} \right\} \right]. \quad (4.14)$$

Los valores absolutos grandes de 4.13 comparados con los cuantiles de la distribución normal estándar conducen al rechazo de la simetría en favor de alguna alternativa que considere sesgo. Resultados de simulación reportados por Pewsey (2002) sugieren que esta prueba basada en la teoría asintótica puede ser usada para tamaños de 50 o más. Para muestras más pequeñas, la versión bootstrap de esta prueba es recomendada en Pewsey *et al.* (2013). Ambas versiones de la prueba asumen que los datos circulares que son analizados están representados en radianes.

4.3. Inferencia para estadísticos descriptivos circulares

4.3.1. Estimación puntual insesgada (bias-corrected)

Los estimadores puntuales para μ , ρ , $\bar{\beta}_2$ y $\bar{\alpha}_2$ son simplemente sus análogos muestrales $\bar{\theta}$, \bar{R} , \bar{b}_2 y \bar{a}_2 . Sin embargo, los resultados para muestras grandes en (4.9) y (4.10), para la distribución asintótica de $\vec{\zeta} = (\bar{\theta}, \bar{R}, \bar{b}_2, \bar{a}_2)^T$, indica que estos cuatro estimados están generalmente sesgados. Sustituyendo los estadísticos poblacionales desconocidos por su correspondientes análogos muestrales, una estimación puntual insesgada para μ es

$$\hat{\mu}_{BC} = \bar{\theta} - \widehat{bias}(\bar{\theta}), \quad \text{donde} \quad \widehat{bias}(\bar{\theta}) = -\frac{\bar{b}_2}{2n\bar{R}^2}. \quad (4.15)$$

Similarmente, los estimados puntuales corregidos para ρ , $\bar{\beta}_2$ y $\bar{\alpha}_2$ están dados por:

$$\hat{\rho}_{BC} = \bar{R} - \widehat{bias}(\bar{R}), \quad \text{donde} \quad \widehat{bias}(\bar{R}) = \frac{1 - \bar{a}_2}{4n\bar{R}}, \quad (4.16)$$

$$\hat{\beta}_{2BC} = \bar{b}_2 - \widehat{bias}(\bar{b}_2) \quad \text{donde} \quad \widehat{bias}(\bar{b}_2) = \frac{1}{n\bar{R}} \left(-\bar{b}_3 - \frac{\bar{b}_2}{\bar{R}} + \frac{2\bar{a}_2\bar{b}_2}{\bar{R}^3} \right) \quad (4.17)$$

$$\hat{\alpha}_{2BC} = \bar{a}_2 - \widehat{bias}(\bar{a}_2) \quad \text{donde} \quad \widehat{bias}(\bar{a}_2) = \frac{1}{n} \left\{ 1 - \frac{\bar{a}_3}{\bar{R}} - \frac{\bar{a}_2(1 - \bar{a}_2) + \bar{b}_2^2}{\bar{R}^2} \right\}. \quad (4.18)$$

Bajo la suposición de una distribución reflexivamente simétrica subyacente, la inferencia para $\bar{\beta}_2$ no será de interés, pues al ser el valor esperado de \bar{b}_2 , esta última medida de sesgo será muy cercana a cero y, por ende, en las ecuaciones (4.15) y (4.18) puede ser considerada como estadísticamente igual a él.

4.3.2. Intervalos de confianza para los estimadores in sesgados

Basándose en la sección 4.3.1, y de nuevo usando los resultados en (4.9) y (4.10), sustituyendo un estadístico poblacional desconocido por su análogo muestral correspondiente, un intervalo del $100(1 - \alpha)\%$ de confianza in sesgado para μ es obtenido mediante

$$\hat{\mu}_{BC} \pm z_{(1-\alpha/2)} \{\widehat{var}(\bar{\theta})\}^{1/2}, \quad \text{donde} \quad \widehat{var}(\bar{\theta}) = \frac{1 - \bar{a}_2}{2n\bar{R}^2}, \quad (4.19)$$

donde $z_{(1-\alpha/2)}$ denota el $(1 - \alpha/2)$ cuantil de la distribución normal estándar.

Similarmente, para ρ , $\bar{\beta}_2$ y $\bar{\alpha}_2$, estos intervalos están dados por:

$$\hat{\rho}_{BC} \pm z_{(1-\alpha/2)} \{\widehat{var}(\bar{R})\}^{1/2}, \quad \text{donde} \quad \widehat{var}(\bar{R}) = \frac{1 - 2\bar{R}^2 + \bar{a}_2}{2n}, \quad (4.20)$$

$$\hat{\beta}_{2BC} \pm z_{(1-\alpha/2)} \{\widehat{var}(\bar{b}_2)\}^{1/2}, \quad \text{donde}$$

$$\widehat{var}(\bar{b}_2) = \frac{1}{n} \left[\frac{1 - \bar{a}_4}{2} - 2\bar{a}_2 - \bar{b}_2^2 + \frac{2\bar{a}_2}{\bar{R}} \left\{ \bar{a}_3 + \frac{\bar{a}_2(1 - \bar{a}_2)}{\bar{R}} \right\} \right], \quad (4.21)$$

y

$$\hat{\alpha}_{2BC} \pm z_{(1-\alpha/2)} \{\widehat{var}(\bar{a}_2)\}^{1/2}, \quad \text{donde}$$

$$\widehat{var}(\bar{a}_2) = \frac{1}{n} \left[\frac{1 - 2\bar{a}_2^2 + \bar{a}_4}{2} + \frac{2\bar{b}_2}{\bar{R}} \left\{ \bar{b}_3 + \frac{\bar{b}_2(1 - \bar{a}_2)}{\bar{R}} \right\} \right]. \quad (4.22)$$

Bajo la suposición de una distribución reflexivamente simétrica subyacente, la inferencia para $\bar{\beta}_2$ no será de interés y la medida \bar{b}_2 en la ecuación (4.22) puede ser considerado como cero.

5

Análisis y modelación de la incidencia de dengue utilizando estadística circular

En este capítulo se estudiará la incidencia de dengue en Yucatán utilizando las herramientas descriptivas y de modelación circulares, que ya se han presentado en los cuatro primeros capítulos.

Para tal fin, se han considerado seis períodos anuales de estudio que se establecen entre los años 2009 y 2015. En un período de estudio en particular se analizan los casos nuevos de dengue (incidencia), registrados semanalmente, que ocurren entre el mes de mayo de un año y el mes de abril del año siguiente. El definir un período de estudio de esta manera permite unificar la información que brindan los meses finales de un año y los iniciales del siguiente, pues el considerar años distintos podría dar lugar a una interpretación errónea del verdadero comportamiento de la incidencia de dengue (considerar un falso comportamiento bimodal en los años pares, por ejemplo), además, se consigue trabajar con datos que pueden ser considerados como unimodales.

Para cada período de estudio, se ha realizado un análisis descriptivo y el ajuste de un modelo circular de probabilidad razonable, y se presentan aquí los resultados obtenidos en cada uno.

5.1. Los datos

La base de datos que presenta la Dirección General de Epidemiología ¹ (DGE) consiste en informes semanales de diferentes *semanas epidemiológicas*. Sin embargo, es el Sistema Nacional de Vigilancia Epidemiológica (SINAVE) el que “capta, registra y analiza los datos de morbilidad, mortalidad, y daños y riesgos en salud — en este caso específicamente para dengue — a través del Sistema Único de Información para la Vigilancia Epidemiológica (SUIVE), apoyado a su vez en: La Notificación Semanal de Casos Nuevos de Enfermedades y los Sistemas Especiales de Vigilancia Epidemiológica”.

La Notificación Semanal de Casos Nuevos de Enfermedades es un sistema de notificación numérico, que integra a las instituciones del Sistema Nacional de Salud, con unidades médicas distribuidas en todo el país. Actualmente, el Sistema Nacional de Salud se alimenta de la concentración de los casos registrados utilizando un formato llamado SUIVE-1-2014, que se genera semanalmente en cada unidad de salud, para ser capturada en la Plataforma de Notificación Semanal de la DGE, la cual opera en el nivel jurisdiccional, estatal y nacional, y permite capturar la información desde la jurisdicción o delegación institucional ².

En septiembre de 2015, eran 20 033 unidades médicas las que se encargaban de generar la información preliminar al corte de determinada semana epidemiológica, con un promedio de cobertura de notificación de 96.4 %.

5.2. Establecimiento de los períodos de estudio

Considere por ejemplo estudiar el comportamiento de la incidencia de dengue en el año 2009 y en el año 2010. Al observar la Figura 5.11, se observa que en ese año hubo un *tramo largo de bajas frecuencias* de casos nuevos de dengue, que abarca de la primera semana de ese año (la primera de enero según el calendario gregoriano) hasta la semana 32 del mismo. Luego, a partir de la semana 33, comienza un tramo que consiste en un grupo de frecuencias altas que comienzan un franco decrecimiento hacia las últimas semanas del 2009. Sin embargo, este comportamiento decreciente no es propio de 2009, sino que continúa hasta la semana 11 del año 2010 (ver Figura 5.12). Después de esto, comienza un *tramo corto de bajas frecuencias* de casos de dengue, que abarca de la semana 12 a la semana 25

¹Panorama epidemiológico de dengue: http://www.epidemiologia.salud.gob.mx/dgae/panodengue/intd_dengue.html

²Subsistema de notificación de Casos Nuevos de Enfermedades 2015: http://www.epidemiologia.salud.gob.mx/dgae/infoepid/inf_morbilidad.html

del 2010; posteriormente, comienza otro tramo, uno de frecuencias grandes que es propio del 2010.

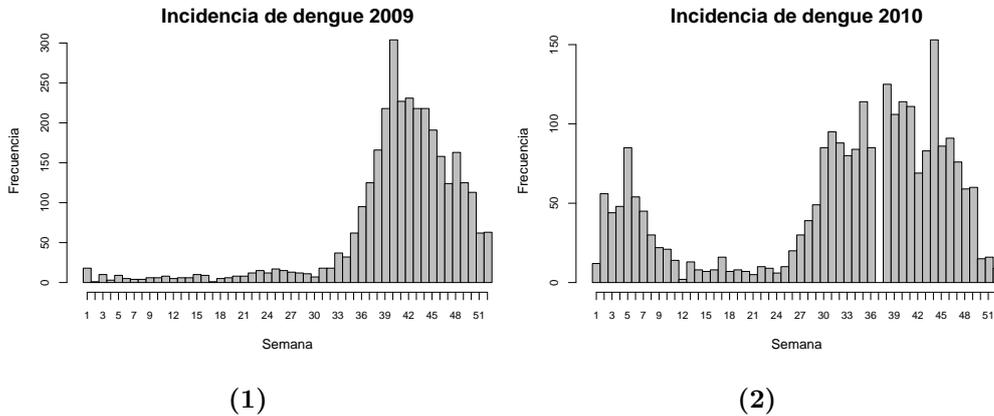


Figura 5.1: Histogramas de frecuencias de incidencia de dengue semanales para los años 2009 y 2010, respectivamente.

Un comportamiento similar al descrito anteriormente, puede ser observado en la Figura 5.2, para los años 2011 y 2012, y en la Figura 5.3 para los años 2013 y 2014.

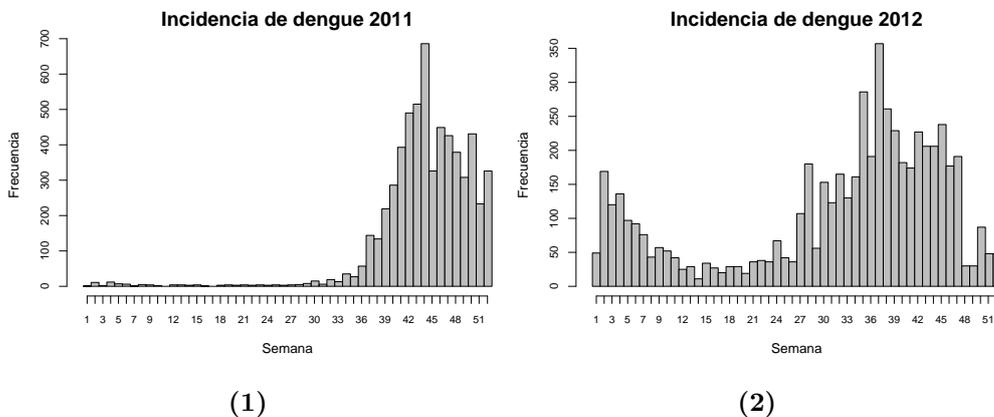


Figura 5.2: Histogramas de frecuencias de incidencia de dengue semanales para los años 2011 y 2012, respectivamente.

Siendo así, si se inicia el estudio del comportamiento de la incidencia de dengue en el año 2009 (2011 ó 2013), éste no se vería reflejado por completo en este año, sino que se requeriría de las primeras semanas del año siguiente para poder observar un comportamiento unimodal. Encasillar el estudio de la incidencia de dengue a un año en particular, por ejemplo 2009 (2011 ó 2013), puede pasar por alto un comportamiento unimodal, o bien, dar la impresión de un comportamiento bimodal para el año 2010 (2012 ó 2014). En el primer caso, se estaría incurriendo

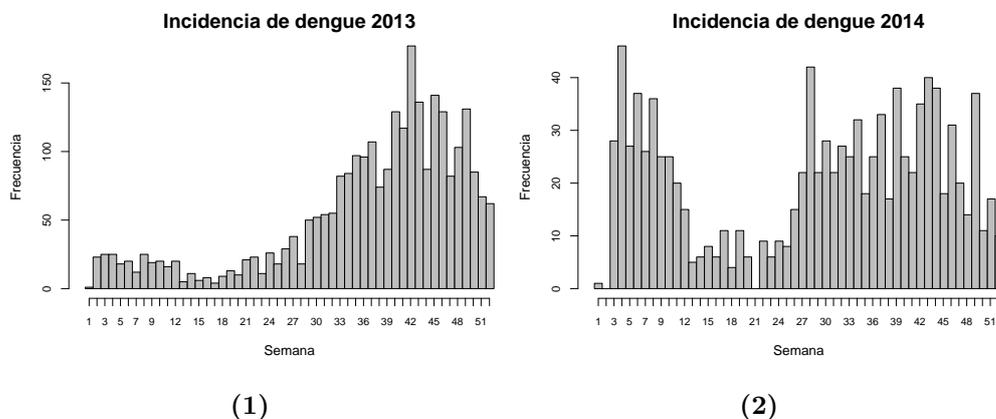


Figura 5.3: Histogramas de frecuencias de incidencia de dengue semanales para los años 2013 y 2014, respectivamente.

en un corte o sesgo en la información, y en el segundo caso, se estaría interpretando de manera errónea la información.

En la búsqueda de un período de estudio que pudiera contener al grupo de frecuencias grandes que comienza en 2009 y termina en 2010, sin incurrir en un sesgo en la información, el siguiente procedimiento fue de utilidad. El *tramo largo de bajas frecuencias* del 2009 (ver Figura 5.11), que abarca de la semana 1 a la 32, fue dividido en dos, y la mitad de ese tramo cayó entre las semanas 16 y 17 de ese año, que corresponden a las últimas semanas del mes de abril. Luego, se dividió en dos el *tramo corto de bajas frecuencias* del 2010 (ver Figura 5.12), que abarca la semana 12 a la 25 de ese año, y la mitad cayó entre las semanas 18 y 19 que corresponden a las primeras semanas del mes de mayo. Claramente, un período que comience durante las dos últimas semanas del mes de abril del 2009, y termine entre las dos primeras semanas del mes de mayo del 2010, contendrá al grupo de frecuencias grandes que comienza en el 2009 y termina en el 2010, sin ningún corte entre ellas (ver Figura 5.41).

Uno podría hacer un procedimiento parecido para encasillar al grupo de altas frecuencias del 2010, en un periodo de entre el *tramo corto de bajas frecuencias* de ese año y el *tramo largo de bajas frecuencias* del año 2011 (ver Figura 5.21). Anteriormente, al dividir el tramo corto del 2010, la mitad cayó entre las dos primeras semanas de mayo. El tramo largo del 2011, que va semana 1 a la 35, al dividirse en dos, la mitad cayó en la semana 17, que es la última semana de abril. Entonces, un período que comience durante las dos primeras semanas de mayo de 2010, y termine durante la última semana del mes de abril de 2011, contendrá al grupo de frecuencias grandes que es propio del año 2010 (ver Figura 5.42).

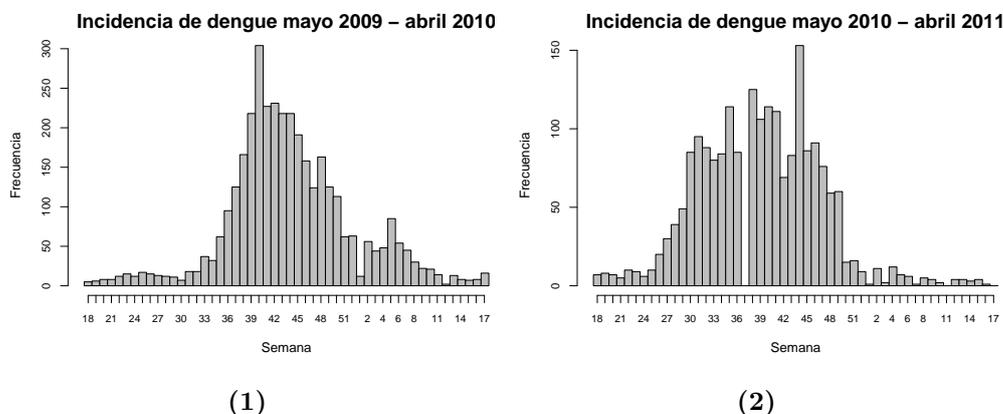


Figura 5.4: Histogramas de frecuencias de incidencia de dengue, considerando un periodo que abarca del mes de mayo de un año, hasta el mes de abril del año siguiente.

En general, se pueden repetir ambos procedimientos para los demás años de estudio. Es decir, encasillar el grupo de frecuencias altas que comienzan en un año impar pero terminan al siguiente, en un período que se establezca entre un tramo largo y otro corto de bajas frecuencias. O bien, encasillar el grupo de frecuencias altas propio de un año par, en un periodo que se establezca entre un tramo corto y otro largo de bajas frecuencias. Si se hace esto, se notará que todos estos períodos tienen inicio en semanas muy cercanas a la semana de inicio del mes de mayo de terminado año. Lo anterior, da la pauta para establecer la primera semana de mayo, del 30 de abril a 6 de mayo (la semana 18 del calendario gregoriano), como el inicio de un período de estudio de incidencia de dengue, en particular, con el fin de partir de un inicio común para los seis períodos de estudio. Ahora bien, con tal de que estos períodos tengan una duración de un año (52 semanas), el final de ellos se ha establecido sea la última semana de abril, del 23 al 29 de abril (la semana 17 del calendario gregoriano), del año siguiente. Además, mayo es un mes de referencia en el calendario meteorológico, pues éste es el mes de inicio de la temporada de lluvias (o de huracanes) a nivel nacional en México. Más generalmente, la temporada de huracanes en el Océano Pacífico Nororiental (Costas de México) abarca del 15 de mayo al 30 de noviembre cada año, mientras que en el Océano Atlántico, que incluye el Golfo de México y el Mar Caribe, se extiende del 1 de junio al 30 de noviembre ³. Y es que, las lluvias, son uno de los principales factores de incidencia de dengue en el estado de Yucatán.

También, al establecer un periodo de estudio del tipo mayo - abril anual, claramente se ha ganado el hecho de que estos datos circulares, o sea, los correspondientes a

³CONAGUA. ¿Qué es un ciclón?: http://smn.cna.gob.mx/index.php?option=com_content&view=article&id=39&Itemid=47.

los casos nuevos de dengue, sean considerados como de comportamiento unimodal como puede verse en las Figuras 5.4 y 5.5.

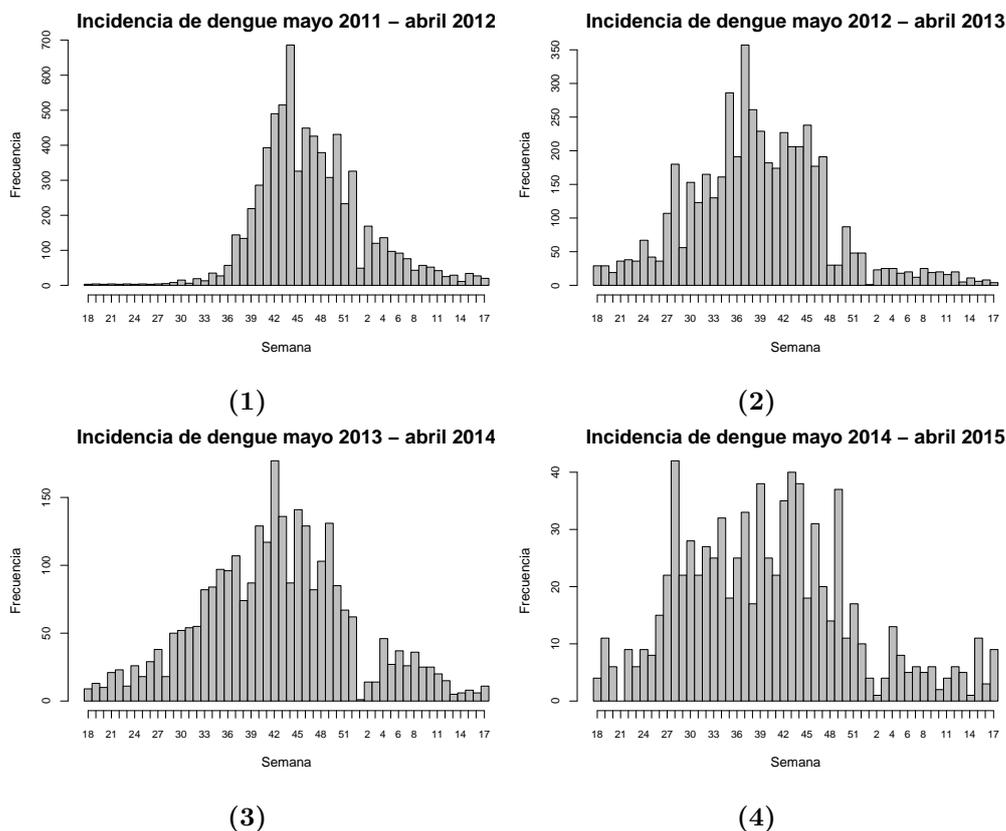


Figura 5.5: Junto con la Figura 5.4, estas gráficas muestran un comportamiento unimodal en los periodos de estudio de la incidencia de dengue, del tipo mayo - abril anual.

5.3. Consideraciones con respecto al tiempo

En este trabajo, no se consideraron los años como bisiestos, en el caso de que existiese uno, aunque de hecho, el único que se presenta es el año 2012. Cada año, de entre 2009 y 2014, se tomó como uno normal. Un *año normal* (no bisiesto), es decir, uno tal cual como lo considera el calendario gregoriano, con mes de inicio enero y mes final diciembre, está conformado por 52 semanas, cada una de siete días, y 1 día más, para hacer un total de 365 días.

En un año normal, la *semana 1* abarca del 1 al 7 de enero, la *semana 2* abarca del 8 al 14 de enero, y así sucesivamente. La *semana 18*, que es la primera de mayo,

en realidad abarca del 30 de abril al 6 de mayo. La *semana 52* abarca del 24 al 30 de diciembre; sin embargo, en este estudio, para fines prácticos, la *semana 52* consistirá de ocho días, o sea del 24 al 31 de diciembre.

La Tabla A.1 del Apéndice A, muestra la ubicación, en el calendario gregoriano, de las semanas contempladas en un período de estudio particular, que barca de la semana 18 a la 52 de un año, hasta la semana 1 a la 17 del siguiente año, es decir, de mayo - diciembre de un año hasta enero - abril del siguiente. Esta Tabla, además, incluye los días de un año normal que conforman una semana numerada en particular.

5.4. Consideraciones con respecto a la base de datos

Una *semana epidemiológica* es una entidad de tiempo que los informes de la DGE manejan para presentar la información de los casos nuevos de dengue que se han registrado y acumulado.

La información contenida en un informe en particular es sobre una semana epidemiológica en específico. En él se presentan los casos de dengue acumulados hasta esa semana epidemiológica. Se presenta el acumulado nacional y el acumulado a nivel de un estado en particular. También se presenta la división entre los casos que se dan por “Fiebre por Dengue” (FD) y los que se dan por “Fiebre Hemorrágica por Dengue” (FHD) ¹.

A nivel nacional, se hacen comparaciones de las cantidades que se analizan con respecto a la semana epidemiológica análoga del año pasado. Se presenta una gráfica en la que se exhibe el avance de los casos confirmados y probables que se dan en el transcurso del año, y también se muestran los casos confirmados y probables del año pasado. Otra gráfica muestra la aparición de los serotipos propios del dengue en un estado en particular. Una tabla resume los casos acumulados de dengue probables, confirmados y las defunciones debidas a esta enfermedad, por cada estado de la república, y se compara con respecto a lo sucedido en el año pasado.

En el informe, sólo se habla de los estados de la república en donde se hayan presentado al menos un caso nuevo de dengue desde el inicio de año. Para un estado en específico, se presentan los casos acumulados por FD, por FHD (o dengue grave), y las defunciones, para esa semana en particular, y estas cantidades se comparan con respecto a la semana análoga del año pasado. También, un gráfico nos muestra

los municipios donde se han presentado los mayores casos de dengue en el estado.

Por lo que se ha observado, una semana epidemiológica particular intenta siempre estar de acorde al lapso de tiempo que abarcaría su semana análoga, con igual número de orden, en el calendario gregoriano. Sin embargo, sucede que siempre ocurren desfases, ya se hacia atrás o hacia adelante, de una semana epidemiológica con respecto a su análoga gregoriana. A veces estos desfases son drásticos. Y más aún, sucede que además de esto, una semana epidemiológica podría no ser estrictamente de 7 días, sino de más o de menos.

En este trabajo, se ha considerado que un desfase de hasta 3 días de una semana epidemiológica con respecto a su análoga ubicada en el calendario gregoriano, diferirá ligeramente de la información y conclusiones a las que se podrían llegar, si se hubiesen registrado los casos nuevos de dengue de acorde al lapso de tiempo estricto que maneja una semana ordenada en el calendario gregoriano.

En el Apéndice C se exhibe una serie de consideraciones, entre desfases y tamaños diferentes de las semanas epidemiológicas, que se han tenido en cuenta con respecto a la base de datos de un año en particular, así como una serie de acciones que se han efectuado a fin de amortiguar los desfases drásticos que se dan entre una semana epidemiológica y su análoga gregoriana. De manera global, con estas acciones, se ha conseguido que si se presentara un desfase de una semana epidemiológica respecto a su análoga gregoriana éste sea de a los más tres días.

La Tabla A.2 del apéndice A, presenta las frecuencias registradas en la semanas epidemiológicas que caen en un período de estudio abril-mayo anual, según la DGE y contemplando las consideraciones hechas en el Apéndice C.

5.5. La incidencia de dengue como datos circulares

En cada período de estudio mayo - abril anual, se ha hecho un análisis de la incidencia de dengue en el estado de Yucatán.

Los casos nuevos de dengue pueden ser analizados como datos circulares, debido a que son un fenómeno de ocurrencia periódica en todos los años. Entonces, se pueden utilizar las herramientas de la estadística circular para su análisis.

Para tal fin, cada período de estudio mayo - abril fue representado por una circunferencia de un círculo unitario. En un período de estudio en particular, la circunferencia correspondiente fue dividida en 365 puntos que representan los 365

días de lapso de un año. A cada día en la circunferencia se le ha asignado un ángulo representante medido en radianes, de la forma matemática, es decir, con dirección de inicio 0° hacia el oriente, y sentido de rotación positivo antihorario. Siendo así, el día 30 de abril tiene como representante el ángulo 0° , el día primero de mayo el ángulo $2\pi/365$, el día dos de mayo el ángulo $2 \times (2\pi/365)$, el día tres de mayo el ángulo $3 \times (2\pi/365)$, y así sucesivamente, hasta terminar con los días del primer año de estudio y empezar con los del siguiente. El día 29 de abril tiene como representante el ángulo $364 \times (2\pi/365)$.

Los datos que nos brinda la Dirección General de Epidemiología ¹, consisten de los casos de dengue acumulados hasta cierta semana epidemiológica (incidencia semanal) de un año normal en particular. La diferencia entre los casos de dengue acumulados en una semana epidemiológica y los casos acumulados en la semana epidemiológica anterior, nos da el número de casos registrados en la primera. Siendo así, los datos que se tienen a disposición son los del número de casos nuevos de dengue en una semana específica, es decir, la incidencia semanal. Por ende, en esta circunferencia, también se han incluido 52 intervalos, cada uno representa una semana en particular, por tanto, cada uno abarca siete días, con excepción de la última semana del primer año de estudio la cuál incluirá ocho días.

El número de casos de dengue se obtiene de manera semanal de la base de datos, sin embargo, no se sabe en qué día específico de la semana han ocurrido. Por ende, se ha convenido ubicar los casos de dengue en un día representante de la semana, y se ha elegido el día mitad de ella, a saber, el cuarto día de la misma. Por ejemplo, la semana 18, la primera de mayo del primer año de estudio, abarca del 30 de abril, con ángulo representante 0° , hasta el 6 de mayo, con ángulo representante $6 \times (2\pi/365)$; por ende, el día representante de esta semana es el cuarto día, 3 de mayo, con ángulo $3 \times (2\pi/365)$. Y así, se aplica este procedimiento en todas las semanas del período mayo - abril con duración de un año, excepto para la semana 52 del primer año de estudio; esta semana abarca del 24 de diciembre, con ángulo $357 \times (2\pi/365)$, hasta el 31 de diciembre, con ángulo $364 \times (2\pi/365)$, pero aquí, esta semana abarca 8 días, por lo que como ángulo representante se ha elegido el promedio de los ángulos del cuarto y quinto día que conforman esta semana, a saber, $[360 \times (2\pi/365) + 361 \times (2\pi/365)] / 2$.

5.6. Uso del diagrama circular en un período de estudio mayo - abril anual

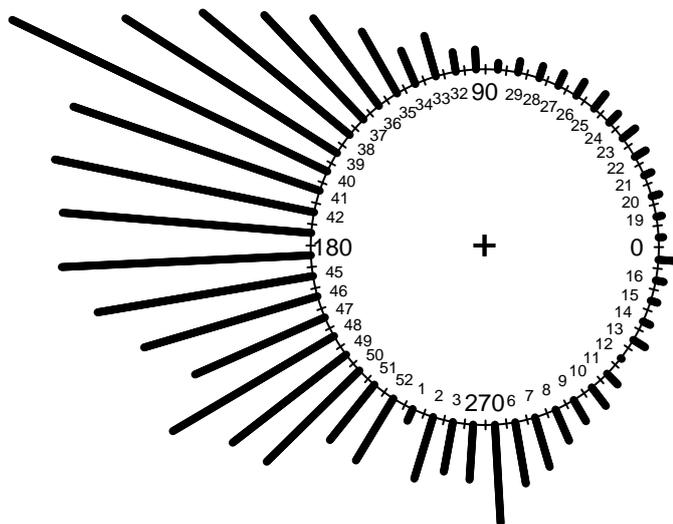


Figura 5.6: Diagrama circular para el periodo mayo 2009 - abril 2010.

La Figura 5.6 muestra el diagrama circular del período de estudio mayo 2009 - abril 2010 de casos reportados de dengue en el estado de Yucatán. Este diagrama servirá de ejemplo para explicar cómo se le dará lectura a la información general contenida en él, cuando se estudie un período en particular.

En este diagrama, los datos circulares son medidos, de la manera matemática, utilizando el oriente como dirección de inicio y sentido de rotación positivo antihorario. La dirección 0° representa el 30 de abril del primer año de estudio.

Como puede observarse en la Figura 5.6, el círculo se ha dividido en 52 intervalos, que en realidad son pequeños arcos en el círculo, y esta división es hecha mediante líneas separadoras pequeñas. La división empieza en la dirección cero.

Para los ángulos, la representación usada en dicho diagrama circular es en grados. Los números más grandes, localizados en las direcciones que apuntan hacia los cuatro puntos cardinales, o sea, 0, 90, 180 y 270, son los referentes angulares.

Los números más pequeños hacen referencia al orden de aparición de una semana

en particular si se le ubica en el calendario gregoriano. Los números 17, 18, 30, 31, 43, 44, 4 y 5, no se pueden ver debido a limitaciones visuales del software. Se ha de tener en cuenta que los números del 18 al 52 corresponden al primer año de estudio y, los números del 1 al 17 corresponden al año siguiente.

Los “rayos de sol” que se pueden ver son en realidad las frecuencias semanales de incidencia de dengue. Cada frecuencia está constituida por un grupo de puntos acomodados en la misma dirección, y están muy cercanos entre sí, debido al gran número de ellos, a tal grado de parecerse en su conjunto a una línea continua.

5.7. Modelos circulares elegidos y herramientas utilizadas

Una vez hechas las observaciones y consideraciones anteriores, se iniciará el análisis de los casos de dengue reportados, de manera semanal en el estado de Yucatán, según la Dirección General de Epidemiología (DGE). Un análisis en particular se ha hecho para cada período de estudio mayo - abril anual, los cuales se han establecido entre los años 2009 y 2015. En cada uno de estos períodos, se ha de tener en cuenta la siguiente simbología.

Estadístico descriptivo circular	Símbolo
Dirección modal muestral	$\check{\theta}$
Dirección mediana muestral	$\tilde{\theta}$
Dirección media muestral	$\bar{\theta}$
Longitud media resultante muestral	\bar{R}
Medida de Sesgo	\bar{b}_2
Sesgo propuesto por Mardia (1972)	\hat{s}
Medida de Curtosis	\bar{a}_2

La contraparte para la curtosis \hat{k} propuesta por Mardia (1972) no ha sido considerada en este estudio debido a que su interpretación es acerca de un modelo circular que no ha sido considerado en este trabajo (ver Sección 2.4).

En cada período se han ajustado cinco distribuciones circulares:

1. Inversa Bastchelet Completa, con base von Mises, con parámetros ξ , κ , μ y λ , que controlan la localización, la concentración, el sesgo y la picudez, respectivamente.
2. Inversa Batschelet Sogada, con parámetro de picudez $\lambda = 0$ (von Mises Sogada).

3. Inversa Bastchelet Simétrica, con parámetro de sesgo $\nu = 0$.
4. Jones-Pewsey, con parámetros μ , κ y ψ , que controlan la localización, la concentración y la forma, respectivamente.
5. Von Mises, con dirección media μ y parámetro de concentración κ .

Los primeros dos modelos consideran una parámetro de sesgo ν . Los últimos tres consideran simetría reflexiva.

Un idea fundamental en modelación estadística es el de *parsimonia*: si dos modelos ofrecen ajustes igual de buenos a los datos, entonces el menos complejo entre los dos (el de menos parámetros) es preferido. Como consecuencia, generalmente interesa comparar los ajustes de modelos que compitan entre sí, con miras a identificar el modelo con menos parámetros que ofrezca un ajuste razonable a los datos. Varios criterios están disponibles para comparar modelos, y los más conocidos involucran el valor máximo de la función de log-verosimilitud (Pewsey *et al.*, 2013).

Existen dos criterios que sirven para comparar modelos que no son necesariamente anidados (que uno sea submodelo de otro). El primero de ellos es el criterio de información de Akaike (1974)

$$AIC = 2k - 2l_{\max}, \quad (5.1)$$

donde k es el número de parámetros en un modelo, y l_{\max} denota el valor máximo de la función log-verosimilitud. El segundo es el criterio de información bayesiano

$$BIC = k \log(n) - 2l_{\max} \quad (5.2)$$

donde n denota el tamaño de la muestra. Este criterio es debido a Schwarz *et al.* (1978). En ambos criterios, el “mejor” modelo será el que tenga el valor más bajo del índice correspondiente.

En este trabajo, el mejor modelo para un período de estudio en particular se ha elegido de acuerdo a los criterios de información AIC y BIC.

La Tabla A.3 del apéndice A, contiene el resumen de todos los ajustes hechos en cada período de estudio mayo - abril anual. Para un período en particular, se exhiben los estimados de los parámetros que rigen cierta distribución (calculados vía máxima verosimilitud), el máximo de la función log-verosimilitud, y los valores AIB y BIC, todos éstos obtenidos para cada uno de los cinco modelos considerados para el ajuste.

A continuación, se exhibirán los resultados obtenidos para cada período de estudio mayo - abril anual.

5.8. Período 1: Mayo 2009 - Abril 2010

Análisis descriptivo

La Figura 5.7 y la Tabla 5.1 resumen los gráficos y estadísticos descriptivos circulares para el período mayo 2009 - abril 2010. En este período se registraron en total 3594 casos de dengue en el estado. La semana con mayor cantidad de casos fue la semana 40 (1ª de octubre) del 2009.

El diagrama circular en la Figura 5.71 y la Tabla 5.2 muestran hacia qué semanas apuntan las direcciones moda $\check{\theta}$, mediana $\tilde{\theta}$ y media muestral $\bar{\theta}$. Ya se ha dicho que en la semana 40 se alcanza la mayor incidencia de dengue, y por ende, la dirección moda $\check{\theta} = 154.8493^\circ$ apunta hacia ahí. La dirección $\tilde{\theta} = 175.5616^\circ$ muestra que es a finales de octubre, en la semana 43, donde se alcanza la mitad de los casos reportados en total para este período. La dirección $\bar{\theta} = 179.913^\circ$ muestra que las grandes frecuencias de incidencia se presentan alrededor de la semana 44, que es la semana de transición entre octubre y noviembre.

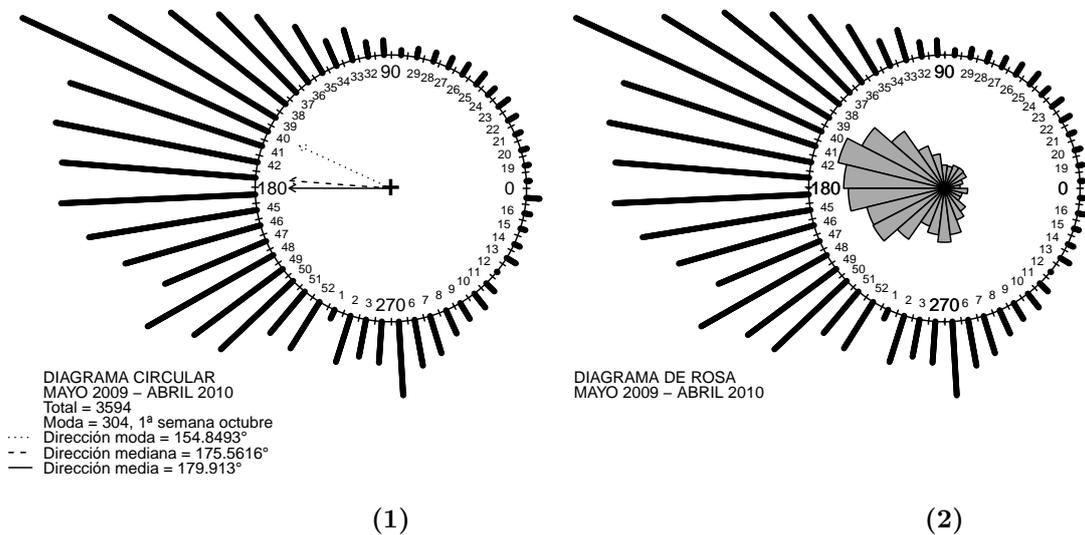


Figura 5.7: Diagramas circular a la izquierda y de rosa a la derecha para el período mayo 2009 - abril 2010.

Período 1: mayo 2009 - abril 2010	
Total	3594
Moda	304
$\check{\theta}$	154.8493°
$\tilde{\theta}$	175.5616°
$\bar{\theta}$	179.9130°
\bar{R}	0.6511
\bar{b}_2	-0.1153
\hat{s}	-0.5597
\bar{a}_2	0.3233

Tabla 5.1: Total, moda y medidas de localización, de dispersión y de sesgo para el período mayo 2009 - abril 2010.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	40	Octubre (1 ^a)
$\tilde{\theta}$	43	Octubre (4 ^a)
$\bar{\theta}$	44	Octubre - Noviembre

Tabla 5.2: Tiempo al que se dirigen los estadísticos descriptivos circulares de localización, para el período mayo 2009 - abril 2010.

La medida $\bar{R} = 0.6511$ ofrece una señal de una concentración de los casos de dengue al alrededor de la dirección media $\bar{\theta}$. Si dividimos el año 2009 en cuatro partes se puede ver, por el diagrama de rosa 5.72, que esta concentración ocurre después de la mitad del tercer cuarto de año hasta antes del final del último cuarto del 2009. Más precisamente, se presenta de la semana 36 (1^a septiembre) hasta la semana 51 (3^a de diciembre) del 2009.

La circunferencia unitaria para este período se ha cortado aproximadamente a la mitad del arco que abarca de la semana 30 del 2009 hasta la semana 12 del 2010, recorriendo la circunferencia en sentido horario, que es el arco en el que se presenta un tramo de bajas frecuencias de casos de dengue. Más precisamente, se ha cortado en el punto con dirección 27° (0.4712 radianes). Con base en este corte, se construye el histograma lineal de la Figura 5.8.

Las medidas $\bar{b}_2 = -0.1153$ y $\hat{s} = -0.5597$ revelan un sesgo en sentido antihorario con respecto de la dirección media, y el diagrama de rosa de la Figura 5.72 confirma esta situación. Este diagrama, junto con el histograma lineal de la Figura 5.8, muestran que a partir de la semana 33 (3^a de agosto) de 2009 empieza un aumento

rápido de los casos de dengue, que se prolonga hasta la semana 42 (3^a de octubre), y luego de aquí, comienza un decrecimiento lento de los casos que continúa hacia la semana 11 (2^a de marzo) del año 2010.

La medida $\bar{a}_2 = 0.3233$ es señal de una picudez no extrema en la forma polar de la distribución de los datos, pues su valor no es cercano a 1 (ver diagrama circular de la Figura 5.71).

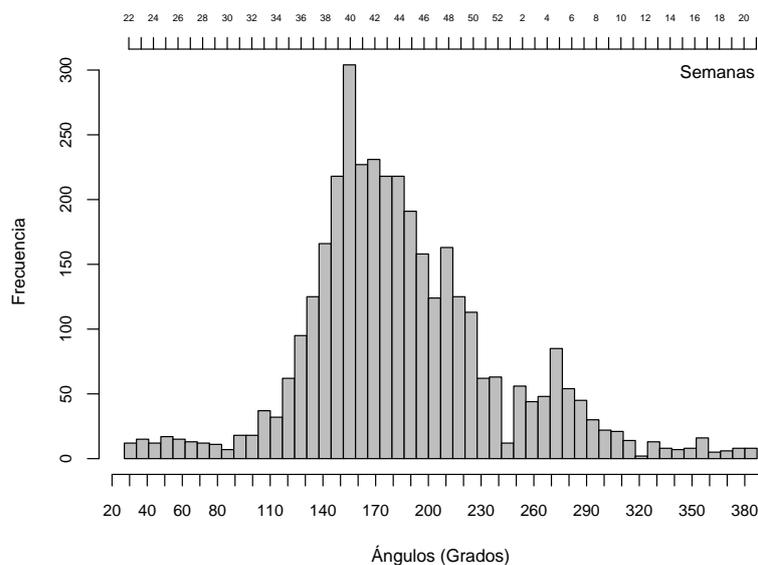


Figura 5.8: Histograma lineal de frecuencias de dengue para el período mayo 2009 - abril 2010, con punto de corte en la dirección 27° (0.4712 radianes). A los valores arriba de 360° del final, se le debe restar esta cantidad para obtener valores en el rango de $(0^\circ, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba calculado es de 0.6511, y arroja un p-valor muy cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de prueba es de 11.9226 y el p-valor es de $9.02586e-33$. Con p-valores muy cercanos a cero, las hipótesis nulas de uniformidad y simetría reflexiva, respectivamente, de estas pruebas se rechazan para cualquier nivel de significancia α frecuentemente utilizado (digamos 0.01, 0.05 o 0.1). Por lo anterior, se espera que un modelo que considere sesgo ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3 muestra que el mejor modelo que se ajusta a los casos de dengue para este período es el de una distribución de la familia Inversa Batschelet, con base

von Mises, en el que se consideran todos los parámetros distintos de cero (modelo completo).

Período 1: mayo 2009 - abril 2010	
Mejor ajuste: Inversa Batschelet Completo	
Parámetro	Estimado
ξ	3.1416
κ	1.6332
ν	0.0683
λ	0.3231

Tabla 5.3: Ajuste de un modelo Inversa Batschelet Completo, con base von Mises, para el período mayo 2009 - abril 2010.

La Tabla 5.3 muestra los estimados de los parámetros que controlan la distribución Inversa Batchelet obtenidos vía máxima verosimilitud. El estimado $\hat{\nu} = 0.0683$, que considera un sesgo en sentido antihorario, aunque cercano a cero, ofrece un mejor ajuste ante un modelo Inverso Batschelet Simétrico ($\nu = 0$) (ver Tabla A.3), es decir, ante uno en donde se se consideré igual a cero. El estimado $\hat{\lambda} = 0.3233$ considera una distribución con picudez no extrema en su presentación polar.

La dirección modal teórica de la distribución, estimada por el modelo a través de la expresión $\xi - 2\nu \pmod{2\pi}$, es de 3.0049 radianes, la cual apunta hacia las semanas 42 y 43 del 2009, que son la tercera y cuarta semana de octubre, respectivamente.

La Figura 5.9 muestra las representaciones polar y lineal del modelo ajustado para este período. La Figura 5.91 muestra el diagrama circular que incluye la dirección modal teórica y la dirección media muestral; se observa que la semana en la que se hubiese esperado se presente la mayor frecuencia de dengue sería la primera o la segunda anterior a la 44, semana alrededor de la cual se concentran las mayores frecuencias de casos de dengue. La Figura 5.92 muestra que hay una mayor masa de probabilidad después de la dirección modal teórica, en sentido antihorario.

Con el fin de validar el modelo, la Figura 5.10 exhibe los gráficos probabilidad-probabilidad (P-P plot, Figura 5.101) y cuantil-cuantil (Q-Q plot, Figura 5.102) de la distribución ajustada para este período. En una gráfica P-P el valor de la función de distribución empírica para cada representante semanal es graficado contra el valor correspondiente de la función de distribución estimada para el modelo propuesto. En su lugar, en una gráfica Q-Q se grafican los valores de la función cuantil empírica y los valores de la función cuantil estimada. El ajuste de una distribución circular es considerado como bueno si los puntos graficados están

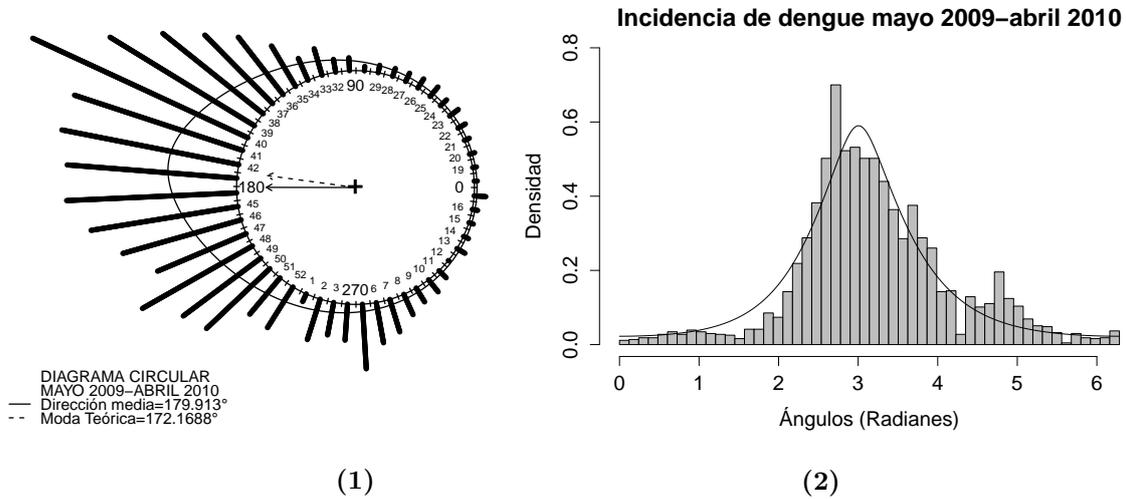


Figura 5.9: Representaciones polar, a la izquierda, y lineal, a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2009 - abril 2010. El histograma lineal comienza en la dirección cero.

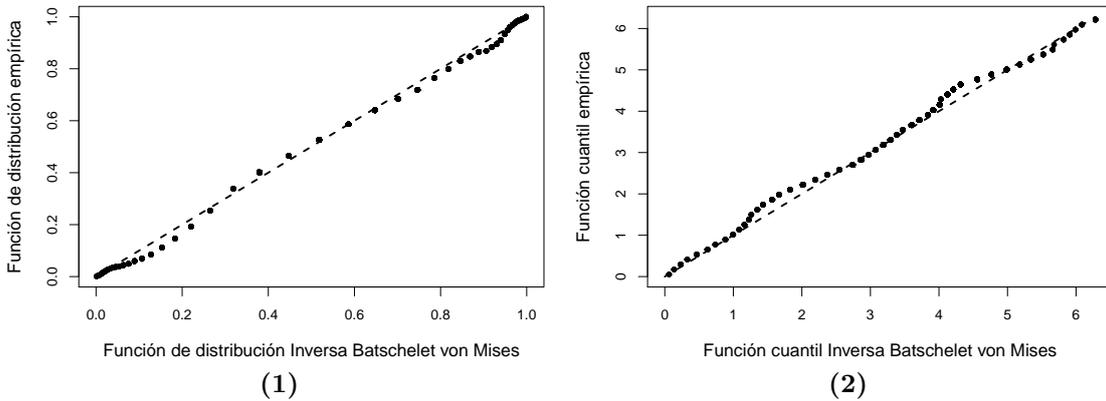


Figura 5.10: Gráficos probabilidad-probabilidad (P-P plot), a la izquierda, y cuantil-cuantil (Q-Q plot), a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2009 - abril 2010.

cerca de la línea de referencia diagonal que conecta a los puntos (0,0) y (1,1) en un gráfico P-P, y los puntos (0, 0) y (2π, 2π) en un gráfico Q-Q.

De la Figura 5.10, en general, se observa que en las gráficas P-P y Q-Q la mayoría de los puntos de ambas están cercanos a la línea de referencia, lo cual es una señal de un ajuste razonable.

En la prueba de uniformidad de Rayleigh, convertida en una de bondad de ajuste cuando se aplica a los datos transformados $2\pi F(\theta_i)$, el valor del estadístico de

prueba calculado es de 0.0123, y arroja un p-valor de 0.5825, el cual es relativamente grande; entonces, a cualquier nivel de significancia α frecuentemente usado, esto indica que no hay evidencia suficiente para rechazar la hipótesis nula de uniformidad, y su vez, esto indica que para la distribución de los casos de dengue en este período, es razonable el ajuste de una Inversa Batschelet Completa con parámetros estimados vía máxima verosimilitud, dados en la Tabla 5.3.

5.9. Período 2: Mayo 2010 - Abril 2011

Análisis descriptivo

La Figura 5.111 y la Tabla 5.4 resumen los gráficos y estadísticos descriptivos circulares para el período mayo 2010 - abril 2011.

El diagrama circular de la Figura 5.111, muestra que se han reportado cero casos en la semana 37 del 2010. Lo anterior es un contraposition a los registros de incidencia que se tienen semanas antes y después de ésta, pues precisamente se encuentra en un tramo de frecuencias semanales altas. Con fin de evaluar qué tanto influye la posible ausencia de datos de incidencia de esta semana en el cálculo de los estadísticos descriptivos circulares se han hecho dos análisis, uno con “registro corregido” y el otro considerando los registros de incidencia tal como lo reporta la DGE. La corrección consiste en considerar en la semana 37 el promedio de las frecuencias que se dan entre las tres semanas anteriores y las tres posteriores a la 37. Al comparar uno y otro análisis, no se encuentran cambios significativos entre los resultados que arrojan los cálculos de los estadísticos descriptivos circulares considerados en el estudio. Con el fin de estar de acuerdo con los números semanales y el total reportados de manera oficial por la DGE, se ha optado trabajar con la base de datos de este período sin corrección.

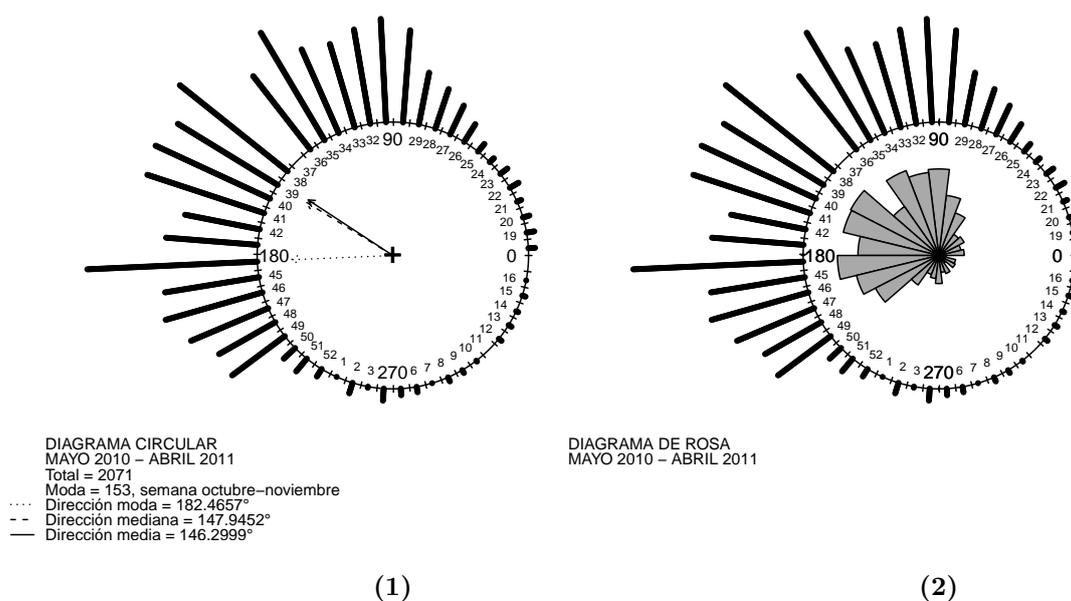


Figura 5.11: Diagramas circular a la izquierda, y de rosa a la derecha para el período mayo 2010 - abril 2011.

En este período, se registraron en total 2071 casos de dengue en el estado. La semana con mayor cantidad de casos fue la 44 (transición entre octubre y noviembre) del 2010.

El diagrama circular en la Figura 5.111 y la Tabla 5.5 muestran hacia qué semanas apuntan las direcciones $\check{\theta}$, $\tilde{\theta}$ y $\bar{\theta}$. Para la semana modal 44 la dirección moda $\check{\theta} = 182.4658^\circ$. La dirección $\tilde{\theta} = 147.9452^\circ$ y la dirección $\bar{\theta} = 146.2999^\circ$ están muy cercanas y de hecho apuntan hacia la misma semana, la 39 (4^a de septiembre), es decir, alrededor de esta semana se acumulan las grandes frecuencias y, además, en ella se alcanza la mitad de los casos reportados en total para este período.

Período 2: mayo 2010 - abril 2011	
Total	2071
Moda	153
$\check{\theta}$	182.4658°
$\tilde{\theta}$	147.9452°
$\bar{\theta}$	146.2999°
\bar{R}	0.6560833
\bar{b}_2	0.03323178
\hat{s}	0.1647683
\bar{a}_2	0.1808922

Tabla 5.4: Total, moda y medidas de localización, de dispersión y de sesgo para período mayo 2010 - abril 2011.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	44	Octubre - Noviembre
$\tilde{\theta}$	39	Septiembre (4 ^a)
$\bar{\theta}$	39	Septiembre (4 ^a)

Tabla 5.5: Tiempo al que se dirigen los estadísticos descriptivos circulares de localización, para el período mayo 2010 - abril 2011.

La medida $\bar{R} = 0.6561$ ofrece una señal de una concentración de los casos alrededor de θ . Si dividimos el 2010 en cuatro partes se puede ver, por el diagrama de rosa 5.112, que esta concentración ocurre antes de la mitad del tercer cuarto hasta después de la mitad del último cuarto del 2010. Más precisamente, se presenta de la semana 30 (4^a de julio) hasta la semana 47 (3^a de noviembre) del 2010.

La circunferencia unitaria para este período se ha cortado aproximadamente a la mitad del arco, que abarca de la semana 1 hasta la semana 17 del 2011 (recorriendo

la circunferencia en sentido antihorario), el cual presenta las más bajas frecuencias. Es decir, se ha cortado en el punto con dirección 301° (5.2534 radianes). Con base en este corte, se construye el histograma lineal de la Figura 5.12.

La medida $\bar{b}_2 = 0.0332$ es cercana a cero, pero su contraparte estandarizada $\hat{s} = 0.1648$ no lo es tanto. Por ende, no hay evidencia muestral contundente de un comportamiento reflexivamente simétrico. Sin embargo, estas medidas evidencian un ligero sesgo en sentido horario respecto de la dirección media (ver diagrama de rosa de la Figura 5.112).

El diagrama de rosa, junto con el histograma lineal de la Figura 5.12, muestran que a partir de la semana 26 (transición junio - julio) del 2010 empieza un aumento rápido de los casos, y el tramo de altas frecuencias se mantiene hasta las semana 49 (1^a de diciembre), después de la cual se registra un descenso abrupto de los casos reportados en el mismo año.

La medida $\bar{a}_2 = 0.1809$ es señal de una picudez no extrema en la forma polar de la distribución de los datos (ver diagrama circular 5.112), pero es menor que la reportada un período antes.

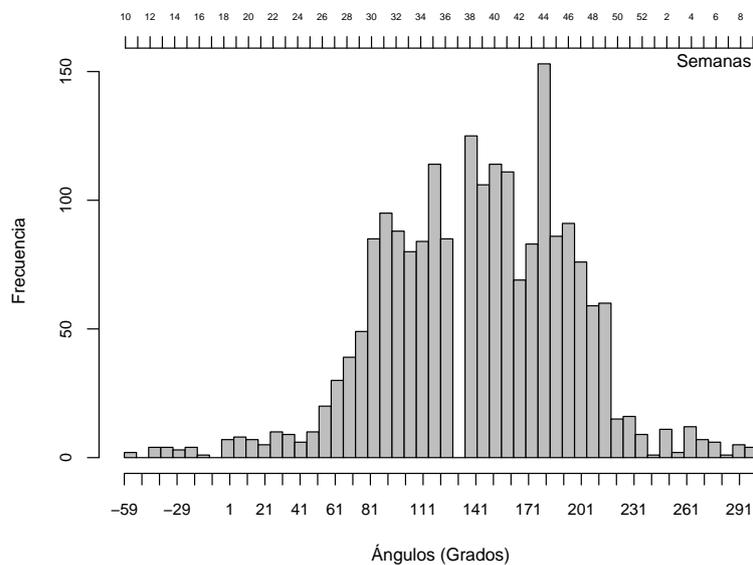


Figura 5.12: Histograma lineal de frecuencias de dengue, para el período mayo 2010 - abril 2011, con punto de corte en la dirección 301° (5.2534 radianes). A los valores negativos del principio hay que sumarle 360° para obtener valores en el rango de $(0, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba es de 0.6561 y arroja un p-valor muy cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de

prueba es de 2.8654 y el p-valor es de 0.0042. Con p-valores muy cercanos a cero, las hipótesis nulas de uniformidad y de simetría reflexiva se rechazan para cualquier nivel de significancia α frecuentemente utilizado. Por lo anterior, se espera que un modelo que considere sesgo ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3 muestra que el mejor modelo que se ajusta a los casos de dengue para este período es el de una distribución Inversa Batschelet Completa.

Período 2: mayo 2010 - abril 2011	
Mejor ajuste: Inversa Batschelet Completo	
Parámetro	Estimado
ξ	2.29223
κ	2.088345
ν	-0.1938533
λ	-0.2857464

Tabla 5.6: Ajuste de un modelo Inversa Batschelet Completo, con base von Mises, para el período mayo 2010 - abril 2011.

La Tabla 5.6 muestra los estimados de los parámetros que controlan la distribución Inversa Batchelet, obtenidos vía máxima verosimilitud. El valor de $\hat{\nu} = -0.1939$ considera un sesgo en sentido horario. El valor $\hat{\lambda} = -0.2857$ considera una distribución con cima plana en su presentación polar.

La dirección modal teórica de la distribución, estimada por el modelo a través de la expresión $\xi - 2\nu(\text{ mod } 2\pi)$, es de 2.6799 radianes, la cual apunta hacia la semana 40 del 2010, que es la primera semana de octubre.

La Figura 5.131 muestra el diagrama circular que incluye la representación polar de la distribución, la dirección modal teórica y la dirección media muestral; se observa que la semana (la 40) en la que se hubiese esperado se presente la mayor frecuencia de dengue, sucede después de la semana (la 39) alrededor de la cual se concentran los mayores casos de dengue de este período. La Figura 5.132 muestra representación lineal de la distribución, en la cual se observa hay una mayor masa de probabilidad antes de la dirección modal teórica, en sentido horario.

De la Figura 5.14, en general, se observa que las gráficas P-P y Q-Q presentan la mayoría de sus puntos muy cercanos a la línea de referencia, lo cual es señal de un ajuste razonable.

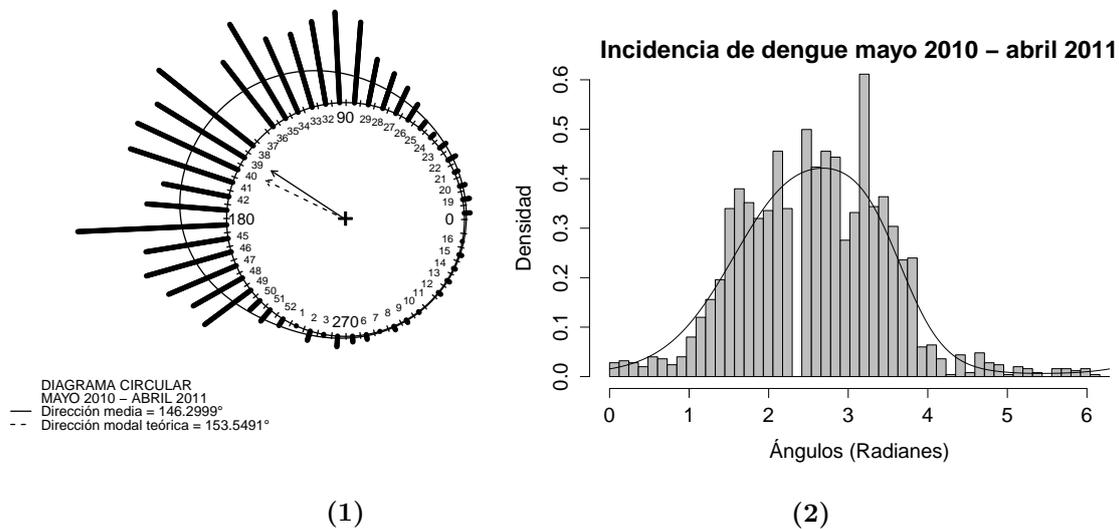


Figura 5.13: Representaciones polar, a la izquierda, y lineal, a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2010 - abril 2011. El histograma lineal comienza en la dirección cero.

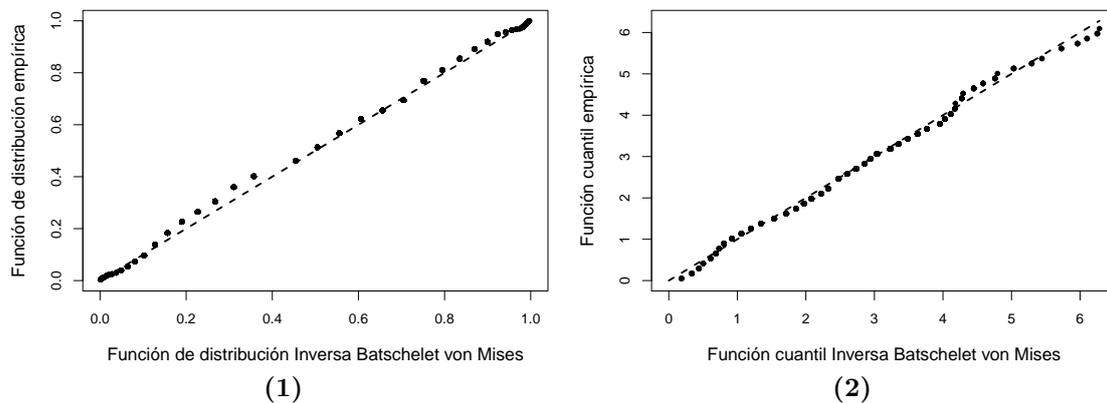


Figura 5.14: Gráficos probabilidad-probabilidad (PP plot), a la izquierda, y cuantil-cuantil (QQ plot), a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2010 - abril 2011.

En la prueba de bondad de ajuste, basado en la de uniformidad de Rayleigh, el valor del estadístico de prueba es de 0.0419 y el p-valor arrojado es de 0.0262. Se concluye que para el comportamiento de los casos de dengue de este período es razonable el ajuste de una distribución Inversa Batschelet Completa, a un nivel de significancia $\alpha < 0.02$.

Interesante es decir que si se hubiese utilizado la base de datos corregida, los resultados de la pasada prueba de bondad de ajuste arrojan un p-valor de aproximadamente 0.2308, con lo cual se puede trabajar a cualquier nivel de significancia

α frecuentemente utilizado. Al parecer, la prueba de bondad de ajuste basado en la de Rayleigh es sensible a la posible ausencia de registro de casos de dengue en la semana 37, que sucede justamente en un tramo de frecuencias semanales altas.

5.10. Período 3: Mayo 2011 - Abril 2012

Análisis descriptivo

La Figura 5.15 y la Tabla 5.7 resumen los gráficos y estadísticos descriptivos circulares para el período mayo 2011 - abril 2012. En este período se registraron en total 7044 casos de dengue en el estado. La semana con mayor cantidad de casos fue la semana 44 (transición entre octubre y noviembre) del 2011.

El diagrama circular en la Figura 5.15(1) y la Tabla 5.8 muestran hacia qué semanas apuntan las direcciones $\tilde{\theta}$, $\tilde{\theta}$ y $\bar{\theta}$. Para la semana modal 44 la dirección moda es $\tilde{\theta} = 182.4657^\circ$. La dirección $\tilde{\theta} = 196.274^\circ$ y la dirección $\bar{\theta} = 197.7495^\circ$ están muy cercanas y de hecho apuntan hacia la misma semana, la 46 (2ª de noviembre), es decir, alrededor de esta semana se acumulan las grandes frecuencias y, además, en ella se alcanza la mitad de los casos reportados en total para este período.

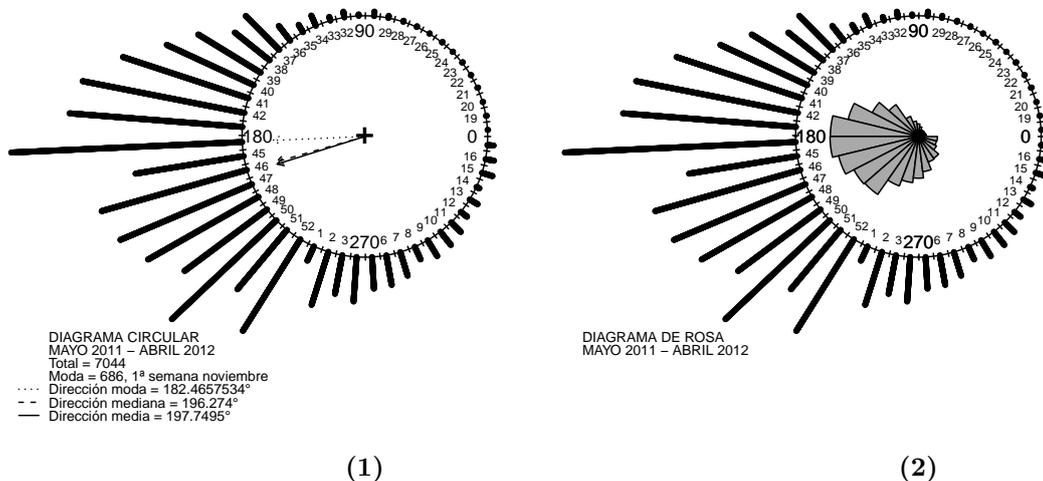


Figura 5.15: Diagramas circular a la izquierda y de rosa a la derecha para el período mayo 2011 - abril 2012.

Período 3: mayo 2011 - abril 2012	
Total	7044
Moda	686
$\check{\theta}$	182.4657534°
$\tilde{\theta}$	196.274°
$\bar{\theta}$	197.7495°
\bar{R}	0.7310412
\bar{b}_2	-0.1158706
\hat{s}	-0.8307007
\bar{a}_2	0.3714658

Tabla 5.7: Total, moda y medidas de localización, de dispersión y de sesgo para período mayo 2011 - abril 2012.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	44	Octubre - Noviembre
$\tilde{\theta}$	46	Noviembre (2 ^a)
$\bar{\theta}$	46	Noviembre (2 ^a)

Tabla 5.8: Tiempo al que se dirigen los estadísticos descriptivos circulares de localización, para el período mayo 2011 - abril 2012.

La medida $\bar{R} = 0.731$ ofrece una señal de concentración de los casos de dengue alrededor de $\bar{\theta}$. Si dividimos el año 2011 en cuatro partes, con ayuda del diagrama de rosa de la Figura 5.152, notaremos que esta concentración ocurre prácticamente en el último cuarto del 2011, de la semana 40 (1^a de octubre) a la 51 (3^a de diciembre).

La circunferencia unitaria para este período se ha cortado aproximadamente a la mitad del arco, que abarca de la semana 18 a la 33 (recorriendo la circunferencia en sentido antihorario), que es el arco en el que se ha observado un tramo de bajas frecuencias de casos de dengue. Más precisamente, se ha cortado en el punto con dirección 55° (0.9599311 radianes). Con base en este corte, se construye el histograma lineal de la Figura 5.16.

Las medidas $\bar{b}_2 = -0.1159$ y $\hat{s} = -0.8307$ revelan un sesgo en sentido antihorario, respecto de la dirección media, y el diagrama de rosa de la Figura 5.152 confirma esta situación. Este diagrama, junto con el histograma lineal de la Figura 5.16, muestran que a partir de la semana 37 (2^a de septiembre) del 2011 empieza un aumento rápido de los casos de dengue, que se prolonga hasta la semana 44

(transición entre octubre y noviembre), y luego de aquí, comienza un decrecimiento lento de los casos que continúa hacia la semana 11 (2^a de marzo) del 2012.

La medida de curtosis $\bar{a}_2 = 0.3715$ es señal de una picudez no extrema en la forma polar de la distribución de los datos (ver diagrama circular 5.151).

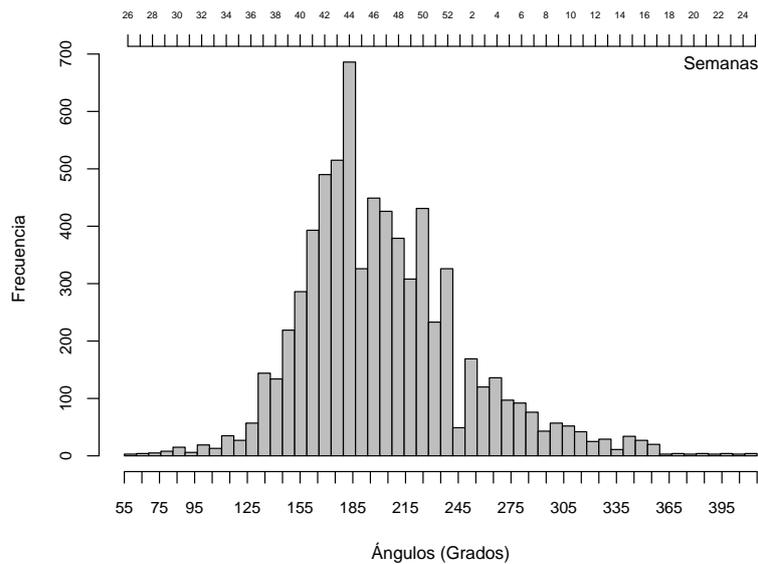


Figura 5.16: Histograma lineal de frecuencias de dengue, para el período mayo 2011 - abril 2012, con punto de corte en la dirección 55° (0.9599311 radianes). A los valores arriba de 360° del final, se les debe restar esta cantidad para obtener valores en el rango de $(0^\circ, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba es de 0.731 y arroja un p-valor muy cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de prueba es de 20.5477 y el p-valor es de $8.069242e-94$. Con p-valores muy cercanos a cero, las hipótesis nulas de uniformidad y simetría reflexiva se rechazan a cualquier nivel de significancia α frecuentemente utilizado. Por lo anterior, se espera que un modelo que considere sesgo ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3 muestra que el mejor modelo que se ajusta a los casos de dengue para este período, es el de una distribución Inversa Batschelet Completa.

Período 3: mayo 2011 - abril 2012	
Mejor ajuste: Inversa Batschelet Completo	
Parámetro	Estimado
ξ	4.217897
κ	2.218277
ν	0.5074749
λ	0.1031728

Tabla 5.9: Ajuste de un modelo Inverso Batschelet Completo, con base von Mises, para el período mayo 2011 - abril 2012.

La Tabla 5.9 muestra los estimados de los parámetros que controlan la distribución Inversa Batchelet obtenidos vía máxima verosimilitud. El estimado $\hat{\nu} = 0.5075$ considera un sesgo medio en sentido antihorario. El estimado $\hat{\lambda} = 0.1032$ considera una distribución con picudez no extrema en su presentación polar.

La dirección modal teórica de la distribución, estimada por el modelo a través de la expresión $\xi - 2\nu \pmod{2\pi}$, es de 3.2029 radianes, la cual apunta hacia la semana 44 del 2011, la cual es la semana transición de octubre a noviembre.

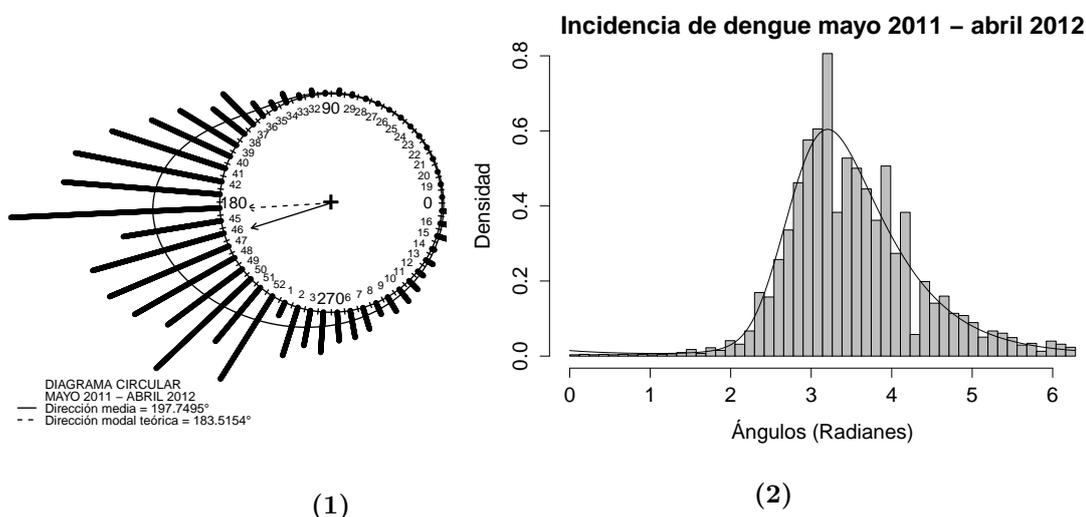


Figura 5.17: Representaciones polar, a la izquierda, y lineal, a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2011 - abril 2012.

La Figura 5.171 muestra el diagrama circular que incluye la forma polar de la distribución, la dirección modal teórica y la dirección media muestral; se observa que la semana (la 44) en la que se hubiese esperado se presente la mayor frecuencia

de dengue, sucede antes de la semana (la 46) alrededor de la cual se concentran los mayores casos de dengue de este período. La Figura 5.172 muestra la representación lineal de la distribución, en la cual se observa hay una mayor masa de probabilidad después de la dirección modal teórica, en sentido antihorario.

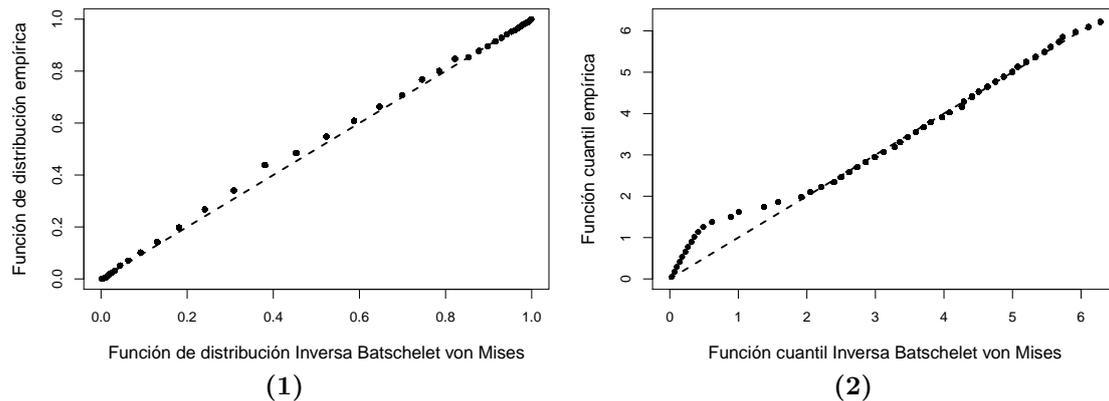


Figura 5.18: Gráficos probabilidad-probabilidad (PP plot), a la izquierda, y cuantil-cuantil (QQ plot), a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2011 - abril 2012.

De la Figura 5.18, en general, se observa que las gráficas P-P y Q-Q presentan la mayoría de sus puntos muy cercanos a la línea de referencia, lo cual es señal de un ajuste razonable.

En la prueba de bondad de ajuste, basado en la de uniformidad de Rayleigh, el valor del estadístico de prueba es de 0.0116 y el p-valor arrojado es de 0.3871. Se concluye que para el comportamiento de los casos de dengue de este período es razonable el ajuste de una distribución Inversa Batschelet Completa, a cualquier nivel de significancia α frecuentemente utilizado.

5.11. Período 4: Mayo 2012 - Abril 2013

Análisis descriptivo

La Figura 5.191 y la Tabla 5.10 resumen los gráficos y estadísticos descriptivos circulares para el período mayo 2012 - abril 2013. En este período se registraron en total 4833 casos de dengue en el estado. La semana con mayor cantidad de casos fue la 37 (2ª de septiembre) del 2012.

El diagrama circular en la Figura 5.191 y la Tabla 5.11 muestran hacia qué semanas apuntan las direcciones $\check{\theta}$, $\tilde{\theta}$ y $\bar{\theta}$. Para la semana modal 37 la dirección moda es $\check{\theta} = 134.137^\circ$. La dirección $\tilde{\theta} = 141.0411^\circ$ y la dirección $\bar{\theta} = 141.8827^\circ$ están muy cercanas, y de hecho apuntan hacia la misma semana, la 38 (3ª de septiembre), es decir, alrededor de esta semana se acumulan las grandes frecuencias y, además, en ella se alcanza la mitad de los casos reportados en total para este período.

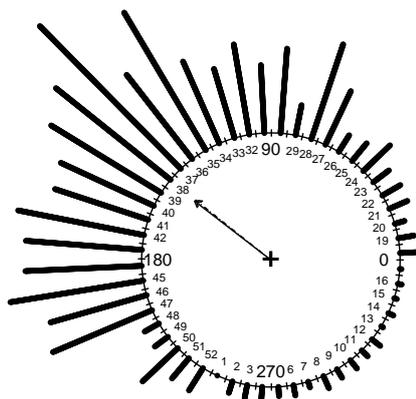


DIAGRAMA CIRCULAR
 MAYO 2012 - ABRIL 2013
 Total = 4833
 Moda = 357, 2ª semana septiembre
 Dirección moda = 134.1369°
 --- Dirección mediana = 141.8827°
 — Dirección media = 141.0411°

(1)

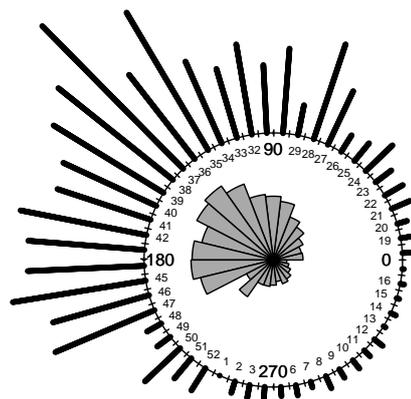


DIAGRAMA DE ROSA
 MAYO 2012 - ABRIL 2013

(2)

Figura 5.19: Diagramas circular a la izquierda y de rosa a la derecha para el período mayo 2012 - abril 2013.

Período 4: mayo 2012 - abril 2013	
Total	4833
Moda	357
$\check{\theta}$	134.137°
$\tilde{\theta}$	141.0411°
$\bar{\theta}$	141.8827°
\bar{R}	0.5999591
\bar{b}_2	0.03493162
\hat{s}	0.1380582
\bar{a}_2	0.2044234

Tabla 5.10: Total, moda y medidas de localización, de dispersión y de sesgo para período mayo 2012 - abril 2013.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	37	Septiembre (2 ^a)
$\tilde{\theta}$	38	Septiembre (3 ^a)
$\bar{\theta}$	38	Septiembre (3 ^a)

Tabla 5.11: Tiempo al que se dirigen los estadísticos descriptivos circulares de localización, para el período mayo 2012 - abril 2013.

La medida $\bar{R} = 0.5999$ ofrece una señal de una concentración de los casos de dengue alrededor de $\bar{\theta}$. Si dividimos el año 2012 en cuatro partes, con ayuda del diagrama de rosa de la Figura 5.192, observaremos que esta concentración es notoria a principios del tercer cuarto hasta después de la mitad del último cuarto del 2012; más precisamente, a partir de la semana 28 (2^a de julio) hasta la 47 (3^a de noviembre).

La circunferencia unitaria para este período se ha cortado aproximadamente a la mitad del arco, que abarca de la semana 1 hasta la semana 17 del 2013 (recorriendo la circunferencia en sentido antihorario), el cual presenta las más bajas frecuencias. Es decir, se ha cortado en el punto con dirección 301° (5.2534 radianes). Con base en este corte, se construye el histograma lineal de la Figura 5.20.

La medida $\bar{b}_2 = 0.0349$ es cercana a cero, pero su contraparte estandarizada $\hat{s} = 0.1380$ no lo es tanto. Por ende, no hay evidencia muestral contundente de un comportamiento reflexivamente simétrico. Sin embargo, estas medidas evidencian un ligero sesgo en sentido horario lejos de la dirección media (ver diagrama de rosa de la Figura 5.192).

El diagrama de rosa, junto con el histograma lineal de la Figura 5.20, muestran que a partir de la semana 27 (1^a de julio) del 2012 empieza un franco aumento de los casos, y el tramo de altas frecuencias se mantiene hasta las semana 47 (3^a de noviembre), después de la cual se registra un descenso abrupto de los casos reportados en el mismo año.

La medida $\bar{a}_2 = 0.2044$ señala una picudez no extrema en la forma polar de la distribución de los datos (ver diagrama circular 5.191), pero es menor que la reportada un período antes.

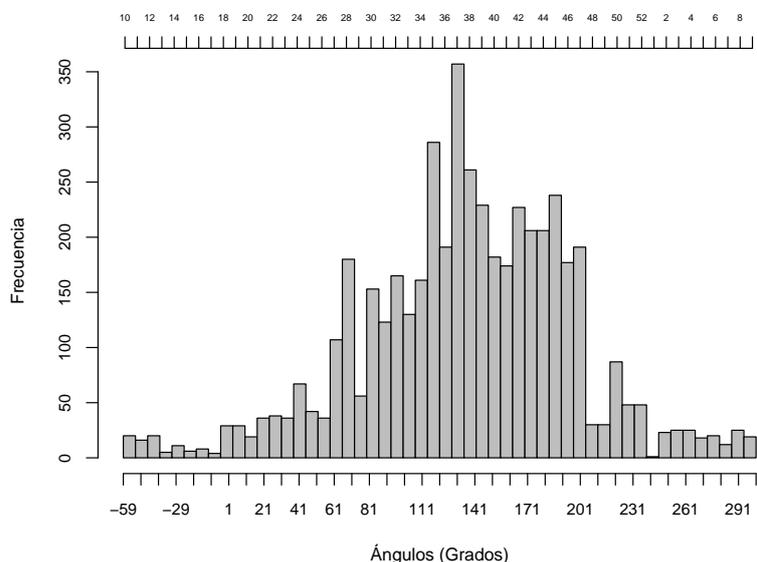


Figura 5.20: Histograma lineal de frecuencias de dengue, para el período mayo 2012 - abril 2013, con punto de corte en la dirección 301° (5.2534 radianes). A los valores negativos del principio hay que sumarles 360° para obtener valores en el rango de $(0, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba es de 0.6 y arroja un p-valor muy cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de prueba es de 4.4471 y el p-valor es de $8.70234e-06$. Con p-valores muy cercanos a cero, las hipótesis nulas de uniformidad y simetría reflexiva se rechazan a cualquier nivel de significancia α frecuentemente utilizado. Por lo anterior, se espera que un modelo que considere sesgo ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3 muestra que el mejor modelo que se ajusta a los casos de dengue para este período, es el de una distribución Inversa Batschelet considerando su parámetro de curtosis $\lambda = 0$ y todos los demás distintos de cero. En la literatura circular, a esta distribución se le suele llamar “von Mises sesgada”.

Período 4: mayo 2012 - abril 2013	
Mejor ajuste: von Mises Sesgado	
Parámetro	Estimado
ξ	2.201661
κ	1.522607
ν	-0.2036836
λ	0

Tabla 5.12: Ajuste de un modelo Inverso Batschelet con $\lambda = 0$, denominado “von Mises sesgado”, para el período mayo 2012 - abril 2013.

La Tabla 5.12 muestra los estimados de los parámetros que controlan la distribución Inversa Batschelet obtenidos vía máxima verosimilitud. El valor de $\hat{\nu} = -0.2036$ considera un sesgo en sentido horario. El valor $\hat{\lambda} = 0$ considera una distribución de cima plana.

La dirección modal teórica de la distribución, estimada por el modelo a través de la expresión $\xi - 2\nu \pmod{2\pi}$, es de 2.6090, la cual apunta hacia la semana 39 del 2012, la cuarta semana de octubre.

La Figura 5.211 muestra el diagrama circular que incluye la forma polar de la distribución, la dirección modal teórica y la dirección media muestral; se observa que la semana (la 39) en la que se hubiese esperado se presente la mayor frecuencia de casos de dengue, sucede después de la semana (la 38) alrededor de la cual se concentran las altas frecuencias de incidencia. La Figura 5.212 muestra la representación lineal de la distribución, en la cual se observa hay una mayor masa de probabilidad antes de la dirección modal teórica, en sentido horario.

De la Figura 5.22, en general, se observa que las gráficas P-P y Q-Q presentan la mayoría de sus puntos muy cercanos a la línea de referencia, lo cual es señal de un ajuste razonable.

En la prueba de bondad de ajuste, basado en la de Rayleigh, el valor del estadístico de prueba es de 0.0135 y el p-valor arrojado es de 0.4141. Se concluye que

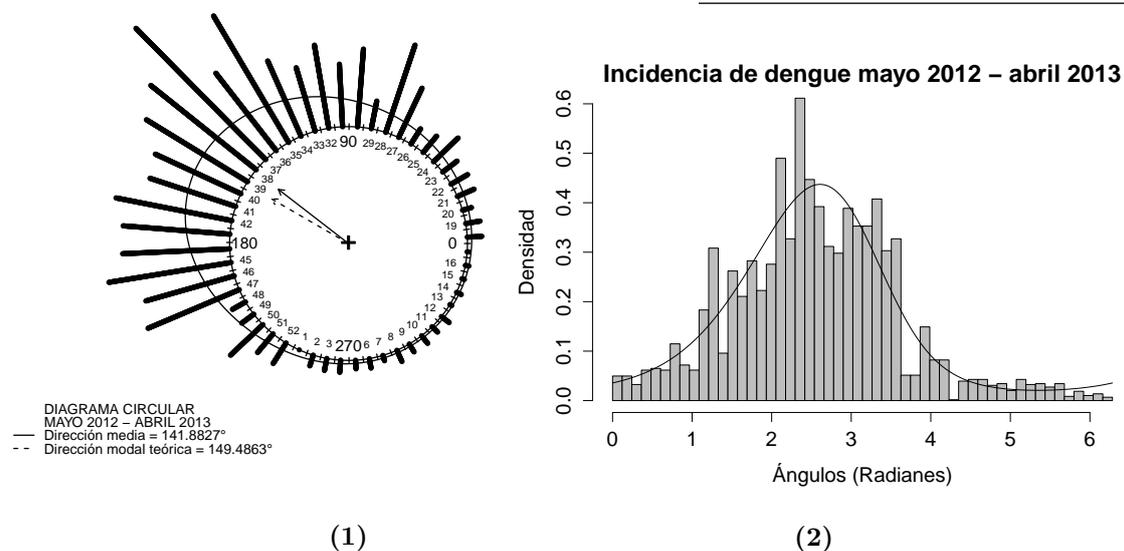


Figura 5.21: Representaciones polar, a la izquierda, y lineal, a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2010 - abril 2011. El histograma lineal comienza en la dirección cero.

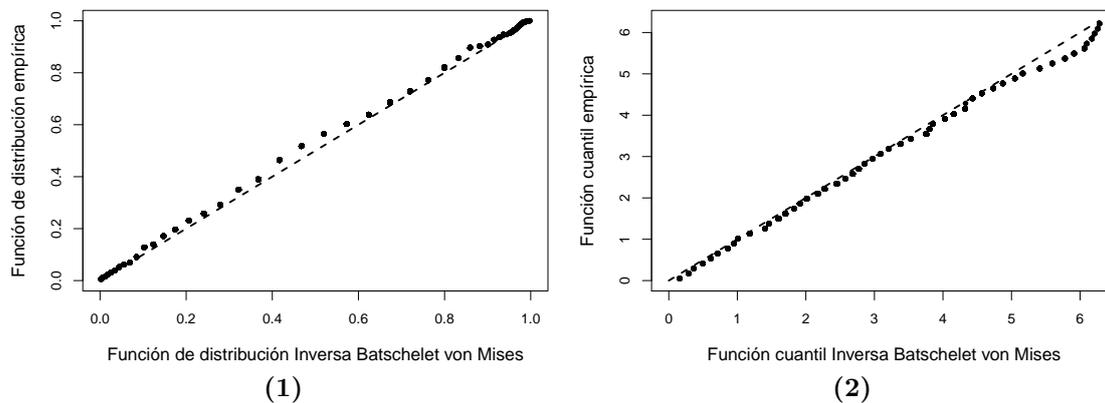


Figura 5.22: Gráficos probabilidad-probabilidad (PP plot), a la izquierda, y cuantil-cuantil (QQ plot), a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2012 - abril 2013.

para el comportamiento de los casos de dengue para este período es razonable el ajuste de una distribución von Mises sesgada, a cualquier nivel de significancia α frecuentemente usado.

5.12. Período 5: Mayo 2013 - Abril 2014

Análisis descriptivo

La Figura 5.23 y la Tabla 5.13 resumen los gráficos y estadísticos descriptivos circulares para el período mayo 2013 - abril 2014. En este período se registraron en total 2822 casos de dengue en el estado. La semana con mayor cantidad de casos fue la semana 42 (3^a de octubre) del 2013.

El diagrama circular en la Figura 5.23(1) y la Tabla 5.14 muestran hacia qué semanas apuntan las direcciones $\check{\theta}$, $\tilde{\theta}$ y $\bar{\theta}$. En este caso $\check{\theta} = \tilde{\theta} = 168.6575^\circ$ y $\bar{\theta} = 167.4262^\circ$ apuntan todas hacia la misma semana, la 42 (3^a de octubre), lo cual significa que esta semana es en la que ocurre la mayor frecuencia de casos de dengue; también, alrededor de ella se acumulan las mayores frecuencias, y además, en ella se alcanza la mitad de los casos reportados en total para este período.

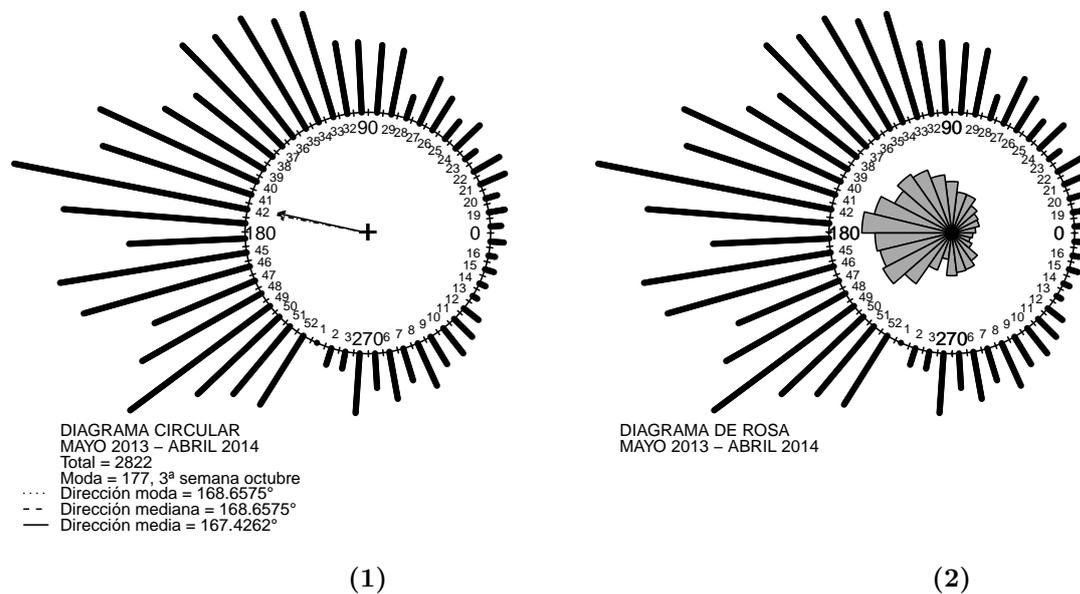


Figura 5.23: Diagramas circular a la izquierda y de rosa a la derecha para el período mayo 2011 - abril 2012.

Período 5: mayo 2013 - abril 2014	
Total	2822
Moda	177
$\check{\theta}$	168.6575°
$\tilde{\theta}$	168.6575°
$\bar{\theta}$	167.4262°
\bar{R}	0.5105622
\bar{b}_2	0.007603004
\hat{s}	0.0222044
\bar{a}_2	0.151142

Tabla 5.13: Total, moda y medidas de localización, de dispersión y de sesgo para período mayo 2013 - abril 2014.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	42	Octubre (3 ^a)
$\tilde{\theta}$	42	Octubre (3 ^a)
$\bar{\theta}$	42	Octubre (3 ^a)

Tabla 5.14: Tiempo al que se dirigen los estadísticos descriptivos circulares de localización, para el período mayo 2013 - abril 2014.

La medida $\bar{R} = 0.511$ indica que los datos exhiben un rasgo más de dispersión que de concentración alrededor $\bar{\theta}$. Los datos están dispersos en un rango amplio alrededor de ella. Si se observa el diagrama de rosa de la Figura 5.232, este rango va desde tiempo antes de la mitad del tercer cuarto del 2013 hasta casi al final del último cuarto; más precisamente, abarca a partir de la semana 30 (4^a de julio) hasta la semana 51 (3^a de diciembre) del 2013.

La circunferencia unitaria se ha cortado aproximadamente a la mitad del arco que abarca de la semana 13 a la 20 (recorriendo la circunferencia en sentido antihorario), que es el arco en el que se presenta un tramo de bajas frecuencias de casos de dengue. Más precisamente, se ha cortado en el punto con dirección 353° (6.161012 radianes). Con base en este corte se construye el histograma lineal de la Figura 5.24.

Las medidas $\bar{b}_2 = 0.0076$ y $\hat{s} = 0.0222$ ambas son cercanas a cero, lo cual es un indicador de que sería razonable considerar en primera instancia un modelo simétrico para la distribución de los datos.

La medida $\bar{a}_2 = 0.1511$ indica que la distribución de los datos tiene una picudez no extrema (ver diagrama circular 5.231), y menor a lo reportada en los períodos anteriores.

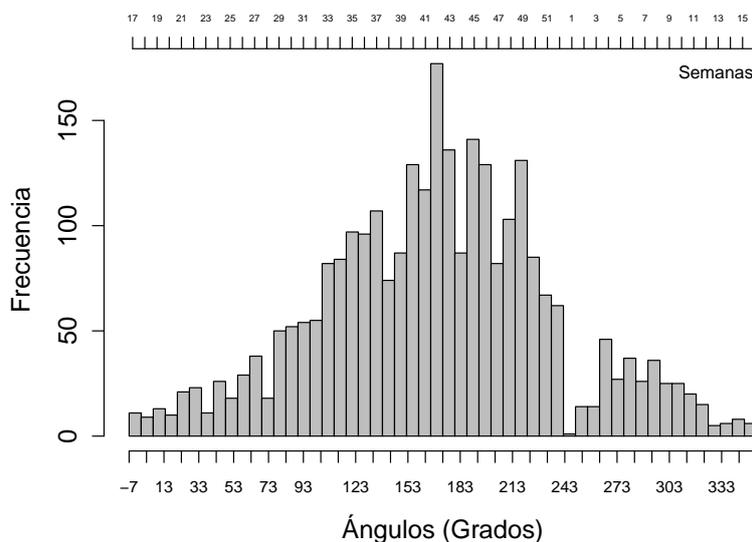


Figura 5.24: Histograma lineal de frecuencias de dengue, para el período mayo 2013 - abril 2014, con punto de corte en la dirección 353° (6.161012 radianes). A los valores negativos del principio hay que sumarles 360° para obtener valores en el rango de $(0, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba es de 0.5106 y arroja un p-valor muy cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de prueba es de 0.6385 y el p-valor es de 0.5232 . De lo anterior, la hipótesis nula de uniformidad es rechazada a cualquier nivel de significancia α frecuentemente utilizado, pues el p-valor es cercano a cero; en contraste, no hay evidencia suficiente para rechazar la hipótesis nula de simetría reflexiva alrededor de la dirección media, a cualquier nivel de significancia α frecuentemente utilizado. Por esto último, y por los valores cercanos a cero de las medidas de sesgo, se espera que un modelo que considere simetría reflexiva ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3 muestra que el mejor modelo que se ajusta a los casos de dengue para este período es el de una distribución von Mises, también llamada Normal Circular, la cual considera simetría reflexiva. Esta Tabla también muestra bajo qué

parámetros esta distribución puede ser considerada como un submodelo Inverso Batschelet ($\xi = \mu, \nu = \lambda = 0$), o un submodelo Jones-Pewsey ($\psi = 0$).

Período 5: mayo 2013 - abril 2014	
Mejor ajuste: von Mises	
Parámetro	Estimado
μ	2.922138
κ	1.19285

Tabla 5.15: Ajuste de un modelo von Mises para el período mayo 2013 - abril 2014.

La Tabla 5.15 muestra los estimados de los parámetros que controlan la distribución von Mises obtenidos vía máxima verosimilitud. En este caso $\hat{\kappa} = A_1^{-1}(\bar{R})$ y $\hat{\mu} = \bar{\theta}$.

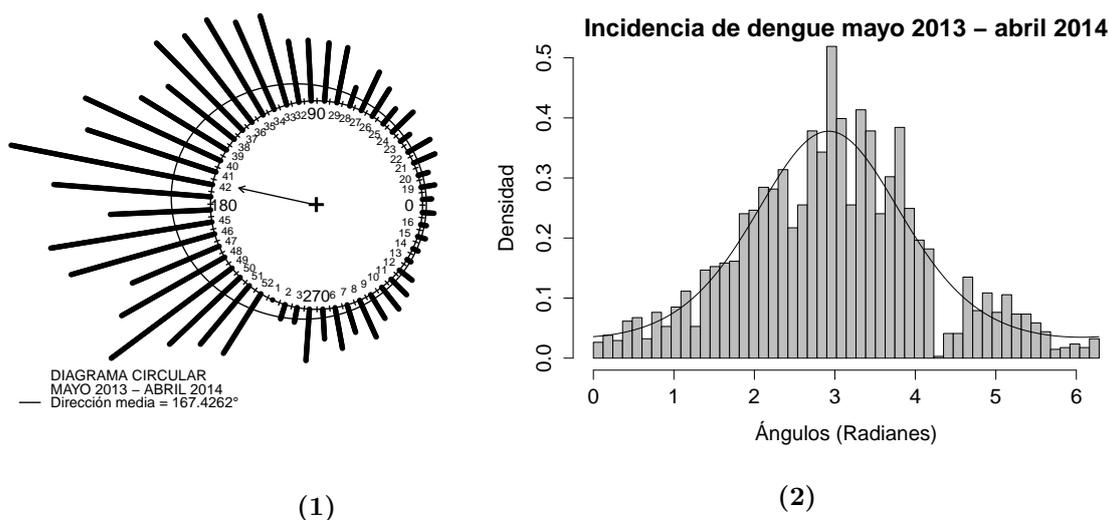


Figura 5.25: Representaciones polar, a la izquierda, y lineal, a la derecha, de la distribución von Mises ajustada para el período mayo 2013 - abril 2014.

La Figura 5.251 muestra el diagrama circular que incluye la forma polar de la distribución y la dirección media. Bajo un modelo von Mises las direcciones modal, mediana y media coinciden, de hecho, visto como un submodelo Inverso Batchelet con $\nu = \lambda = 0$, la dirección modal teórica es $\xi - 2\nu = \xi = \mu$. La Figura 5.252 muestra la representación lineal de la distribución, en la cual se observa hay un equilibrio de las masas de probabilidad, antes y después de la dirección media.

La Figura 5.26 exhibe que las gráficas P-P y Q-Q presentan la mayoría de sus puntos cerca de la línea de referencia, los cual es un indicio de un ajuste razonable.

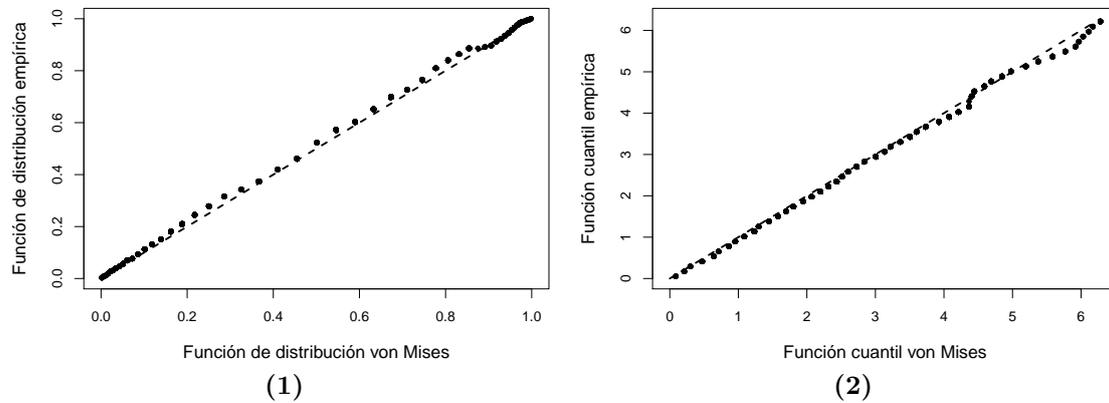


Figura 5.26: Gráficos probabilidad-probabilidad (PP plot), a la izquierda, y cuantil-cuantil (QQ plot), a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período mayo 2013 - abril 2014.

En la prueba de bondad de ajuste, basado en la de uniformidad de Rayleigh, el valor del estadístico de prueba es de 0.0172 y el p-valor arrojado es de 0.4325. Se concluye que para el comportamiento de los casos de dengue para este período el ajuste de una distribución von Mises es razonable a cualquier nivel de significancia α frecuentemente utilizado.

5.13. Período 6: Mayo 2014 - Abril 2015

Análisis descriptivo

La Figura 5.271 y la Tabla 5.16 resumen los gráficos y estadísticos descriptivos circulares para el período mayo 2014 - abril 2015. En este período se registraron en total 830 casos de dengue en el estado. La semana con mayor cantidad de casos fue la 28 (2ª de julio) de 2014.

El diagrama circular en la Figura 5.271 y la Tabla 5.14 muestran hacia qué semanas apuntan las direcciones $\check{\theta}$, $\tilde{\theta}$ y $\bar{\theta}$. Para la semana modal 28 la dirección moda es $\check{\theta} = 72^\circ$. La dirección $\tilde{\theta} = 147.9452^\circ$, muestra que es a finales de septiembre, en la semana 39, donde se alcanza la mitad de los casos reportados en total para este período. La dirección $\bar{\theta} = 142.0618^\circ$, muestra que las grandes frecuencias de incidencia se dan alrededor de la semana 38 (3ª de septiembre).

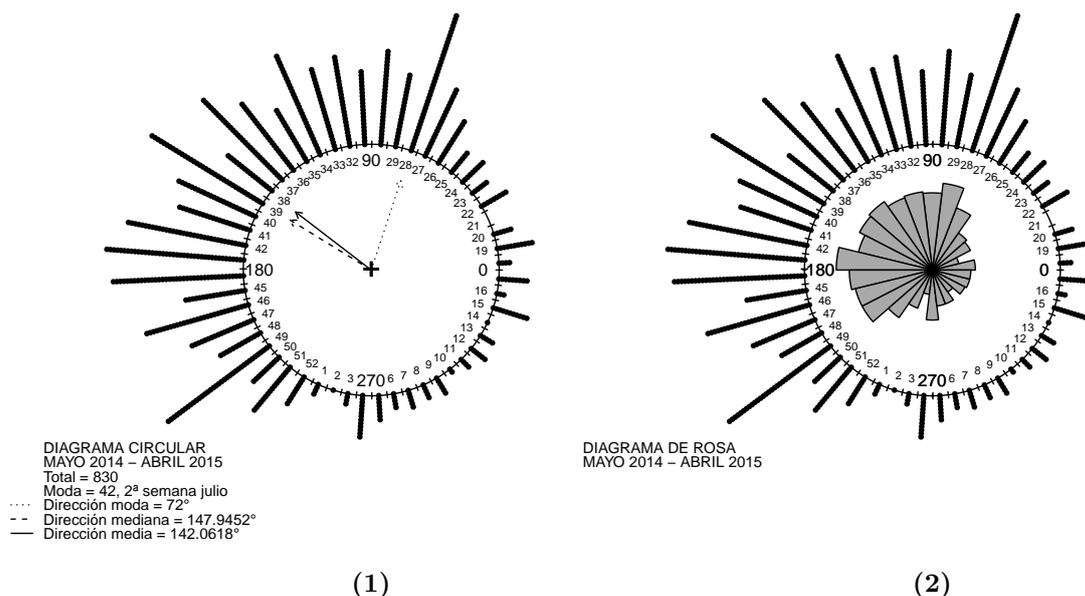


Figura 5.27: Diagramas circular a la izquierda y de rosa a la derecha para el período mayo 2014 - abril 2015.

Período 6: mayo 2014 - abril 2015	
Total	830
Moda	42
$\check{\theta}$	72°
$\tilde{\theta}$	147.9452°
$\bar{\theta}$	142.0618°
\bar{R}	0.4327795
\bar{b}_2	0.02071404
\hat{s}	0.0484883
\bar{a}_2	0.03968642

Tabla 5.16: Total, moda y medidas de localización, de dispersión y de sesgo para período mayo 2014 - abril 2015.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	28	Julio (2 ^a)
$\tilde{\theta}$	39	Septiembre (4 ^a)
$\bar{\theta}$	38	Septiembre (3 ^a)

Tabla 5.17: Tiempo al que se dirigen los estadísticos descriptivos circulares de localización, para el período mayo 2014 - abril 2015.

La medida $\bar{R} = 0.4327$ es señal, ya no de concentración, sino de una dispersión de los casos de dengue en un amplio rango de tiempo. Según el diagrama de rosa de la Figura 5.272, este rango es notorio desde la semana 26 (transición entre junio y julio) hasta la semana 49 (1^a de diciembre) del 2014, es decir, las frecuencias de dengue se presentan de manera considerable, casi de manera uniforme, en prácticamente la segunda mitad del 2014.

La circunferencia unitaria se ha cortado aproximadamente a la mitad del arco, que abarca de la semana 2 hasta la semana 21 (recorriendo la circunferencia en sentido antihorario); notar que la semana 2 a la 18 pertenecen al año 2015, y la semana 18 a la 21 son del 2014. Este arco presenta las más bajas frecuencias. Es decir, se ha cortado en el punto con dirección 318° (5.550147 radianes). Con este corte, se construye el histograma lineal de la Figura 5.28.

Las medidas $\bar{b}_2 = 0.0207$ y $\hat{s} = 0.0485$ son cercanas a cero y esto es un indicador de que sería razonable considerar en primera instancia un modelo simétrico para la distribución de los datos.

El histograma lineal de la Figura 5.28, muestra que las frecuencias se empiezan a elevar a partir de la semana 26 (transición entre junio y julio). A partir de aquí empieza un tramo de frecuencias que alternan en cantidad entre elevadas y no tan elevadas. Este comportamiento, aparentemente se termina después de la semana 49 (1ª de diciembre). A partir de aquí, las frecuencias comienzan un franco decrecimiento.

La medida $\bar{a}_2 = 0.0407$ es una cantidad que señala un aparente acercamiento a un comportamiento uniforme en los datos para este período.

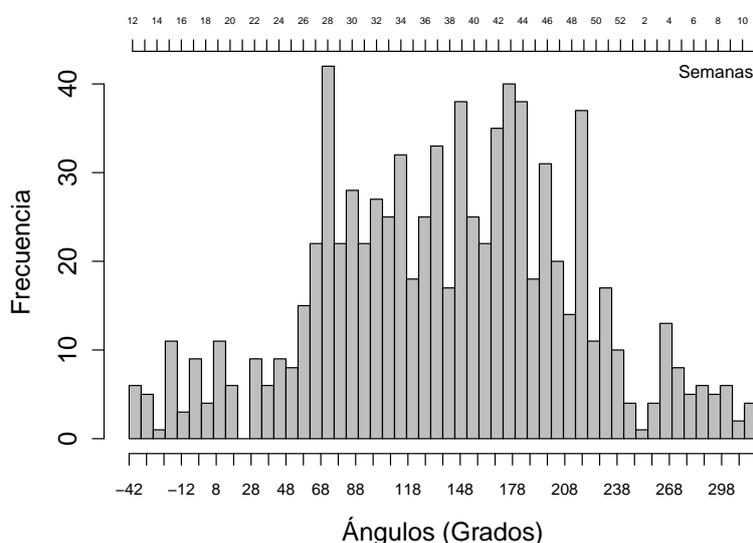


Figura 5.28: Histograma lineal de frecuencias de dengue, para el período mayo 2014 - abril 2015, con punto de corte en la dirección 318° (5.550147 radianes). A los valores negativos del principio hay que sumarles 360° para obtener valores en el rango de $(0, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba es de 0.4328 y arroja un p-valor muy cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de prueba es de 0.8948 y el p-valor es de 0.3709 . De lo anterior, la hipótesis nula de uniformidad es rechazada a cualquier nivel de significancia α frecuentemente utilizado, pues el p-valor es cercano a cero; en contraste, no hay evidencia suficiente para rechazar la hipótesis nula de simetría reflexiva alrededor de la dirección media, a cualquier nivel de significancia α frecuentemente utilizado. Por esto último, y por los valores cercanos a cero de las medidas de sesgo, se espera que un modelo que considere simetría reflexiva ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3 muestra que el mejor modelo que se ajusta a los caso de dengue para este período es el de una distribución Inversa Batschelet, considerando su parámetro de sesgo $\nu = 0$ y todos los demás distintos de cero, denominada la distribución “Inversa Batschelet Simétrica”.

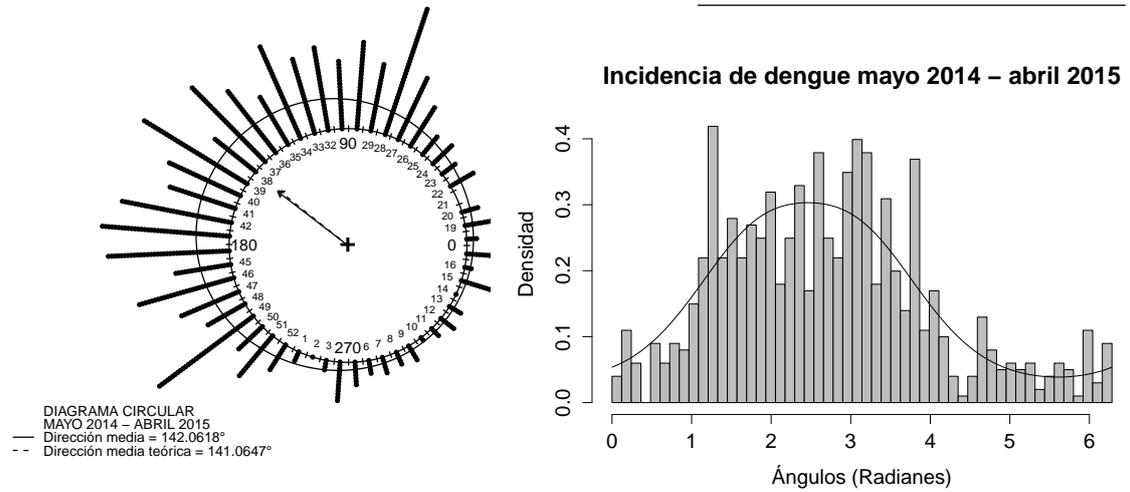
Período 6: mayo 2014 - abril 2015	
Mejor ajuste: Inversa Batschelet Simétrica	
Parámetro	Estimado
ξ	2.462044
κ	1.029367
ν	0
λ	-0.2890412

Tabla 5.18: Ajuste de un modelo Inverso Batschelet con $\nu = 0$, denominado “Inverso Batschelet Simétrico”, para el período mayo 2014 - abril 2015.

La Tabla 5.18, muestra los estimados de los parámetros que controlan la distribución Inversa Batchelet, obtenidos vía máxima verosimilitud. El valor de $\hat{\nu} = 0$ considera simetría reflexiva alrededor de la dirección media. El valor $\hat{\lambda} = -0.2890$ considera una distribución de cima plana en su presentación polar.

Cuando $\nu = 0$ (no se considera sesgo), una distribución Inversa Batschelet tiene como dirección media teórica $\xi(\text{mod } 2\pi)$, que coincidirá con la dirección modal y la dirección mediana poblacionales. La Figura 5.211 muestra el diagrama circular que incluye la dirección media teórica estimada y la dirección media muestral. Ambas son muy cercanas entre si. De hecho, se podría decir que la dirección media estimada vía máxima verosimilitud es una corrección mínima de la dirección media muestral.

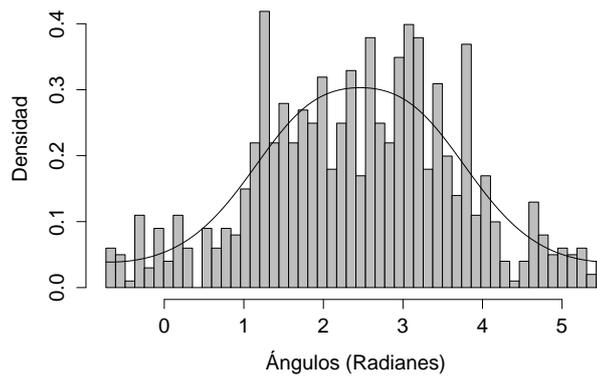
Las Figuras 5.292 y 5.293 exhiben la representación lineal de la distribución para este período, pero considerando diferentes punto de corte en $(0, 2\pi]$, en 0 en el caso de la primera y 5.5501 radianes en la segunda. La Figura 5.293 exhibe el comportamiento reflexivamente simétrico modelado para los datos, y por ende se observa hay un equilibrio de las masas de probabilidad antes y después de la dirección media.



(1)

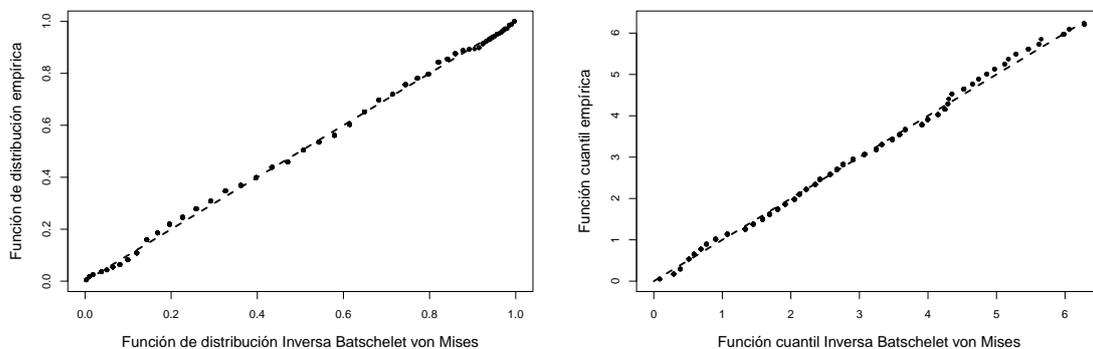
(2)

Incidencia de dengue mayo 2014 – abril 2015



(3)

Figura 5.29: Representaciones polar (izquierda), lineal con inicio en la dirección cero (derecha), y lineal con inicio en la dirección 318° (abajo), de la distribución Inversa Batschelet Simétrica ajustada para el período mayo 2014 - abril 2015.



(1)

(2)

Figura 5.30: Gráficos probabilidad-probabilidad (PP plot), a la izquierda, y cuantil-cuantil (QQ plot), a la derecha, de la distribución Inversa Batschelet Simétrica ajustada para el período mayo 2014 - abril 2015.

La Figura 5.30 exhibe que las gráficas P-P y Q-Q presentan la mayoría de sus puntos muy cerca de la línea de referencia, lo cual es una señal de un ajuste razonable.

En la prueba de bondad de ajuste, basado en la de uniformidad de Rayleigh, el valor del estadístico de prueba es de 0.0236 y el p-valor arrojado es de 0.629. Se concluye que para el comportamiento de los casos de dengue de este período el ajuste de una distribución Inversa Batschelet Simétrica es razonable, a cualquier nivel de significancia α frecuentemente utilizado.

6

Comparación entre los períodos de estudio

A continuación mostraremos las similitudes y diferencias encontradas entre los 6 períodos de estudio mayo - abril anuales. Hay que tener en cuenta que cuando se hable de “períodos impares” se estará haciendo referencia a los períodos de estudio mayo - abril anuales con rango 2009 - 2010, 2011 - 2012 y 2013 - 2014, que son precisamente los períodos que tienen como primer año de estudio uno impar, y que tienen número impar en el orden que se les ha asignado para estudiarlos en este trabajo; estos son 1, 3 y 5. Análogamente, cuando se hable de “períodos pares” se hará referencia a los períodos mayo - abril anuales con rango 2010 - 2011, 2012 - 2013 y 2014 - 2015, que tienen como primer año de estudio uno par, o sea, los que aparecen en el orden de 2, 4 y 6.

6.1. Comparación entre gráficos circulares

Las Figuras B.1, B.2 y B.3 del apéndice B ofrecen una comparación visual entre los períodos de estudio a través de los diagramas circulares, los diagramas de rosa, y los histogramas lineales, respectivamente. En cada uno de estos conjuntos de figuras, los gráficos de la izquierda pertenecen a los períodos impares 1, 3 y 5 en orden de arriba abajo. Los gráficos de la derecha pertenecen a los períodos pares 2, 4 y 6 en orden de arriba abajo.

La Figura B.1 muestra los diagramas circulares de los 6 períodos de estudio, y en ellos se incluyen las direcciones media, mediana y modal muestrales.

La dirección modal muestral se refiere al representante angular de la semana con

mayor frecuencia de casos de dengue en un período en particular. La Figura 6.1 muestra las modas muestrales obtenidas en los seis períodos de estudio; en conjunto los valores de la dirección modal muestran un comportamiento irregular. Sin embargo, en la Figura puede observarse que los valores de la moda para los períodos impares son cercanos entre sí y de hecho, apuntan hacia las semanas 40, 44 y 42, respectivamente; para los períodos pares se exhibe un comportamiento descendente a partir del período 2 y los valores respectivos apuntan a las semanas 44, 37 y 28.

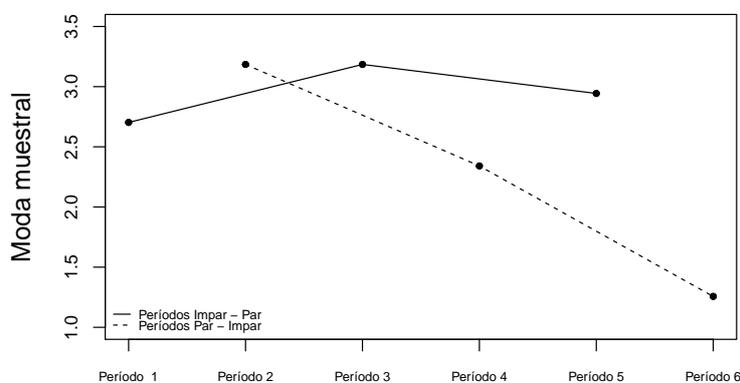


Figura 6.1: Valores de la dirección modal muestral obtenidos durante los 6 períodos de estudio mayo - abril anuales.

En cuanto a las direcciones media y mediana se puede decir que, en general, sus respectivos valores se encuentran muy cerca uno del otro en un período en particular, así que los comentarios se reducirán a la dirección media muestral.

En los períodos impares se puede ver que las direcciones medias varían, pero todas están relativamente cerca de 180° ; éstas hacen referencia a las semanas 44, 46 y 42, en orden de arriba abajo. Para los períodos pares, las direcciones medias están relativamente lejos de 180° , sin embargo, tienen valores muy similares entre ellas; éstas hacen referencia a las semanas 39, 38 y 38 (de nuevo) en orden de arriba abajo.

La Figura B.2 muestra los diagramas de rosa. Para los períodos impares 1 y 3 se exhibe, en general, una concentración que se hace notoria después de la dirección de 90° , de hecho, se podría decir que cerca de la mitad del arco de 90° a 180° ; esta concentración deja de ser notoria poco antes de terminar el primer año de estudio (antes de la semana 52). Ambos diagramas de rosa exhiben un sesgo en sentido antihorario.

Para los períodos pares 2 y 4, se exhibe una concentración que se hace notoria antes de la dirección de 90° , y que deja de serlo antes de terminar el primer año

de estudio, pero que lo hace mucho antes en comparación con los períodos impares pasados. Ambos diagramas de rosa exhiben un pequeño sesgo en sentido horario.

Para los últimos dos períodos no se exhibe una clara concentración, más bien se percibe cierto grado de dispersión de los casos de dengue.

La Figura B.3 muestra los histogramas lineales construidos a partir de un punto de corte ideal en los diagramas circulares. En estos histogramas, ir de izquierda a derecha equivale a recorrer la circunferencia unitaria en sentido antihorario.

Para los períodos impares 1 y 3, se exhibe una estructura similar de la distribución de los casos de dengue. De manera general, se puede decir que después de un tramo corto de bajas frecuencias, éstas comienzan a crecer rápidamente hasta alcanzar su punto máximo, para luego comenzar a decrecer de manera lenta. El sesgo antihorario es evidente. Ambos histogramas no comienzan en el mismo punto de corte, de hecho, si uno observa los diagramas circulares B.11 y B.13 a detalle, pareciera que para poder ver de manera visual las similitudes de ambas distribuciones hay que aplicar una rotación a alguna de ellas con respecto a la otra.

Para los períodos pares 2 y 4, también se exhibe cierta estructura similar en la distribución de los casos de incidencia, de hecho ambos histogramas comienzan en el mismo punto de corte. En general, se tiene que después de un tramo de frecuencias bajas éstas comienzan un franco crecimiento y, después de alcanzar un punto máximo, comienza un tramo corto en que se mantienen frecuencias altas, para luego tener una caída abrupta de los casos de dengue.

Cuando se esperaba tener una estructura similar a la vista en los períodos 1 y 3, en el período 6 se tiene un histograma que logra reflejar una simetría reflexiva. Este histograma consiste en que, desde el inicio, se exhibe un franco crecimiento de las frecuencias que, después de alcanzar un punto máximo, comienzan ahora un franco decrecimiento.

También, cuando se esperaba tener una estructura similar a la vista en los períodos 2 y 4, en el período 6 se tiene un histograma que logra reflejar cierto tipo de simetría reflexiva. Este histograma consiste en que después de un tramo corto de frecuencias relativamente bajas, comienza un tramo de frecuencias más altas que entre ellas exhiben un comportamiento alternado, y que se logran mantener hasta dar paso a otro tramo corto de frecuencias relativamente bajas.

6.2. Comparación muestral

Período	Estadísticos descriptivos circulares			
	$\bar{\theta}$	\bar{R}	\bar{b}_2	\bar{a}_2
(1) Mayo 2009 - Abril 2010	(3.1396,3.1406) 3.1401	(0.6351,0.6670) 0.6510	(-0.134,-0.0968) -0.1154	(0.2999,0.3467) 0.3233
(2) Mayo 2010 - Abril 2011	(2.5527,2.5541) 2.5534	(0.6387,0.6732) 0.6559	(0.0106,0.056) 0.0333	(0.1529,0.2081) 0.1805
(3) Mayo 2011 - Abril 2012	(3.4511,3.4517) 3.4514	(0.7219,0.7401) 0.7310	(-0.1266,-0.1052) -0.1159	(0.3559,0.3869) 0.3714
(4) Mayo 2012 - Abril 2013	(2.4758,2.4768) 2.4763	(0.5860,0.6138) 0.5999	(0.0196,0.0503) 0.0349	(0.1852,0.2235) 0.2043
(5) Mayo 2013 - Abril 2014	(2.9213,2.9230) 2.9221	(0.4897,0.5311) 0.5104	(-0.0157,0.031) 0.0076	(0.1262,0.1758) 0.151
(6) Mayo 2014 - Abril 2015	(2.4775,2.4813) 2.4794	(0.3929,0.4713) 0.4321	(-0.0244 0.0662) 0.0209	(-0.0086,0.0855) 0.0384

Tabla 6.1: Intervalos del 95 % de confianza para los estadísticos descriptivos circulares $\bar{\theta}$, \bar{R} , \bar{b} , y \bar{a}_2 para cada período de estudio. Abajo de cada intervalo, está el estimado puntual del respectivo estadístico.

La Tabla 6.1 muestra los estimados de los estadísticos descriptivos circulares $\bar{\theta}$, \bar{R} , \bar{b} , y \bar{a}_2 con sus respectivos intervalos de confianza del 95 %. Estos intervalos se han obtenido usando los resultados debido a Pewsey (2004), vistos en la Sección 4.3.2. Estos resultados también ofrecen una corrección insesgada a los estimados de los estadísticos circulares convencionales, sin embargo, la diferencia entre los estimados originales y sus correcciones es prácticamente cero, por lo que se han tomado los valores originales de estos estadísticos descriptivos.

Notar que en el caso $\bar{\theta}$, los resultados de Pewsey (2004) ofrecen intervalos de confianza de poca amplitud, esto indica que el error estándar es pequeño y, por lo tanto, el rango de posibles valores para $\bar{\theta}$ es pequeño.

6.2.1. Sobre la dirección media muestral $\bar{\theta}$

La Tabla 6.2 muestra que, en general, la dirección mediana muestral resulta ser muy cercana a la dirección media muestral, por lo tanto, reducimos nuestro análisis sólo acerca de la dirección media muestral.

Período	Descriptivos circulares de localización					
	Dirección media muestral $\bar{\theta}$			Dirección mediana muestral $\tilde{\theta}$		
	Estimado	Semana	Mes	Estimado	Semana	Mes
1	3.1401	44	Octubre-Noviembre	3.0641	43	Octubre (4 ^a)
2	2.5534	39	Septiembre (4 ^a)	2.5821	39	Septiembre (4 ^a)
3	3.4514	46	Noviembre (2 ^a)	3.4256	46	Noviembre (2 ^a)
4	2.4763	38	Septiembre (3 ^a)	2.4616	38	Septiembre (3 ^a)
5	2.9221	42	Octubre (3 ^a)	2.9436	42	Octubre (3 ^a)
6	2.4794	38	Septiembre (3 ^a)	2.5821	39	Septiembre (4 ^a)

Tabla 6.2: Semanas a las que apuntan las direcciones media y mediana muestrales y su ubicación en los meses de un año normal. Los meses mencionados corresponden al primer año de estudio.

Sobre los valores de $\bar{\theta}$ estimados para cada período de estudio considerado, la Tabla 6.2 muestra que éstos se mueven globalmente en un rango de aproximadamente (2.4, 3.5) radianes, el cual incluye representantes de días pertenecientes a las semanas 38, 39, 42, 44 y 46, las cuales caen entre los meses de septiembre (mitad) y noviembre (mitad).

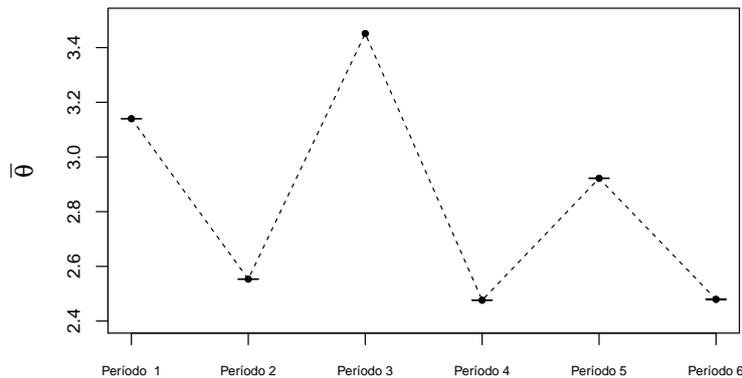


Figura 6.2: Valores estimados para la dirección media muestral $\bar{\theta}$ obtenidos durante los 6 períodos de estudio mayo - abril anuales.

Además, la Figura 6.2 muestra una gráfica en la que se exhibe un comportamiento alternado entre los valores de $\bar{\theta}$. Se muestra que los estimados para los períodos impares son relativamente más grandes que los correspondientes a los períodos pares. Se alterna entre un valor para un período impar y otro relativamente más chico para un período par.

Si en esta misma gráfica observamos de manera independiente los estimados de $\bar{\theta}$ para los períodos pares, notaremos que éstos parecieran tener un valor similar, están muy cerca de 2.5 radianes, de hecho pueden ser incluidos en un rango de aproximadamente (2.4, 2.6) radianes, que incluyen representantes de días pertenecientes a las semanas 38 y 39, las cuales se ubican en la segunda parte del mes de

septiembre.

En contraste, los estimados de $\bar{\theta}$ para los períodos impares se aprecia cierta variación entre ellos. Estos valores están arriba de 3 radianes, de hecho, pueden ser incluidos en un rango de aproximadamente (2.9, 3.5) radianes, que incluyen representantes de días pertenecientes a las semanas 42, 44 y 46, las cuales se ubican en el rango después de la mitad del mes de octubre hasta después de la mitad del mes de noviembre.

6.2.2. Sobre la longitud media resultante muestral \bar{R}

Período	Estimado	Semanas	Rango de semanas	Meses que abarca
1	0.6511	36 - 51	16	Septiembre (1ª) - Diciembre (3ª)
2	0.6560	30 - 47	18	Julio (4ª) - Noviembre (3ª)
3	0.731	40 - 51	12	Octubre (1ª) - Diciembre (3ª)
4	0.5999	28 - 47	20	Julio (2ª) - Noviembre (3ª)
5	0.511	30 - 51	22	Julio (4ª) - Diciembre (3ª)
6	0.4327	26 - 49	24	Transición Junio-Julio - Diciembre (1ª)

Tabla 6.3: Rango de la concentración de los casos de dengue en casa período de estudio. Los meses mencionados corresponden al primer año de estudio.

La Tabla 6.3 ofrece los estimados para \bar{R} y el rango de concentración en el tiempo. De manera global, entre 2009 y 2014, se podría decir que las altas frecuencias de casos de dengue ocurren en un rango de entre principios de julio (semana 26) hasta la tercera semana de diciembre (la 51).

Aunque la semana inicial en la que se hace notoria la concentración difiere entre los períodos de estudio, llama la atención que las semanas finales son similares entre los períodos pares e impares. Para los períodos impares, la concentración deja de ser notoria después de la tercera semana de diciembre (semana 51). Para los períodos pares, la concentración deja de ser notoria después de la semana 47 o 49, que son la tercera semana de noviembre y la primera de diciembre, respectivamente; sin embargo, la semana final 49 pertenece al último período de estudio, en la que ya no hay concentración sino una considerable dispersión de los casos de dengue en el rango respectivo.

Por otra parte, la Figura 6.3 muestra los estimados junto con su respectivo intervalo de confianza del 95%. Notar que la longitud de los intervalos en los períodos 1 y 2 ofrecen una inferencia similar, de hecho los estimados para esos períodos son similares y hacen referencia a un concentración considerable de los casos de dengue alrededor de la dirección media. Sin embargo, en el período 3 hay un salto hacia un estimado relativamente más grande que los dos pasados, lo que indica

una concentración alta. Después del período 3, se observa un decrecimiento en los valores estimados, lo que indica que los casos de dengue se concentran menos, o bien, que se van dispersando cada vez más alrededor de la dirección media. De hecho, los dos últimos intervalos son cada vez más amplios; en el intervalo del estimado del período 5, ya se consideran valores de dispersión (abajo de 0.5) para los estimados de \bar{R} , mientras que en intervalo para el período 6, son todos valores referentes a una dispersión baja en los casos de dengue (ver Tabla 6.1).

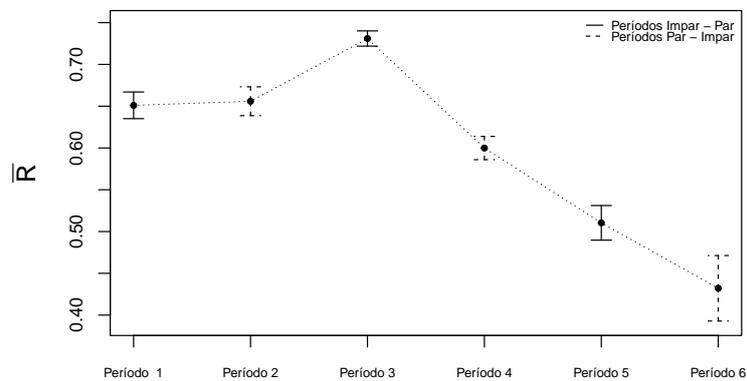


Figura 6.3: Valores estimados para longitud media muestral \bar{R} obtenidos durante los 6 períodos de estudio mayo - abril anuales.

6.2.3. Sobre la medida de sesgo \bar{b}_2

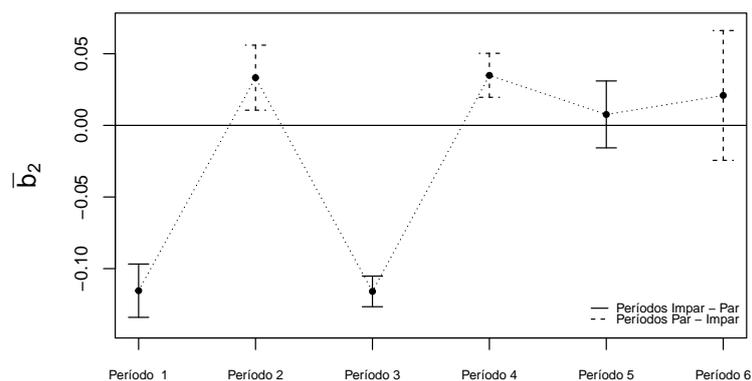


Figura 6.4: Valores estimados para la medida de sesgo \bar{b}_2 obtenidos durante los 6 períodos de estudio mayo - abril anuales.

La Figura 6.4 contiene la gráfica de los valores estimados para la medida de sesgo \bar{b}_2 , cada uno junto con su respectivo intervalo del 95 % de confianza. Se puede observar

un comportamiento alternado que es más marcado entre los períodos 1 a 4. Se podría decir que los estimados de \bar{b}_2 para los períodos impares son relativamente más chicos que los correspondientes a los períodos pares.

Los estimados de la medida de sesgo para los períodos 1 y 3 son relativamente grandes y negativos e indican un sesgo antihorario; están relativamente alejados de cero y sus intervalos de confianza nos indican que son significativamente distintos de cero. Además, estos dos estimados son muy similares entre sí (ver Tabla 6.1). Sin embargo, cuando para el período 5 se esperaba un valor similar a los obtenidos en 1 y 4, se da un salto muy considerable entre estos valores y el último, el cual ahora tiene signo positivo, aunque es relativamente cercano a cero y su intervalo de confianza, además de considerar valores positivos, sigue considerando valores negativos.

En contraste, los estimados para los períodos pares 2 y 4 son positivos y relativamente cercanos a cero, aunque sus intervalos de confianza nos indican que son significativamente distintos de él; estos valores son similares entre sí (ver Tabla 6.1). Para el período 6 se tiene un estimado acorde a lo que se esperaría según lo observado en los períodos pares anteriores; este valor está más cerca de cero y por sí mismo puede ser incluido en un amplio intervalo de confianza que ahora considera tanto valores positivos como negativos.

Los intervalos de confianza de los estimados para los períodos 5 y 6 no sólo incluyen valores positivos y negativos, sino que también al cero, lo cual es un indicador de que la distribución subyacente de los casos de dengue, de ambos períodos, es cercana a una distribución que considere simetría reflexiva alrededor de la dirección media.

Período	p-valor
1	9.02586e-33
2	0.004164311
3	8.069242e-94
4	8.70234e-06
5	0.5231729
6	0.3708811

Tabla 6.4: P-valor de la prueba de simetría reflexiva para muestras grandes de Pewsey aplicada a cada período de estudio.

La Tabla 6.4 exhibe los p-valores de la prueba de simetría reflexiva de Pewsey, para muestras grandes, aplicada a cada período de estudio. La hipótesis nula de simetría reflexiva alrededor de la dirección media es rechazada en los períodos del 1 al 4 a cualquier nivel de significancia α frecuentemente usado. En contraste, no

hay evidencia suficiente para rechazar la pasada hipótesis nula en los períodos 5 y 6 a cualquier nivel de significancia α razonable.

6.2.4. Sobre la medida de curtosis \bar{a}_2

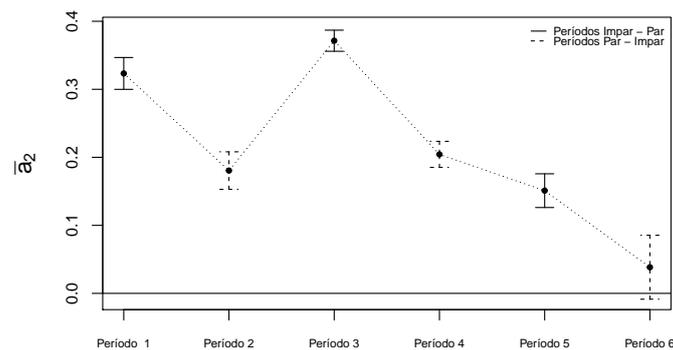


Figura 6.5: Valores estimados para la medida de curtosis \bar{a}_2 obtenidos durante los 6 períodos de estudio mayo - abril anuales.

La Figura 6.5 muestra los valores estimados para \bar{a}_2 graficados juntos con sus respectivos intervalos de confianza del 95%. De manera global se podría decir que estos estimados consideran una distribución con picudez no extrema (los valores cercanos a 1 corresponden a una máxima picudez), pues sus valores se dan en el rango de aproximadamente (0.038,0.38), inclusive el valor cercano a cero de algunos de ellos exhibe una característica de uniformidad.

Hay un evidente comportamiento alternado entre los estimados de los períodos del 1 al 4; los estimados para los períodos impares son relativamente más grandes que los períodos pares. Sin embargo, después del período 4 se exhibe un comportamiento decreciente. De hecho, para el último período el estimado se incluye en un amplio intervalo que ya considera al cero casi al extremo, el cual es un indicador de un comportamiento cercano a la uniformidad.

Sin embargo, la prueba de simetría uniformidad de Rayleigh, para una alternativa unimodal, es rechazada cuando se aplica a todos los períodos, arrojando p valores muy cercanos a cero.

6.3. Comparación entre los modelos ajustados

Los cuatro primeros períodos consideran un grado de sesgo a contemplar al ajustar a los datos un modelo circular de probabilidad; en los períodos 1 y 3 se tiene que considerar un sesgo antihorario, mientras que en los períodos 2 y 4 un sesgo horario.

La familia de distribuciones flexibles Inversa Batschelet ofrece modelos que toman en cuenta no sólo la localización y la concentración, sino también el sesgo y la curtosis. Cuando los cuatro parámetros que controlan estas características en la distribución Inversa Batschelet son todos distintos de cero, entonces se dice que es un modelo completo de esta familia de distribuciones. A manera de evaluar la flexibilidad de estas distribuciones se ha ajustado un modelo completo a cada conjunto de datos de cada período de estudio.

Período	Parámetros			
	ξ	κ	ν	λ
(1) Mayo 2009 - Abril 2010	(3.0067,3.2765) 3.1416	(1.5655,1.7008) 1.6332	(-0.0108,0.1475) 0.0683	(0.2598,0.3864) 0.3231
(2) Mayo 2010 - Abril 2011	(2.1168,2.4677) 2.2922	(1.9333,2.2434) 2.0883	(-0.3305, -0.0572) -0.1939	(-0.3601,-0.2113) -0.2857
(3) Mayo 2011 - Abril 2012	(-2.1311,-1.9995) -2.0653	(2.1431,2.2934) 2.2183	(0.4659,0.5491) 0.5075	(0.0616,0.1447) 0.1032
(4) Mayo 2012 - Abril 2013	(2.0860,2.3133) 2.1997	(1.4757,1.6044) 1.5400	(-0.2872,-0.1237) -0.2055	(-0.0952,0.0200) -0.0376
(5) Mayo 2013 - Abril 2014	(2.7211,3.0103) 2.8657	(1.1160,1.2525) 1.1842	(-0.1485,0.0590) -0.0447	(-0.0596,0.1466) 0.0435
(6) Mayo 2014 - Abril 2015	(1.9205,2.7275) 2.3240	(0.8964,1.1577) 1.0270	(-0.4936,0.2349) -0.1293	(-0.4596,-0.0902) -0.2749

Tabla 6.5: Intervalos del 95 % de confianza para los parámetros ξ , κ , ν y λ del modelo Inversa Batschelet Completo (base von Mises) que se ajusta para cada período de estudio. Abajo de cada intervalo está el estimado puntual del respectivo parámetro.

La Tabla 6.5 contiene los estimados de los parámetros ϵ , κ , ν y λ para cada período de estudio. Como ya se ha dicho, cada modelo es completo. Aprovechando que estos parámetros son estimados vía máxima verosimilitud, sus respectivos intervalos de confianza son obtenidos utilizando teoría asintótica normal.

La Figura 6.6 muestra las gráficas de comparación de los estimados calculados de un parámetro en particular. Para el parámetro de localización ξ , la gráfica en la Figura 6.61 exhibe un comportamiento parecido al de su contraparte muestral $\bar{\theta}$ en la Figura 6.2. El rango de valores para ξ no es el mismo que el de $\bar{\theta}$, de hecho el primero es un poco más amplio. Básicamente, se tiene un comportamiento alternado donde los estimados para los períodos impares son relativamente más grandes que los de los períodos pares. Los estimados para los períodos pares tienen

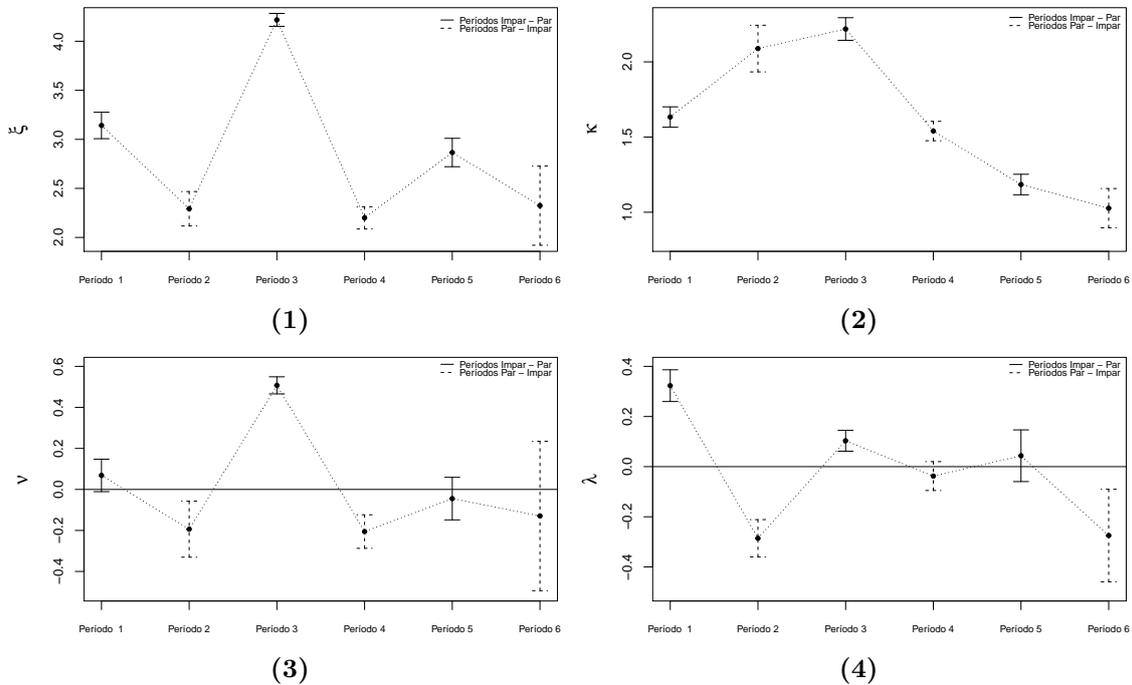


Figura 6.6: Comparación de los valores obtenidos en los 6 períodos de estudio cuando a cada uno se le ajusta un modelo Inversa Batschelet Completo: (a) para el parámetro de localización ξ , (b) para el parámetro de concentración κ , (c) para el parámetro de sesgo ν , (d) para el parámetro de curtosis λ .

valores cercanos entre sí. En esta ocasión, sí se tienen intervalos de confianza para cada valor, y en el caso par, los valores obtenidos pueden ser todos ubicados en el primer intervalo de confianza par.

Para el parámetro de concentración κ , la gráfica en la Figura 6.62 tiene un comportamiento decreciente al final, similar a lo que sucede con su contraparte muestral \bar{R} en la Figura 6.3. Tienen diferentes rangos de valores.

Para el parámetro de sesgo ν , la Figura 6.63 muestra una gráfica que se asemeja a la de ξ y que refleja un comportamiento alternado. Ahora, si consideramos $-\nu$ en la Figura 6.7, éste exhibe un comportamiento similar al visto para su contraparte muestral \bar{b}_2 en la Figura 6.4. Sólo que en el caso de $-\nu$, para el período de estudio 1 se tiene un estimado que está relativamente cerca del cero y su intervalo de confianza lo incluye, dejando la posibilidad de que este valor pueda considerarse como estadísticamente cero. También, los últimos dos intervalos de confianza incluyen al cero y los estimados de los períodos 5 y 6 están relativamente cerca de él, lo cual, como ya se dijo, es un indicador de que pondrían tomarse como estadísticamente cero para considerar una distribución cercana a una reflexivamente simétrica.

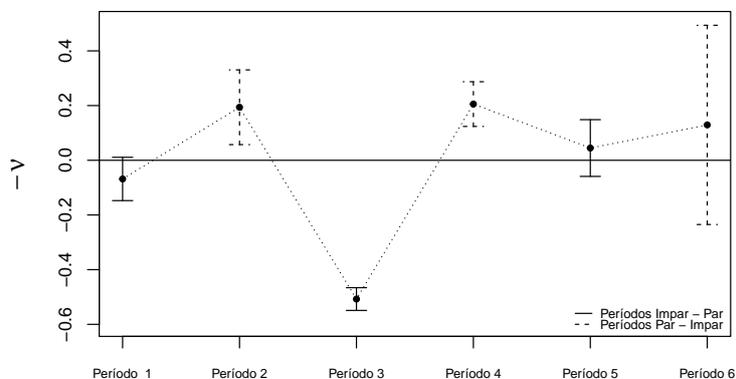


Figura 6.7: Comparación de los valores para $-\nu$ obtenidos en los 6 períodos de estudio cuando a cada uno se le ajusta un modelo Inversa Batschelet Completo.

Para el parámetro de curtosis λ , la Figura 6.64 muestra una gráfica un tanto diferente a la de su contraparte muestral \bar{a}_2 . Aquí, la gráfica de λ exhibe un comportamiento alternado. Los estimados para los períodos impares son positivos y, de acuerdo con los valores observados, esto indica que la forma polar de una distribución particular puede tomar valores no extremos de picudez; de hecho, el intervalo de confianza del estimado del período 5 incluye al cero, lo cual es un indicador de poder considerarlo estadísticamente igual a él. En contraste, los estimados para los períodos pares son negativos, lo que indica que la forma polar de una distribución particular considera valores que indican una cima plana (sin pico evidente); de hecho, el intervalo de confianza del estimado del período 4 incluye al cero, lo cual es un indicador de poder considerarlo estadísticamente igual a él.

Los modelos Inversa Batschelet Completos no siempre ofrecen los mejores ajustes según los números-criterio AIC y BIC de la Tabla A.3 del Apéndice A. De hecho, como hemos visto, algunos de los estimados nos “invitan” a considerarlos estadísticamente cero; aunque, por ejemplo, en el caso del ajuste para el período 1, no considerar estadísticamente cero el parámetro de sesgo ν ofrece un mejor ajuste. Sin embargo, considerar λ estadísticamente cero en el período 4, λ y ν en el período 5, y ν en el período 6, ofrece un mejor ajuste que el modelo completo (ver Tabla A.3).

La Tabla 6.6 ofrece los nuevos estimados del modelo Inversa Batschelet al tomar en cuenta las consideraciones antes mencionadas. Los primeros tres períodos consideran un modelo Inversa Batschelet Completo, el período 4 considera un submodelo “von Mises Sesgado”, el período 5 considera un submodelo von Mises, y el período 6 considera un submodelo Batschelet Simétrico.

La Tabla 6.7 muestra los p-valores de la prueba de bondad de ajuste, basado en la

Período	Parámetros			
	ξ	κ	ν	λ
(1) Mayo 2009 - Abril 2010	3.1416	1.6332	0.0683	0.3231
(2) Mayo 2010 - Abril 2011	2.2922	2.0883	-0.1939	-0.2857
(3) Mayo 2011 - Abril 2012	4.2179	2.2183	0.5075	0.1032
(4) Mayo 2012 - Abril 2013	2.2017	1.5226	-0.2037	0
(5) Mayo 2013 - Abril 2014	2.9221	1.1929	0	0
(6) Mayo 2014 - Abril 2015	2.4620	1.0293	0	-0.2890

Tabla 6.6: Estimados de máxima verosimilitud de los parámetros ξ , κ , ν y λ del modelo Inversa Batschelet (base von Mises). Estos parámetros han dado el mejor ajuste para cada período de estudio.

de uniformidad de Rayleigh, aplicada a cada ajuste en su período correspondiente. Todas los p-valores no son significativos, lo que indica que es razonable el ajuste hecho en cada período. En el caso del período 2, una corrección en la base de datos permite tener un p-valor más grande (ver Sección 5.9).

Período	Ajuste	p-valor
1	Inversa Batschelet Completa	0.5825
2	Inversa Batschelet Completa	0.0262
3	Inversa Batschelet Completa	0.3871
4	Inversa Batschelet $\lambda = 0$	0.4141
5	Inversa Batschelet $\lambda = \nu = 0$	0.4325
6	Inversa Batschelet $\nu = 0$	0.629

Tabla 6.7: P-valor de la prueba de bondad de ajuste, basada en la de uniformidad de Rayleigh para muestras unimodales, aplicada a cada período de estudio.

Un modelo Inversa Batschelet permite saber cuál será la dirección modal teórica a través de la expresión $\xi - 2\nu(\text{mod } 2\pi)$. La Tabla 6.8 muestra las direcciones modales teóricas estimadas para cada período de estudio y la Figura 6.8 muestra estos valores graficados. De manera global, los estimados se pueden ubicar en un intervalo de aproximadamente (2.46, 3.21) radianes, el cual incluye representantes de días pertenecientes a las semanas 38 a la 44. Esta gráfica exhibe un claro comportamiento alternado. Las direcciones modales estimadas para los períodos impares son relativamente más grandes que las de los períodos pares. Por sí mismas, las direcciones modales de los períodos impares varía y pueden ubicarse en un intervalo de aproximadamente (2.92, 3.21) radianes, que incluye representantes de días pertenecientes a las semanas 42 a la 44. Análogamente, por sí mismas las direcciones modales de los períodos pares parecen tener un comportamiento decreciente y pueden ubicarse en un intervalo de aproximadamente (2.46, 2.68)

radianes, que incluyen representantes de días pertenecientes a las semanas 38 a la 40.

Período	Dirección modal teórica $\xi - 2\nu$		
	Estimado	Semana	Mes
1	3.0049	42 - 43	Octubre (3 ^a - 4 ^a)
2	2.6799	40	Octubre (1 ^a)
3	3.2029	44	Octubre - Noviembre
4	2.6090	39	Septiembre (4 ^a)
5	2.9221	42	Octubre (3 ^a)
6	2.4620	38	Septiembre (3 ^a)

Tabla 6.8: Valores estimados de la dirección modal teórica y las semanas a las que apuntan.

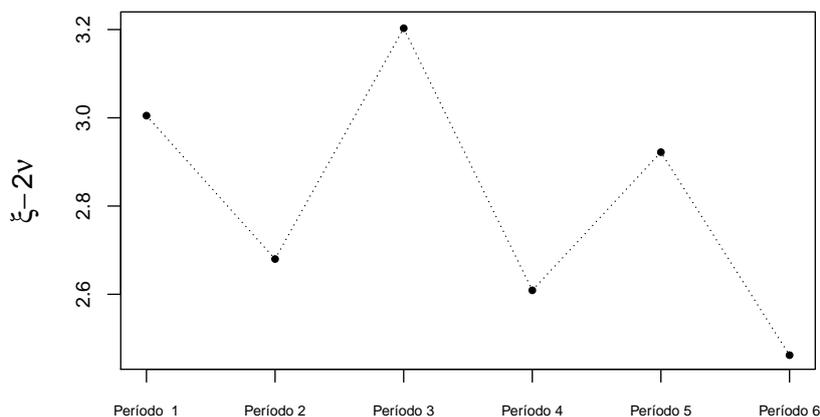


Figura 6.8: Comparación de los valores para $\xi - 2\nu$ obtenidos en los 6 períodos de estudio.

6.4. Resumen de los casos de dengue en dos períodos de estudio generales

El ciclo mayo - abril anual ha sido estudiado abarcando años que se dan en el rango de 2009 a 2015 y, a su vez, esto ha dado lugar a seis períodos de estudio. La información en ellos puede ser resumida en un sólo “círculo resumen” que simbolice el ciclo mayo - abril anual. Sin embargo, resumir todo en un sólo ciclo enmascararía las similitudes cualitativas encontradas entre los períodos pares e impares, así como las diferencias cualitativas observadas entre cada grupo.

Por ende, se han considerado dos circunferencias resumen, una para los períodos impares llamada el “Período resumen 1”, y otra para los períodos pares llamada

el “Período resumen 2”. En cada uno de ellos se ha hecho un análisis descriptivo y se ha ajustado un modelo circular.

Aunque en los dos primeros períodos impares se ha considerado sesgo antihorario en sus respectivos conjuntos de datos, para el último no ha habido suficiente evidencia para rechazar la hipótesis nula de simetría reflexiva alrededor de la dirección media; de hecho, considerar que la distribución de los datos es reflexivamente simétrica ha dado el mejor ajuste de un modelo de probabilidad circular. Por esto, el último período impar se ha considerado también en el Período Resumen 1, pues aunque no posee sesgo antihorario, al menos estadísticamente hablando no considera sesgo horario, además de que la dirección media muestral es relativamente cerca a 180° como lo son las respectivas para los dos primeros períodos impares.

Por un argumento análogo al anterior, se ha incluido el último período par al Período Resumen 2. Sólo que aquí, la direcciones medias muestrales de los tres períodos pares están relativamente lejanas a 180° y exhiben valores cercanos entre sí.

6.4.1. Resumen 1: Mayo Año Impar - Abril Año Par

Análisis descriptivo

La Figura 6.9 y la Tabla 6.9 resumen los gráficos y estadísticos descriptivos circulares para la unión de los datos de los períodos 1, 3 y 5, es decir, mayo 2009 - abril 2010, mayo 2011 - abril 2012, y mayo 2013 - abril 2014. Para este período resumen, se han registrado en total 13460 casos de dengue en el estado. La semana con mayor cantidad de casos fue la semana 44 (transición entre octubre y noviembre) del Año Impar.

El diagrama circular en la Figura 6.91 y la Tabla 6.10, muestran hacia qué semanas apuntan las direcciones $\tilde{\theta}$, $\hat{\theta}$ y $\bar{\theta}$. Si se hablara de un año impar normal, se tendría que la $\tilde{\theta} = \hat{\theta} = 182.4658^\circ$, es decir, es la semana 44 (transición entre octubre y noviembre) en la que se da la mayor frecuencia de dengue además de que en ella se alcanza la mitad de los casos registrados; por último, la dirección $\bar{\theta} = 188.2392^\circ$ mostraría que las grandes frecuencias de incidencia se dan alrededor de la semana 45, a principios de noviembre.

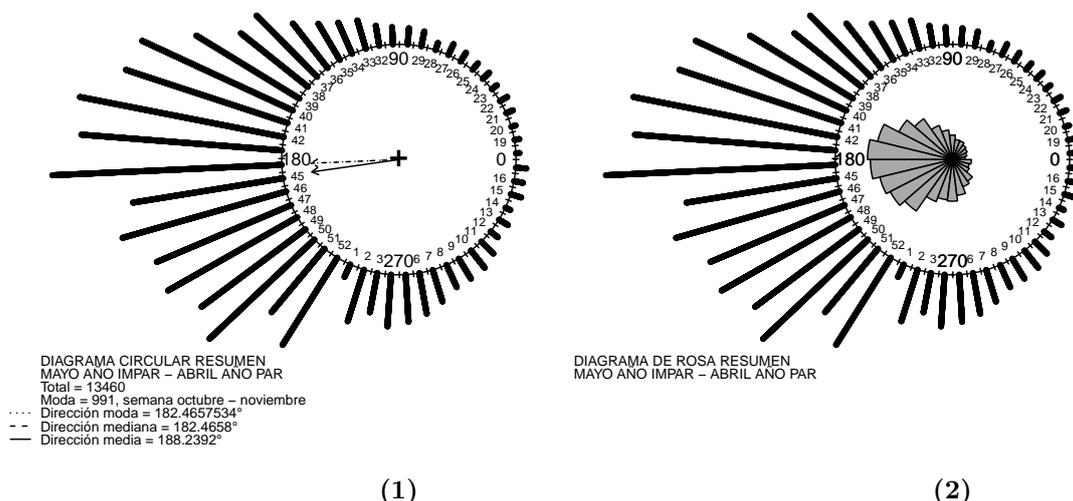


Figura 6.9: Diagramas circular a la izquierda y de rosa a la derecha para el período resumen 1.

Período Resumen 1: mayo año impar - abril año par	
Total	13460
Moda	991
$\check{\theta}$	182.4658°
$\tilde{\theta}$	182.4658°
$\bar{\theta}$	188.2392°
\bar{R}	0.6493943
\bar{b}_2	-0.06807845
\hat{s}	-0.3279301
\bar{a}_2	0.3021823

Tabla 6.9: Total, moda y medidas de localización, de dispersión y de sesgo para el período resumen 1.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	44	Octubre - Noviembre
$\tilde{\theta}$	44	Octubre - Noviembre
$\bar{\theta}$	45	Noviembre (1 ^a)

Tabla 6.10: Semanas a las que se dirigen los estadísticos descriptivos circulares de localización, para el período resumen 1.

La medida $\bar{R} = 0.6494$ ofrece una señal de una concentración de los casos de dengue. Si se dividiera un año impar normal en cuatro partes, con ayuda del diagrama de rosa de la Figura 6.92, se notaría que esta concentración ocurre después de la mitad del tercer cuarto hasta antes del final del último cuarto del año. Es decir, se presenta de la semana 36 (1^a de septiembre) hasta la semana 51 (3^a de diciembre) del año.

La circunferencia unitaria para este período resumen se ha cortado aproximadamente a la mitad del arco, que abarca de la semana 1 a la 29 (recorriendo la circunferencia en sentido antihorario), que es el arco en el que se presenta un tramo de bajas frecuencias de casos de dengue. Más precisamente, se ha cortado en el punto con dirección 38° (0.6632 radianes). Con base en este corte, se construye el histograma lineal de la Figura 6.10.

Las medidas $\bar{b}_2 = -0.1159$ y $\hat{s} = -0.8307$ revelan un sesgo en sentido antihorario respecto de dirección media, y el diagrama de rosa de la Figura 5.152 confirma esta situación. Este diagrama, junto con el histograma lineal de la Figura 5.16, muestran que a partir de la semana 37 (2^a de septiembre) del año impar empieza

un aumento rápido de los casos de dengue, que se prolonga hasta la semana 44 (transición entre octubre y noviembre), y luego de aquí, comienza un decrecimiento lento de los casos que continúa hacia la semana 11 (3^a de marzo) del año par.

La medida de curtosis $\bar{a}_2 = 0.3715$ es señal de un picudez no extrema en la forma polar de la distribución de los datos (ver diagrama circular de la Figura 6.91).

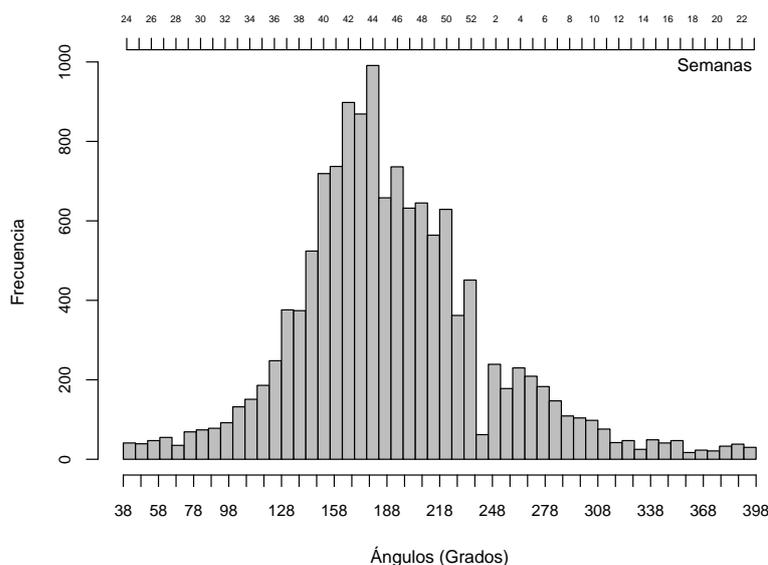


Figura 6.10: Histograma lineal de frecuencias de dengue, para el período resumen 1, con punto de corte en la dirección 38° (0.6632251 radianes). A los valores arriba de 360° del final, se les debe restar esta cantidad para obtener valores en el rango de $(0^\circ, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba calculado es de 0.6494 y arroja un p-valor cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de prueba es de 14.2805 y el p-valor es de $2.896533e-46$. Con p-valores muy cercanos a cero, las hipótesis nulas de uniformidad y de simetría reflexiva se rechazan para cualquier nivel de significancia α frecuentemente utilizado. Por lo anterior, se espera que un modelo que considere sesgo ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3, muestra que el mejor modelo que se ajusta a los casos de dengue para este período resumen es el de una distribución Inversa Batschelet Completa, es decir, con todos sus parámetros distintos de cero.

Período Resumen 1: mayo año impar - abril año par	
Mejor ajuste: Inversa Batschelet Completo	
Parámetro	Estimado
ξ	3.720231
κ	1.659224
ν	0.2971027
λ	0.2241965

Tabla 6.11: Ajuste de un modelo Inverso Batschelet Completo, con base von Mises, para el período resumen 1.

La Tabla 6.11, muestra los estimados de los parámetros que controlan la distribución Inversa Batschelet obtenidos vía máxima verosimilitud. El estimado $\hat{\nu} = 0.2971$ considera un sesgo en sentido antihorario. El estimado $\hat{\lambda} = 0.2241$ considera una distribución con picudez no extrema en su presentación polar.

La dirección modal teórica de la distribución, estimada por el modelo a través de la expresión $\xi - 2\nu(\text{mod } 2\pi)$, es de 3.1260 radianes, la cual apunta hacia las semanas 43 y 44, que son la cuarta semana de octubre y la semana transición de octubre a noviembre, respectivamente.

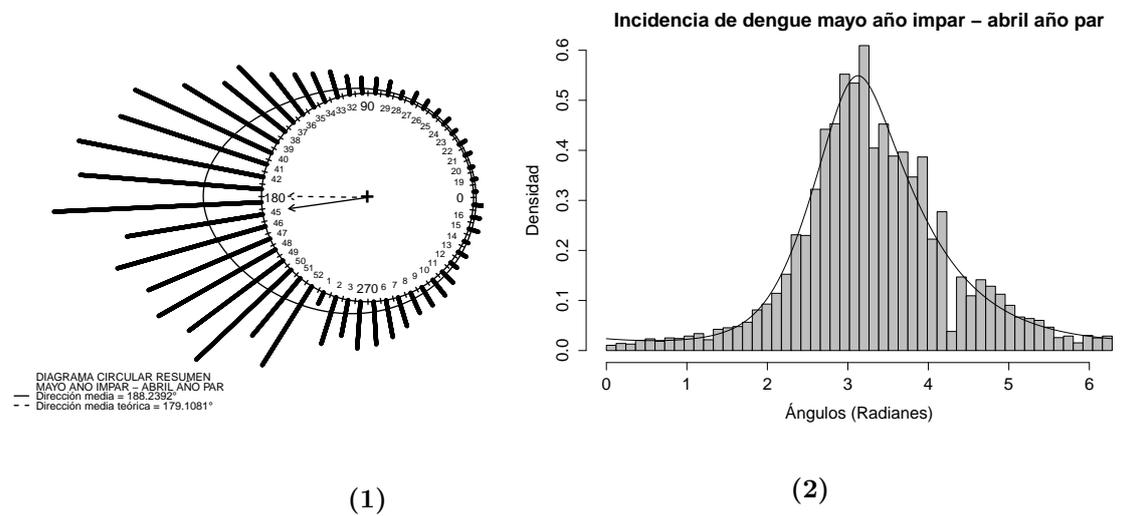


Figura 6.11: Representaciones polar, a la izquierda, y lineal, a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período resumen 1.

La Figura 6.111, muestra el diagrama circular que incluye la forma polar de la distribución, la dirección modal teórica y la dirección media muestral. Si se hablara

de un año impar normal, se observaría que la semana en la que se hubiese esperado se presente la mayor frecuencia de dengue sería la primera o la segunda anterior a la 45, semana alrededor de la cual se concentran las mayores frecuencias de casos de dengue. La Figura 6.112 muestra la representación lineal de la distribución, en la cual se observa hay una mayor masa de probabilidad después de la dirección modal teórica, en sentido antihorario.

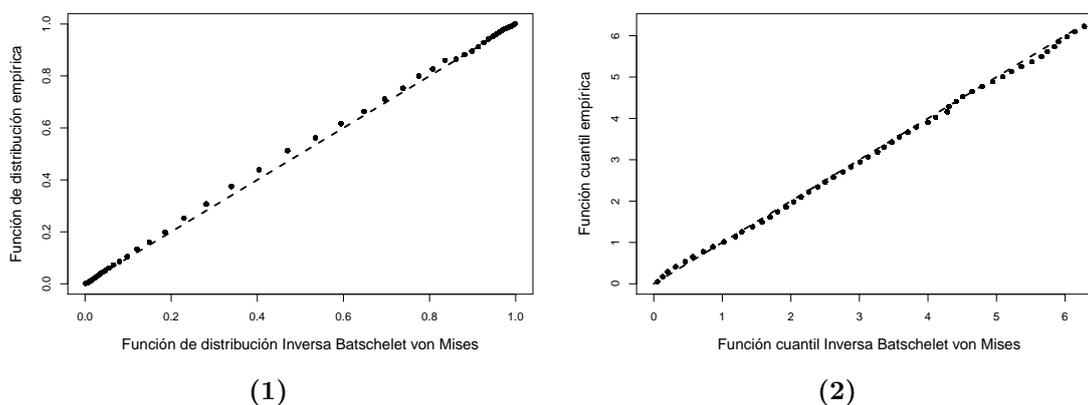


Figura 6.12: Gráficos probabilidad-probabilidad (PP plot), a la izquierda, y cuantil-cuantil (QQ plot), a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período resumen 1.

De la Figura 6.12, en general, se observa que las gráficas P-P y Q-Q presentan la mayoría de sus puntos muy cercanos a la línea de referencia, lo cual es señal de un ajuste razonable.

En la prueba de bondad de ajuste, basado en la de uniformidad de Rayleigh, el valor del estadístico de prueba es de 0.0036 y el p-valor arrojado es de 0.8377. Se concluye que para el comportamiento de los casos de dengue de este período es razonable el ajuste de una distribución Inversa Batschelet Completa, a cualquier nivel de significancia α frecuentemente utilizado.

6.4.2. Período Resumen 2: Mayo Año Par - Abril Año Impar

Análisis descriptivo

La Figura 6.131 y la Tabla 6.12, resumen los gráficos y estadísticos descriptivos circulares para la unión de los datos de los períodos 2, 4 y 6, es decir, mayo 2010 - abril 2011, mayo 2012 - abril 2013, y mayo 2014 - abril 2015. En este período resumen, se registraron en total 7734 casos de dengue en el estado. La semana con mayor cantidad de casos fue la 35 (transición de agosto a septiembre) del Año Par.

El diagrama circular en la Figura 6.131 y la Tabla 6.13, muestran hacia qué semanas apuntan las direcciones $\tilde{\theta}$, $\hat{\theta}$ y $\bar{\theta}$. Para la semana modal 37, la dirección moda es $\tilde{\theta} = 134.137^\circ$. La dirección $\hat{\theta} = 141.0411^\circ$ y la dirección $\bar{\theta} = 141.8827^\circ$, están muy cercanas, y de hecho apuntan hacia la misma semana, la 38 (3ª de septiembre), es decir, alrededor de esta semana se acumulan las grandes frecuencias y, además, en ella se alcanza la mitad de los casos reportados en total para este período.

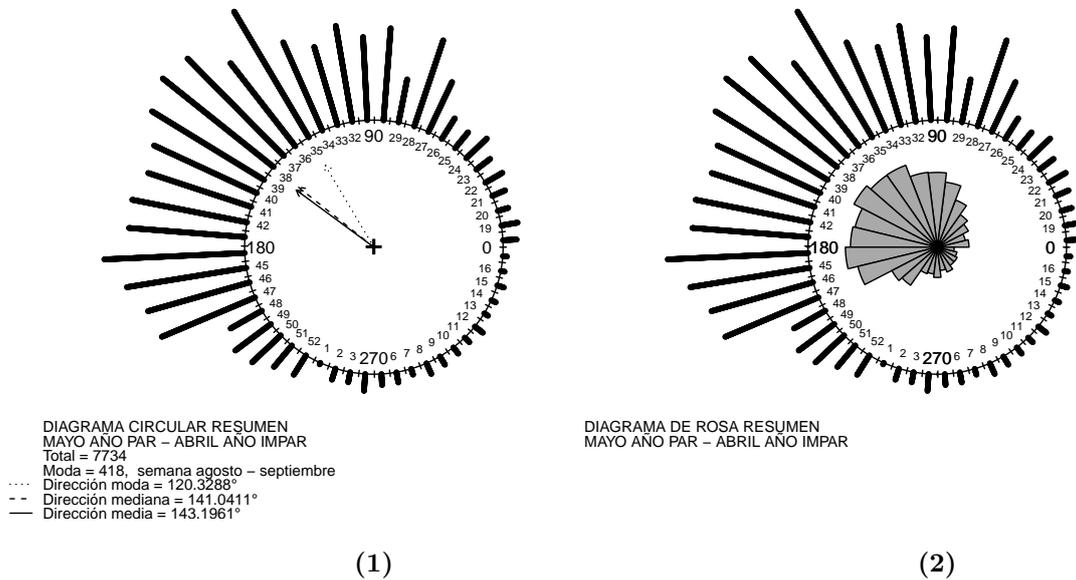


Figura 6.13: Diagramas circular a la izquierda y de rosa a la derecha para el período resumen 2.

Período Resumen 2: mayo año par - abril año impar	
Total	7734
Moda	418
$\check{\theta}$	120.3288°
$\tilde{\theta}$	141.0411°
$\bar{\theta}$	143.1961°
\bar{R}	0.5966812
\bar{b}_2	0.03208866
\hat{s}	0.1252792
\bar{a}_2	0.1801475

Tabla 6.12: Total, moda y medidas de localización, de dispersión y de sesgo para período resumen 2.

Estadístico de localización	Semana de ubicación	Mes
$\check{\theta}$	35	Agosto - Septiembre
$\tilde{\theta}$	38	Septiembre (3 ^a)
$\bar{\theta}$	38	Septiembre (3 ^a)

Tabla 6.13: Tiempo al que se dirigen los estadísticos descriptivos circulares de localización, para el período resumen 2.

La medida $\bar{R} = 0.5967$ ofrece una señal de una concentración de los casos de dengue. Si se dividiera un año par normal en cuatro partes, al observar el diagrama de rosa de la Figura 6.132, la concentración ocurre a principios del tercer cuarto hasta después de la mitad del último cuarto de ese año; más precisamente, a partir de la semana 28 (2^a de julio) hasta la semana 47 (3^a de noviembre) del año.

La circunferencia unitaria para este período resumen se ha cortado aproximadamente a la mitad del arco que abarca de la semana 1 hasta la semana 17 del año impar (recorriendo la circunferencia en sentido antihorario), el cual presenta las más bajas frecuencias. Es decir, se ha cortado en el punto con dirección 301° (5.2534 radianes). Con base en este corte, se construye el histograma lineal de la Figura 6.14.

La medida $\bar{b}_2 = 0.0320$ es cercana a cero, pero su contraparte estandarizada $\hat{s} = 0.1380$ no lo es tanto. Por ende, no hay evidencia muestral contundente de un comportamiento reflexivamente simétrico. Sin embargo, estas medidas evidencian un ligero sesgo en sentido horario, con respecto de la dirección media (ver diagrama de rosa de la Figura 6.132).

El diagrama de rosa, junto con el histograma lineal de la Figura 6.14, muestran que, en general, a partir de la semana 27 (1^a de julio) empieza un franco aumento de los casos, y el tramo de altas frecuencias se mantiene hasta las semana 47 (3^a de noviembre), después de la cual se registra un descenso abrupto de los casos reportados para este período resumen.

La medida $\bar{a}_2 = 0.1801$ señala un picudez no extrema en la forma polar de la distribución de los datos (ver diagrama circular de la Figura 6.131), pero es menor la reportada en el período resumen 1.

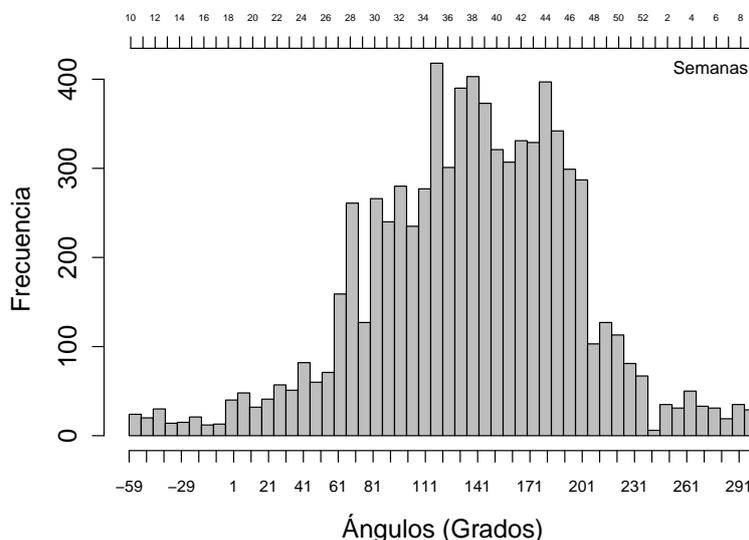


Figura 6.14: Histograma lineal de frecuencias de dengue, para el período resumen 2, con punto de corte en la dirección 301° (5.2534 radianes). A los valores negativos del principio hay que sumarles 360° para obtener valores en el rango de $(0, 360^\circ)$.

Al aplicar la prueba de uniformidad de Rayleigh a estos datos, el valor del estadístico de prueba es de 0.6 y arroja un p-valor muy cercano a cero. Por otra parte, para la prueba de simetría reflexiva de Pewsey, el valor del estadístico de prueba es de 4.4471 y el p-valor es de $8.70234e-06$. Con p-valores muy cercanos a cero, las hipótesis nulas de uniformidad y simetría reflexiva se rechazan a cualquier nivel de significancia α frecuentemente utilizado. Por lo anterior, se espera que un modelo que considere sesgo ofrezca el mejor ajuste a los datos.

Modelación circular

La Tabla A.3, muestra que el mejor modelo que se ajusta a los casos de dengue contenidos en este resumen 2, es el de una distribución Inversa Batschelet Completa, es decir, con todos sus parámetros distintos de cero.

Período Resumen 2: mayo año par - abril año impar	
Mejor ajuste: Inversa Batschelet Completo	
Parámetro	Estimado
ξ	2.223296
κ	1.57123
ν	-0.2096247
λ	-0.1187337

Tabla 6.14: Ajuste de un modelo Inverso Batschelet Completo, con base von Mises, para el período resumen 2.

La Tabla 6.14, muestra los estimados de los parámetros que controlan la distribución Inversa Batchelet, obtenidos vía máxima verosimilitud. El valor de $\hat{\nu} = -0.2096$ considera un sesgo en sentido horario. El valor $\hat{\lambda} = -0.1187$ considera una distribución con cima plana en su presentación polar.

La dirección modal teórica de la distribución, estimada por el modelo a través de la expresión $\xi - 2\nu \pmod{2\pi}$, es de 2.6425, la cual apunta hacia las semanas 39 y 40, que son la cuarta semana de septiembre y la primera de octubre, respectivamente.

La Figura 6.111 muestra el diagrama circular que incluye la forma polar de la distribución, la dirección modal teórica y la dirección media muestral. Si se hablara de un año impar normal, se observaría que la semana en la que se hubiese esperado se presente la mayor frecuencia de dengue sería la primera o la segunda después de la 38, semana alrededor de la cual se concentran los mayores frecuencias de casos de dengue. La Figura 6.152 muestra la forma lineal de la distribución, y evidencia que hay una mayor masa de probabilidad antes de la dirección modal teórica, en sentido horario.

De la Figura 6.16, en general, se observa que las gráficas P-P y Q-Q presentan la mayoría de sus puntos muy cercanos a la línea de referencia, lo cual es señal de un ajuste razonable.

En la prueba de bondad de ajuste, basado en la de Rayleigh, el valor del estadístico de prueba es de 0.0117 y el p-valor arrojado es de 0.3488. Se concluye que para el

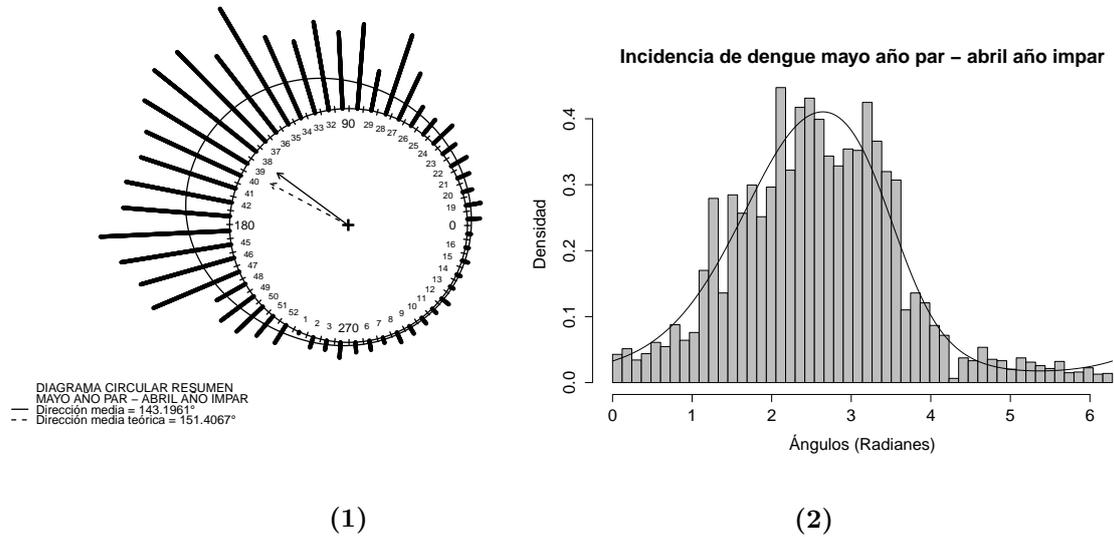


Figura 6.15: Representaciones polar, a la izquierda, y lineal, a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período resumen 2. El histograma lineal comienza en la dirección cero.

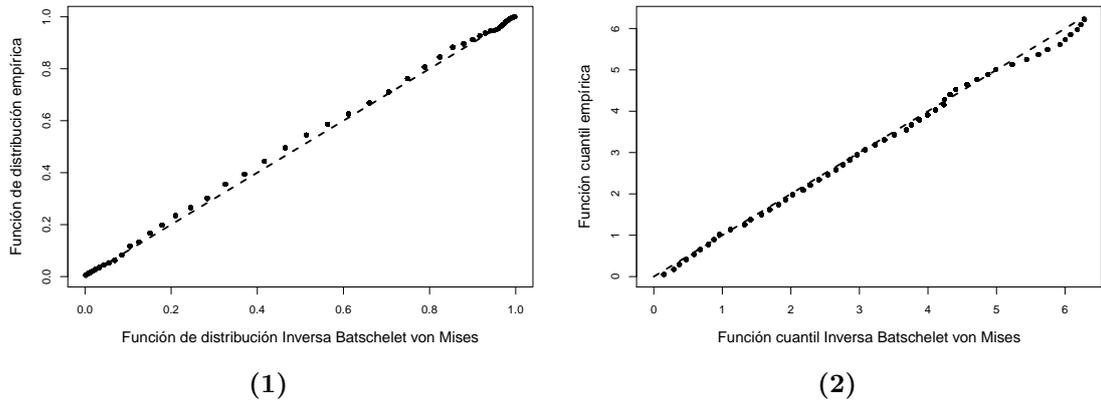


Figura 6.16: Gráficos probabilidad-probabilidad (PP plot), a la izquierda, y cuantil-cuantil (QQ plot), a la derecha, de la distribución Inversa Batschelet Completa ajustada para el período resumen 2.

comportamiento de los casos de dengue para este período es razonable el ajuste de una distribución Inversa Batschelet Completa, a cualquier nivel de significancia α frecuentemente usado.

6.5. Comparación entre los períodos de estudio y los períodos generales

Las Figuras conjuntas B.7 y B.8 del Apéndice B ofrecen una comparación visual entre los períodos resumen 1 y resumen 2.

En el diagrama circular del período resumen 1 de la Figura B.71, podemos observar que la dirección media muestral está relativamente cerca de 180° , comportamiento que hereda de los períodos impares. En contraparte, para el resumen 2 en el diagrama circular de la Figura B.72 la dirección media muestral está relativamente lejos de 180° , como sucede en los períodos pares.

El diagrama de rosa del período resumen 1 de la Figura B.73 muestra un sesgo en sentido antihorario de la distribución de los casos de dengue, como sucede en los períodos 1 y 3. En la Figura B.74 para el resumen 2, se exhibe un sesgo horario heredado de los períodos 2 y 4.

El histograma lineal del resumen 1 en la Figura B.75 muestra un comportamiento de los casos de dengue heredado de los períodos 1 y 3, que consiste en que el que se exhibe un franco crecimiento de las frecuencias, y después de alcanzar el máximo, inmediatamente después comienza un franco decrecimiento de las frecuencias. En contraparte, el histograma lineal del resumen 2 en la Figura B.76 muestra un comportamiento heredado de los períodos 2 y 4, que consiste en el franco crecimiento de las frecuencias hasta alcanzar el máximo, luego sigue un período corto en que las frecuencias altas se mantienen, para después dar paso a una caída casi abrupta de las frecuencias.

Se han ajustado modelos Batschelet Completos en estos períodos resumen. La representación polar de la distribución del resumen 1 en la Figura B.82 exhibe una punta baja y que la dirección modal teórica está antes de la dirección media, en sentido horario, como en los períodos impares 1 y 3. En contraste, la representación polar del resumen 2 en la Figura B.82 exhibe una cima plana como sucede en los períodos pares, y que la dirección modal teórica está después de la dirección media, en sentido antihorario, como sucede en los períodos 2 y 4.

La forma lineal de la distribución para el resumen 1 en la Figura B.83 muestra que el modelo Inverso Batschelet considera un sesgo antihorario como los períodos impares 1 y 3, mientras que en la Figura B.84 para el resumen 2 se considera un sesgo horario como en los períodos 2 y 4.

La Tabla A.4 del apéndice A muestra los estimados de los estadísticos descriptivos circulares del resumen 1 y el 2, junto con sus intervalos del 95% de confianza. La Figura 6.17 muestra la comparación visual de los estimados de los resúmenes

y los que se han obtenido previamente en los períodos de estudio. La Tabla 6.15 muestra que los estimados de los estadísticos calculados para el período resumen 1 son, en general, una especie de promedio de los valores ya estimados para un estadístico particular en los períodos impares; análogamente, sucede algo similar con el resumen 2 respecto a los períodos pares.

La Tabla A.5 del Apéndice A muestra los estimados de los parámetros cuando a los datos de los resumen 1 y 2 se les ajusta un modelo Inverso Batschelet Completo, con sus intervalos del 95 % de confianza. La Figura 6.18 ofrece una comparación visual de los parámetros estimados de los resúmenes y los que se han obtenido previamente en los períodos de estudio.

La Tabla A.6 muestra los valores estimados de los parámetros que han dado el mejor ajuste de un modelo Inverso Batschelet (no necesariamente completo). La Tabla 6.16 muestra que los valores estimados de los parámetros Batschelet Inverso calculados para el período resumen 1, en general, son una especie de promedio de los valores distintos de cero estimados para un parámetro en particular en los períodos impares; análogamente, sucede algo similar con el resumen 2 respecto a los períodos pares.

Descriptivos circulares	Estimados para el período resumen 1	Promedio de los descriptivos en los períodos impares	Diferencia
$\bar{\theta}$	3.2854	3.1712	0.1142
\bar{R}	0.6494	0.6308	0.0186
\bar{b}_2	-0.0681	-0.0746	0.0065
\bar{a}_2	0.3022	0.2819	0.0203
Descriptivos circulares	Estimados para el período resumen 2	Promedio de los descriptivos en los períodos pares	Diferencia
$\bar{\theta}$	2.4992	2.5031	0.0038
\bar{R}	0.5966	0.5626	0.0340
\bar{b}_2	0.0321	0.0297	0.0024
\bar{a}_2	0.1801	0.1411	0.0390

Tabla 6.15: Arriba, se muestra la comparación entre un estadístico descriptivo circular estimado en el período resumen 1 y el promedio de los valores estimados en los períodos impares para ese mismo estadístico. Abajo, se hace la misma comparación pero ahora en el caso del resumen 2 respecto a los períodos pares.

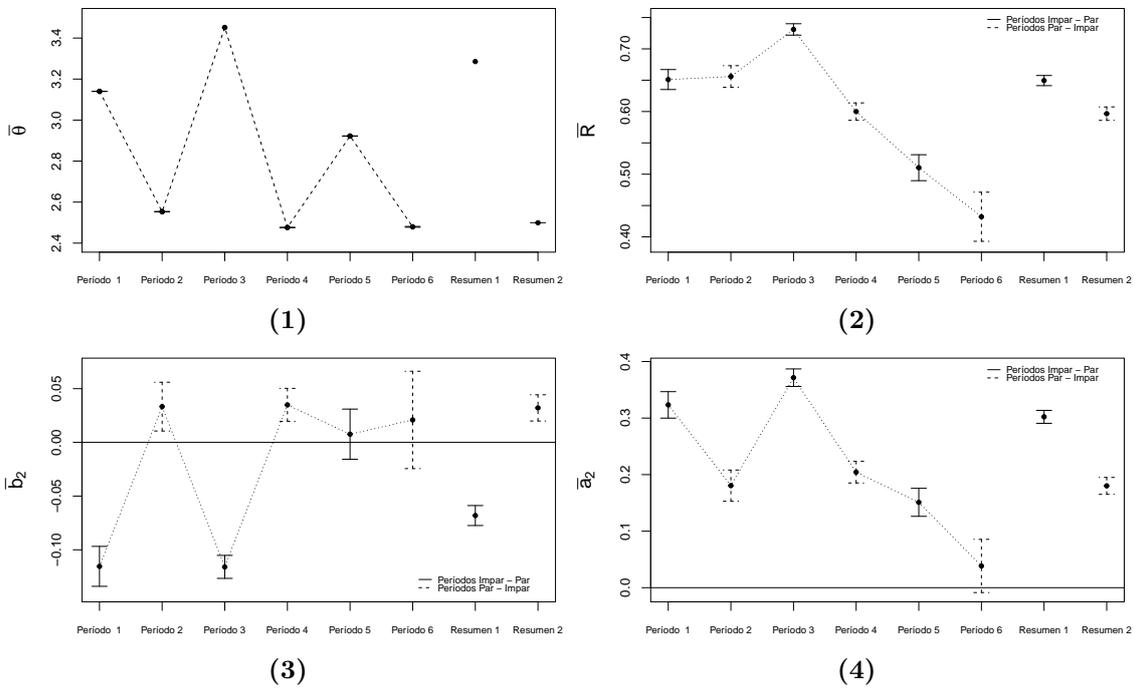


Figura 6.17: Comparación gráfica de los estadísticos descriptivos circulares de los seis períodos de estudio y de los dos generales (resumen 1 y 2): (1) para la dirección media muestral $\bar{\theta}$, (2) para la longitud media muestral \bar{R} , (3) para la medida de sesgo \bar{b}_2 , (4) para la medida de curtosis \bar{a}_2 .

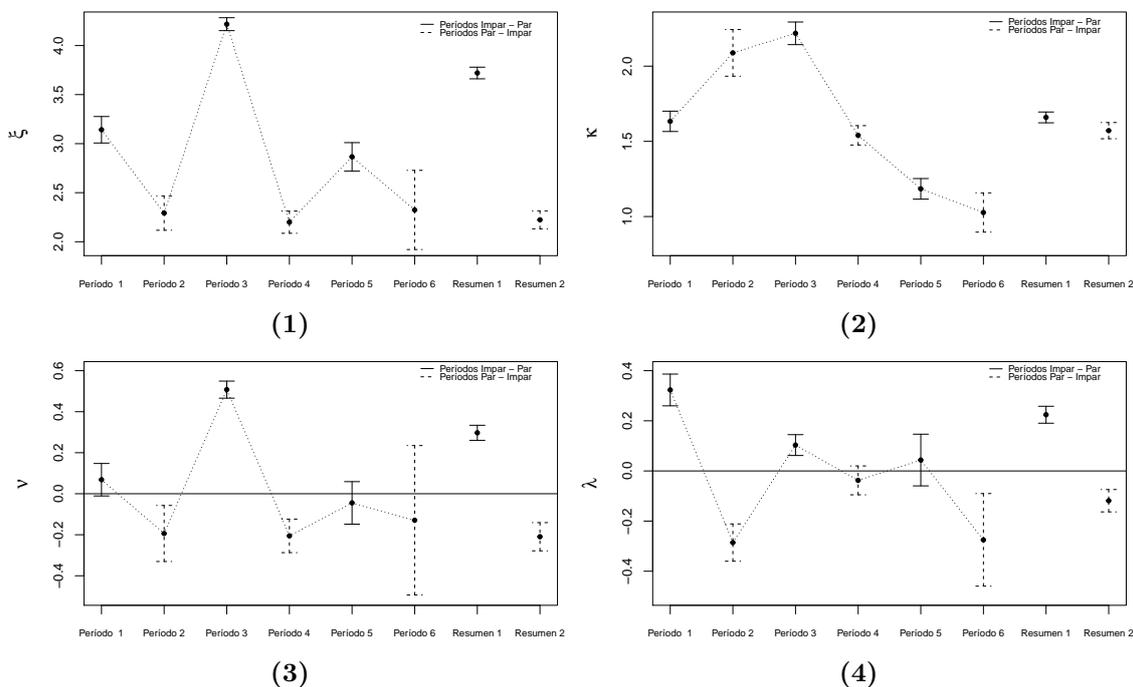


Figura 6.18: Comparación de los valores obtenidos en los seis períodos de estudio y de los dos generales cuando a cada uno se le ajusta un modelo Inversa Batschelet Completo: (1) para el parámetro de localización ξ , (2) para el parámetro de concentración κ , (3) para el parámetro de sesgo ν , (4) para el parámetro de curtosis λ .

Parámetros Inversa Batschelet	Estimados para el período resumen 1	Promedio de los estimados de los períodos impares	Diferencia
ξ	3.7202	3.4272	0.2930
κ	1.6592	1.6814	0.0222
ν	0.2971	0.2879	0.0092
λ	0.2242	0.2131	0.0110
Parámetros Inversa Batschelet	Estimados para el período resumen 2	Promedio de los estimados de los períodos pares	Diferencia
ξ	2.2233	2.3186	0.0953
κ	1.5712	1.5468	0.0245
ν	-0.2096	-0.1988	0.0109
λ	-0.1187	-0.2874	0.1687

Tabla 6.16: Arriba, se muestra la comparación entre un parámetro estimado Inverso Batchelet en el período resumen 1 y el promedio de los valores estimados en los períodos impares para ese mismo parámetro. Abajo, se hace la misma comparación pero ahora en el caso del resumen 2 respecto a los períodos pares.

7

Discusión y conclusiones

En este último Capítulo, se discute sobre las características que ofrece el modelar la incidencia de dengue en un período mayo-abril anual utilizando estadística circular, y se comparan los métodos utilizados aquí con los utilizados en otro trabajo de investigación que también trata la incidencia de dengue con estadística circular. Se discute además sobre las características de los números semanales de dengue que se encuentran en la base de datos en bruto ofrecida por la Dirección General de Epidemiología. Se hace un ejercicio de interpretación de los resultados obtenidos tanto en el Capítulo 5 y como en el Capítulo 6. Se analiza el comportamiento del principal factor de incidencia, la precipitación pluvial, y el posible efecto que tiene, a su vez, en el comportamiento de los casos de dengue observado en los períodos de estudio. Se presentan las conclusiones generales tras analizar los resultados a los que se han llegado a lo largo de este trabajo de investigación. Por último, se consideran algunas labores a llevarse a cabo en el futuro, motivadas por los resultados del presente trabajo.

7.1. Sobre la utilización de la estadística circular como herramienta de estudio

Como se ha dicho en la *Introducción*, los datos de incidencia de dengue ya han sido trabajados con técnicas comúnmente usadas como lo son las series de tiempo, y otras como los árboles de regresión. La estadística circular ofrece una alternativa para modelar que toma en cuenta la periodicidad de los datos circulares en un lapso de tiempo. Por ende, si tenemos datos de un evento periódico que se ha observado en diferentes ciclos de un período conocido, toda la información recabada en ellos

puede ser utilizada para ajustar un modelo global que describa el comportamiento de estos datos en el período de interés, por ejemplo, uno mayo-abril anual.

Los datos semanales de los casos de dengue en un año pueden ser modelados utilizando series de tiempo. En este caso, la información que nos ofrecen los números de casos de dengue semanales es utilizada para pronosticar el número de casos que se esperan para una semana próxima inmediata. Sin embargo, no se pueden hacer conclusiones globales acerca de todo un período de estudio o partes de él, por ejemplo, no se podría decir qué cantidad de casos de dengue se esperan para un lapso de tiempo constituido por un conjunto de semanas, digamos, un mes. Lo anterior es debido a que estos métodos son diseñados para obtener pronósticos extrapolares o intrapolares, siempre para *una observación* precedida, precisamente, de una serie lineal de observaciones que ya han sido registradas.

Si bien, al ajustar un modelo circular no se tendrán conclusiones referentes a números de casos semanales esperados, sí se tendrán probabilidades de incidencia de dengue, que pueden ser transformadas e interpretadas como los porcentajes de incidencia de dengue que se esperan del total verdadero, y que puede ser obtenidos para una semana particular y, por ende, para un mes particular de período de estudio. Es decir, de acuerdo al historial de los datos en un tiempo observado, podríamos tener conclusiones referentes a todas las semanas de un período de estudio, y no sólo referentes a una en particular. En este trabajo, los porcentajes mensuales de incidencia de dengue se han calculado para cada período de estudio, y los resultados se muestran en la Tabla A.7 del Apéndice A.

En un período de estudio mayo-abril anual particular se han hecho un análisis descriptivo y una modelación circular. Para tal efecto se dividió el círculo en 365 partes. Sin embargo, la naturaleza de los datos de incidencia semanal es agrupado, es decir, se tiene el número de casos de dengue que se obtienen en una sola semana. En el caso de datos agrupados, se tiene correcciones a considerar en el análisis descriptivo, como los que se muestra en la Sección 2.5 para la medida de concentración \bar{R} . En cuanto a la modelación, hay que tener algunas consideraciones para el cálculo adecuado de los parámetros de máxima verosimilitud para datos agrupados; básicamente, se tiene que hacer ajustes en las funciones computacionales que hacen estos cálculos para un modelo específico, y Pewsey *et al.* (2013) ofrecen un ejemplo de este trabajo con respecto al cálculo de los estimadores de máxima verosimilitud de los parámetros de la familia de distribuciones simétricas Jones-Pewsey. Sin embargo, las pasadas correcciones y consideraciones no se han tomado en cuenta en este trabajo debido a que la cantidad de datos que se tienen en un período de estudio en particular, es decir, la cantidad de casos registrados, es grande, además de que esta cantidad está repartida en una cantidad de arcos de clase también grande, pues son 52 los intervalos en los que se ha dividido la

circunferencia unitaria. Todo esto se conjuga para que las cantidades resultantes a las que se llegan en este trabajo sean muy exactas, es decir, cercanas a las cantidades que se pudiesen haber obtenido si los datos se hubiesen registrado con más precisión, por ejemplo, en días y no en semanas.

Aunque, también es cierto que debido a la naturaleza agrupada de los datos se han llegado a algunas conclusiones que carecen de sentido en el contexto en que se ubica este trabajo de investigación; por ejemplo, en la secciones 5.8, 6.4.1 y 6.4.2, al calcular la dirección modal teórica estimada por el modelo Inversa Batschelet con base von Mises a través de la expresión $\xi - 2\nu \pmod{2\pi}$, sucede que esta dirección apunta precisamente hacia el espacio que separa al punto representante del último día de una primera semana, y al punto representante del primer día de la semana siguiente, es decir, no apunta hacia una semana específica en sí. Por ende, en estas secciones se han considerado dos posibles semanas en las que el modelo estima pudiera presentarse la mayor frecuencia de dengue, semanas que, por supuesto, son contiguas.

Son escasos los trabajos que se han realizado en los que se aplique estadística circular para estudiar o dar solución a un problema perteneciente a otras áreas de estudio diferentes de las matemáticas. Debido a lo anterior, no es posible comparar el presente trabajo de investigación con muchos y diversos trabajos, y más aún, con trabajos que atiendan problemas de naturaleza similar al que se trata en este presente trabajo. Sin embargo, existe uno que, además de analizar la incidencia de dengue para un lugar y tiempo determinados, ha inspirado y se ha tomado como base para este presente trabajo de investigación; es el trabajo de Jaimes Jaimes y Pineda Garavito (2012). A continuación, discutiremos sobre las similitudes, pero sobretodo las diferencias, entre este presente trabajo y el de Jaimes Jaimes y Pineda Garavito (2012), a fin de compararlos y exhibir, por ejemplo, el incremento que se ha tenido, de un tiempo a la fecha, de las novedosas herramientas que ofrece la estadística circular para el tratamiento de datos en la circunferencia unitaria.

En el presente trabajo de investigación, la idea de utilizar estadística circular en el estudio de la incidencia de dengue, se ha inspirado en el trabajo de tesis de Jaimes Jaimes y Pineda Garavito (2012), en el cual se exhiben dos ejemplos de aplicación de datos circulares. En uno de dichos ejemplos, se aplica estadística circular en el estudio de la incidencia de dengue en el área metropolitana de Bucaramanga, ciudad capital del departamento de Santander, Colombia. En el presente trabajo, las ideas de estudiar la incidencia semanal como datos circulares y el uso de los días de la mitad de la semana como los representantes semanales, son tomadas del trabajo de Jaimes Jaimes y Pineda Garavito (2012).

En Jaimes Jaimes y Pineda Garavito (2012), la información se refiere a un “área

metropolitana” mientras que en el presente trabajo la información se refiere a un estado de la república mexicana, de modo tal que aquí la base de datos es más grande, y por lo tanto, las conclusiones que se derivan se esperan sean más razonables y precisas. Jaimes Jaimes y Pineda Garavito (2012) hacen un análisis descriptivo y un ajuste de modelo circular por años, mientras que en este presente trabajo estas tareas se han hecho para períodos de estudio mayo-abril anuales. La base de datos en aquél trabajo no exhibe ciclos en el comportamiento de la incidencia de dengue al observarlo por años, es decir, no parece haber lapsos de tiempo en los que se esperan números de casos de dengue altos o bajos; el comportamiento de la incidencia de dengue es irregular en cada año. En contraste, en el presente trabajo, la base de datos exhibe la presencia de ciclos claramente establecidos, en los que hay lapsos de tiempo en los que se espera observar números de casos de dengue altos o bajos.

El análisis descriptivo de los datos hecho por Jaimes Jaimes y Pineda Garavito (2012) incluye el uso de las direcciones media y modal muestrales, así como de el rango y la varianza circulares, además del uso de los diagramas circular y de rosa. En cambio, en este presente trabajo, el análisis descriptivo es más amplio e incluye el uso de las direcciones media, mediana y modal muestrales, las medidas descriptivas de sesgo y curtosis circulares, el uso de los diagramas circular y de rosa, el uso de los histogramas lineales, además del uso de pruebas de clasificación de modelos diseñados para datos circulares, como los de uniformidad de Rayleigh y de simetría reflexiva de Pewsey.

En cuanto a la modelación circular, Jaimes Jaimes y Pineda Garavito (2012) se limitan a discutir qué tan razonable es el ajuste de una distribución circular von Mises a los datos de incidencia, y la bondad de ajuste sólo es evaluada utilizando los gráficos probabilidad-probabilidad (P-P plots). En contraste, en este presente trabajo, se han utilizado cinco modelos circulares para encontrar el que mejor se ajuste a los datos de incidencia; el mejor modelo se he encontrado utilizando los criterios de información AIC y BIC, y una vez seleccionado, el buen ajuste es evaluado utilizando una prueba de bondad de ajuste para datos circulares (basada en la prueba de Rayleigh), además del uso de los gráficos cuantil-cuantil (Q-Q plots) y probabilidad-probabilidad (P-P plots).

En el trabajo de Jaimes Jaimes y Pineda Garavito (2012), debido el comportamiento irregular de los casos de dengue en cada año, al final sólo se exhibe en qué meses del año se han presentado los mayores números de casos de dengue, con la ayuda de un gran círculo unitario que engloba la información que se tiene en los años que se han considerado en su estudio. En el presente trabajo, la presencia de casos de incidencia en prácticamente todo el lapso de tiempo que abarcan los períodos mayo-abril anuales permite obtener porcentajes de incidencia de dengue,

que pueden ser mensuales o semanales; aquí, el comportamiento de la incidencia de dengue da como resultado dos grandes círculos unitarios que resumen la información contenida en los llamados períodos impares y pares. Este aparente comportamiento dual en la incidencia de dengue en Yucatán permite, con ayuda del trabajo de modelación circular realizado en este presente estudio, tener al menos una idea sobre lo que se espera observar en los períodos de estudio mayo-abril anuales venideros.

7.2. Sobre los períodos de estudio

Son seis los períodos mayo - abril anuales que se han considerado en este estudio. Se han encontrado características de similitud cualitativas y distribucionales entre los períodos denominados impares, y entre los períodos denominados pares, y ya han sido comentados en el Capítulo 6.

Es importante señalar que estas similitudes se dan, sin importar los totales de casos de dengue registrados en cada período, los cuales varían según se muestra en la Tabla 7.1. De manera global, la Figura 7.1 muestra que el valor registrado para un período impar es relativamente más grande que el período inmediato posterior, el cual es necesariamente par.

Período	Total
1	3594
2	2071
3	7044
4	4833
5	2822
6	830

Tabla 7.1: Total de casos registrados en los períodos de estudio.

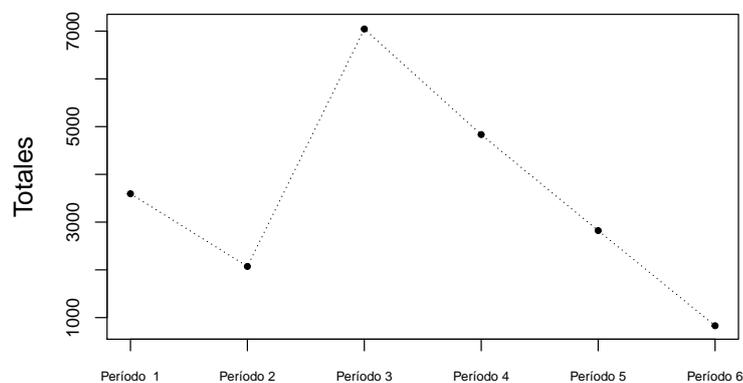


Figura 7.1: Total de casos registrados en los períodos de estudio.

La Figura 7.1 muestra un repunte considerable del total de casos registrados de dengue en el período 3 respecto a los períodos 1 y 2, para luego dar paso un franco decrecimiento de los totales que se ha mantenido hasta el período 6. Del período 3 al período 6 ha habido una reducción de casi el 88% de los casos registrados.

Describiendo los aumentos y decrementos de los totales de incidencia, paso a paso, se puede decir que del período 1 al 2 hubo una reducción de casi el 42% de los

casos registrados; del período 2 al 3 hubo un aumento de casi el 240 %; del período 3 al 4 hubo una reducción de casi el 31 %; del período 4 al 5 hubo una reducción de casi el 42 %; y del período 5 al 6 hubo una reducción de casi el 71 % de los casos registrados.

Hay marcadas similitudes entre los períodos 1 y 3, y, sin embargo, hubo un aumento del casi el doble de los casos registrados en el 3 respecto al 1. Algo similar pasó en el período 4 respecto al 2, y es que en el 4 se superó el doble de los casos registrados respecto al 2.

Aún hay similitudes encontradas entre los períodos 1 y 3, y el 5, aunque del 1 al 5 hubo una reducción del casi el 21 % de los casos de dengue, y del 3 al 5 hubo una reducción de casi el 60 %; también hay similitudes encontradas entre los períodos 2 y 4, y el 6, aunque del 2 al 6 hubo una reducción de casi el 60 %, y del 4 al 6 hubo una reducción de casi el 83 %.

De hecho, pareciera que la estructura que presentan las distribuciones muestrales de los períodos 1 y 3 (ver histogramas lineales de las Figuras B.31 y B.33) está “incluida a escala” en la estructura distribucional muestral del período 5 (ver histograma lineal de la Figura B.35). Las frecuencias semanales de los histogramas lineales de los períodos 1 y 3 parecieran reescalarsen de tal forma que en el histograma del período 5 se de un efecto visual de que las frecuencias del principio, hasta antes de la moda muestral, se eleven a tal punto de dar un impresión de simetría. En realidad pasa todo lo contrario, es la caída de los casos semanales, en las denominadas frecuencias grandes, principalmente las de después de la moda muestral, las que se reescalan a tal punto de quedar en armonía con las frecuencias del principio y dar la impresión de simetría. Algo similar sucede en la estructura que presentan las distribuciones muestrales de los períodos 2 y 4 (ver histogramas lineales de las Figuras B.32 y B.34) respecto a la distribución muestral del período 6 (ver histograma lineal de la Figura B.36); un reescalamiento, sobre todo en las grandes frecuencias de los histogramas lineales 2 y 4, da la impresión visual de un rasgo de simetría en las frecuencias semanales del histograma del período 5.

Sobre los períodos impares

Las similitudes son marcadas en los primeros períodos impares 1 y 3. Entre las similitudes cualitativas se encuentran el que posean sesgo antihorario respecto de la dirección media, que sus direcciones medias muestrales estén relativamente cerca de 180° , y posean valores de concentración considerables. Entre las distribucionales se encuentran que los valores máximos de los modelos circulares ajustados están cerca de 0.6 (0.5897 para el período 1, 0.6040 para el período 3), que posean picudez

no extrema en la forma polar de la distribución, y que las direcciones modales que estiman sus modelos (y que son los que maximizan la densidad) se encuentran muy cerca de 180° .

El período 5 ya no muestra un sesgo antihorario, sino un comportamiento reflexivamente simétrico alrededor de la dirección media, que también está cerca de 180° , y el modelo ajustado indica que esta misma dirección es también la dirección modal de la distribución, además de que se exhibe un comportamiento de dispersión más de concentración. Aquí, la representación polar de la densidad ajustada tiene cima plana (sin pico visualmente evidente), y el máximo ya no se está cerca de 0.6 sino por debajo de 0.4 (en realidad es 0.3778).

La Tabla A.7 del Apéndice A, muestra los porcentajes mensuales del total real de incidencia que se hubiese esperado registrar, de manera honesta, en un período de estudio correspondiente. Estos porcentajes han sido calculados usando los modelos circulares de probabilidad ajustados en cada período. En esta Tabla, los meses arriba del 5 % del total de incidencia se han considerado como interesantes y se han coloreado de amarillo, pues se ha observado que al tener en cuenta sólo a estos meses, en conjunto, ya se alcanza arriba del 80 % de la incidencia de dengue en un período en particular.

En la Tabla A.7 llama la atención que en cada uno de los períodos impares exista un mes “punta”, es decir, un mes con el mayor porcentaje del total real esperado en el período correspondiente. Estos meses se han coloreado de verde.

En el caso del período 1, el rango de meses arriba del 5 % fue de 6, y abarcó de agosto del 2009 a enero del 2010. Juntos, en estos meses se esperaba registrar un 88.67 % del total. Sólo en el mes de octubre, el mes punta, se esperaba un 28.94 % del total.

En el caso del período 3, en el rango de meses arriba del 5 % fue de 5, y abarcó de septiembre del 2011 a enero del 2012. Juntos, en estos meses se esperaba registrar un 90.21 % del total. El mes punta fue noviembre, en el que se esperaba un 28.45 % del total.

En el caso del período 5, el rango de meses arriba del 5 % fue de 7, y abarcó de julio del 2013 a enero del 2014, y en ellos se esperaba un 87.37 % del total. El mes punta fue octubre con un 19.88 % del total.

La Tabla 7.2 muestra un resumen de la información acerca de los porcentajes mensuales para los períodos impares. Los meses punta de los períodos 1 y 3 fueron diferentes, aunque cercanos; fueron octubre y noviembre, respectivamente, aunque los porcentajes de incidencia en ellos fueron similares, cercanas al 30 %. En el período 5, el mes punta volvió a ser octubre, pero aquí hubo una pérdida de

Período	Mes punta	Acumulado de mes punta	Meses arriba del 5 %	Acumulado de los meses arriba del 5 %
1	octubre	28.94 %	agosto - enero (6)	88.67 %
3	noviembre	28.45 %	septiembre - enero (5)	90.21 %
5	octubre	19.88 %	julio - enero (7)	87.37 %
Resumen 1	octubre	25.71 %	agosto - enero (6)	88.53 %

Tabla 7.2: Mes punta y meses arriba del 5 % de incidencia de dengue, para un período impar.

porcentaje respecto a lo observado en los meses punta anteriores, y sólo se esperaba casi un 20 % de incidencia.

En la pasada Tabla, también se puede observar que del período 1 al 3 hubo una especie de compactación de la distribución, la cual para este último período acumula casi el 90 % del total en sólo cinco meses. Así, el comportamiento visto en el diagrama circular de la Figura B.13 para el período 3, es en realidad una compactación de las grandes frecuencias hacia el final del año 2011, en lugar de una rotación, respecto al diagrama circular de la Figura B.11 para el período 1. El período 5 presentó una ampliación del rango de meses arriba del 5 %, fueron 7; sin embargo, también se puede notar que en ese entonces se requirieron de más meses para alcanzar un acumulado cercano a lo que se tenía para 6 meses en el período 1, o un tanto cercano a lo que se tenía para 5 meses en el período 3.

La Tabla A.7 también exhibe los porcentajes mensuales para el período resumen 1: Mayo Año Impar - Abril Año Par. Este período resumen hereda la existencia de un mes punta visto en los períodos impares. La Tabla 7.2 muestra que de acuerdo al historial global del comportamiento de los casos de dengue visto en estos tres períodos, el mes punta ha sido octubre con un porcentaje esperado del 25.71 % del total (el cual es una especie de promedio de los valores obtenidos en los meses punta de los períodos impares.). También, de este período resumen, el historial global de los casos de dengue sólo cuenta como meses significativos a los que están en el rango de agosto de un año par a enero de un año impar, y juntos acumulan un porcentaje esperado del 88.53 % del total.

Además, en la Tabla 7.2, es interesante observar que el último mes arriba del 5 % de incidencia en los períodos impares es enero, inclusive esta característica es heredada por el período resumen 1.

En la Tabla 7.3 se observa que las semanas a las que apuntan las medidas de localización, en general, pueden ser englobadas en el rango de la semana 42 (3^a de octubre) a la semana 46 (2^a de noviembre), es decir, de la segunda mitad de octubre hasta la segunda mitad de noviembre. Para los períodos impares 1, 3 y 5, en general, se puede ver que la semana en la que se alcanza la mitad del total de los

Período	Localización					
	Dirección media muestral $\hat{\theta}$		Dirección mediana muestral $\tilde{\theta}$		Dirección modal teórica	
	Semana	Mes	Semana	Mes	Semana	Mes
1	44	Octubre-Noviembre	43	Octubre (4 ^a)	42 - 43	Octubre (3 ^a -4 ^a)
3	46	Noviembre (2 ^a)	46	Noviembre (2 ^a)	44	Octubre-Noviembre
5	42	Octubre (3 ^a)	42	Octubre (3 ^a)	42	Octubre (3 ^a)
Resumen 1	45	Noviembre (1 ^a)	44	Octubre-Noviembre	43 - 44	Octubre (4 ^a) - Octubre-Noviembre

Tabla 7.3: Semanas a las que apuntan las direcciones media muestral, mediana muestral, modal teórica y su ubicación en los meses de un año normal.

casos registrados en un período en particular, es decir, las semana a la que apunta las dirección mediana muestral, será igual o la inmediata anterior a la semana alrededor de la cual se concentran las grandes frecuencias de dengue, es decir, la semana a la cual apunta la dirección media muestral. Algo parecido sucede con la semana modal con respecto a la semana media; la semana en la que se espera se presente la mayor frecuencia de incidencia de dengue, es decir, la semana a la que apunta la dirección modal estimada por el modelo circular ajustado, será igual o a lo más la segunda anterior a la semana a la que apunta la dirección media muestral.

En la Tabla 7.3, también se tiene que, de acuerdo a la historial global del comportamiento de los datos en estos tres períodos, el resumen 1 estima que alrededor de la semana 45, la primera de noviembre, se han concentrado las mayores frecuencias de dengue. También, se tiene que se alcanzaron la mitad del total de los casos en la semana 44, la semana transición entre octubre y noviembre; y, que la mayor frecuencia de dengue se habrá dado entre la semana 43, la cuarta de octubre, y la semana 44.

Período	Distribución (en sentido antihorario)			
	De la antimoda a la moda		Del cero a la moda	
	Probabilidad	Complemento	Probabilidad	Complemento
1	0.4864	0.5136	0.4833	0.5167
3	0.3908	0.6092	0.3917	0.6083
5	0.5	0.5	0.4922	0.5078
Resumen 1	0.4385	0.5615	0.4381	0.5619

Tabla 7.4: Distribución de probabilidad calculada de la antimoda a la moda, y de la cero a la moda, recorriendo el círculo en sentido antihorario.

La Tabla 7.4 ofrece la probabilidad acumulada que se calcula de la dirección antimoda a la dirección modal, y de la dirección cero a la moda, ambas recorriendo el círculo en sentido antihorario. Se observa que para cada período impar (inclusive para el período resumen 1), no hay mucha diferencia entre las distribuciones calculadas con uno y otro inicio, y esto es porque el arco corto que separa al cero de

la antimoda no acumula mucha masa de probabilidad.

Calcular la probabilidad acumulada de la dirección cero a la moda, equivale a calcular la probabilidad de registrar al menos un caso de dengue entre el principio de mayo hasta la semana modal. En general, se concluye que en los períodos impares, hay más posibilidad de incidencia a partir de la semana modal o después de ella, hasta terminar el período. Esto refleja la característica de sesgo en sentido antihorario de los períodos 1 y 3.

En el período 1, se esperaba registrar casi un 52% del total casos de dengue durante la semana modal (a partir de la semana 42) hasta el final del período. En el período 3, se esperaba registrar casi un 60% del total de incidencia durante la semana modal (la semana 42) hasta el final del período. Y en el período 5, había casi la misma posibilidad de incidencia desde principios de mayo hasta la semana modal, o de la semana modal hasta terminar el período.

De acuerdo al historial global del comportamiento de los datos en los períodos 1, 3 y 5, el resumen 1 estima que se espera hasta un 56% del total de la moda (a partir de la semana 43) hasta terminar el período.

Sobre los períodos pares

También hay similitudes marcadas encontradas en los primeros períodos pares 2 y 4. Entre las similitudes cualitativas se encuentran el que posean un pequeño sesgo horario con respecto a la dirección media, que sus direcciones medias muestrales estén relativamente lejos (y con valores muy cercanos entre sí) de 180° , y que posean valores de concentración considerables. Entre las distribucionales se encuentran que los valores máximos de los modelos circulares ajustados están ligeramente arriba de 0.4 (0.4214 para el período 2, 0.4371 para el período 4), que posean una cima plana en la forma polar de la distribución, y que las direcciones modales que estiman sus modelos (y que son los que maximizan la densidad) se encuentran relativamente lejos de 180° .

El período 6 ya no muestra un sesgo horario, sino un comportamiento reflexivamente simétrico alrededor de la dirección media, que también está relativamente lejos de 180° . Aquí ya no se exhibe un comportamiento de concentración sino uno de dispersión. La representación polar de la densidad ajustada sigue teniendo cima plana, pero el máximo ya no está por arriba de 0.4 sino cerca de 0.3 (en realidad es 0.3031).

En la Tabla A.7 del Apéndice A llama la atención la existencia de dos meses “punta”, es decir, dos meses que contienen los mayores porcentajes del total real

esperado de incidencia de dengue, en un período par particular, y la diferencia entre estos dos porcentajes es de menos de un punto porcentual. Estos meses se han coloreado de verde.

En el caso del período 2, el rango de meses arriba del 5% fue de 6, y abarcó de julio a diciembre del 2010. Juntos, en estos meses se esperaba registrar un 91.74% del total. Los meses punta, septiembre y octubre, con 20.93% y 21.86% respectivamente, juntos concentraron un 42.79%.

En el caso del período 4, el rango de meses arriba del 5% siguió siendo de 6, y abarcó de julio a diciembre del 2012. Juntos, en estos meses se esperaba un 86.73% del total. Los meses punta, septiembre y octubre, con 21.42% y 21.40% respectivamente, juntos concentraron un 42.82%.

En el caso del período 6, el rango de meses arriba del 5% fue de 7, y abarcó de junio a diciembre de 2014, y en ellos se esperaba un 84.63% del total. Los meses punta, septiembre y octubre, con 15.58% y 15.51% respectivamente, juntos concentraron un 31.09%. Sin embargo, el mes de agosto tiene un porcentaje muy cercano al de los meses punta, con un 14.87% del total.

Período	Meses punta	Acumulado de meses punta	Meses arriba del 5 %	Acumulado de los meses arriba del 5 %
2	septiembre octubre	20.93 % 21.86 %	julio - diciembre (6)	91.74 %
4	septiembre octubre	21.42 % 21.40 %	julio - diciembre (6)	86.73 %
6	septiembre octubre	15.58 % 15.51 %	junio - diciembre (7)	84.63 %
Resumen 2	septiembre octubre	20.29 % 20.84 %	junio - diciembre (6)	87.18 %

Tabla 7.5: Meses punta y meses arriba del 5% de incidencia de dengue, para un período par.

La Tabla 7.5 muestra un resumen de la información acerca de los porcentajes mensuales para los períodos pares. Los meses punta han sido los mismos en todos los períodos pares: septiembre y octubre. Los porcentajes alcanzados en los meses punta en el período 2 son similares a los alcanzados en los meses punta del período 4, los cuales son cercano o están arriba del 21%. En cuanto a los meses punta del período 6, aquí hubo una pérdida de porcentaje respecto a los anteriores, al tener valores cercanos al 15.5% del total.

El rango de meses arriba del 5% fue el mismo en los períodos 2 y 4, pero se registra una pérdida del porcentaje de incidencia en este último. En el período 6 el rango de meses se amplía, fueron 7; sin embargo, también se puede notar que en ese entonces se requirieron de más meses para alcanzar un porcentaje un tanto cercano al que se tuvo en 6 meses en el período 4.

La Tabla A.7 también exhibe los porcentajes mensuales para el período resumen 2: Mayo Año Par - Abril Año Impar. Este período resumen hereda la existencia de dos meses punta vistos en los períodos pares. La Tabla 7.5 muestra que de acuerdo con el historial global del comportamiento de los casos de dengue en estos tres períodos, los meses punta han resultado ser, como era de esperarse, septiembre y octubre, con porcentajes poco arriba del 20%. También, de este resumen, el historial del comportamiento de los datos sólo cuenta como meses significativos a los que están en el rango de julio a diciembre de un año par, y juntos acumulan un porcentaje de incidencia del 87.18%.

Además, en la Tabla 7.5 es interesante observar que el último mes arriba del 5% de incidencia en los períodos pares es diciembre, inclusive esta característica es heredada por el período resumen 2.

Período	Localización					
	Dirección media muestral $\bar{\theta}$		Dirección mediana muestral $\tilde{\theta}$		Dirección modal teórica	
	Semana	Mes	Semana	Mes	Semana	Mes
2	39	Septiembre (4 ^a)	39	Septiembre (4 ^a)	40	Octubre (1 ^a)
4	38	Septiembre (3 ^a)	38	Septiembre (3 ^a)	39	Septiembre (4 ^a)
6	38	Septiembre (3 ^a)	39	Septiembre (4 ^a)	38	Septiembre (3 ^a)
Resumen 2	38	Septiembre (3 ^a)	38	Septiembre (3 ^a)	39 - 40	Septiembre (4 ^a) - Octubre (1 ^a)

Tabla 7.6: Semanas a las que apuntan las direcciones media muestral, mediana muestral, y modal teórica, y su ubicación en los meses de un año normal.

En la Tabla 7.6 se observa que las semanas a las que apuntan las medidas de localización, en general pueden ser englobadas en el rango de la semana 39 (3^a de septiembre) a la semana 40 (1^a de octubre), es decir, de la segunda mitad de septiembre a principios de octubre. Para los períodos pares 2, 4 y 6, en general, se puede ver que la semana en la que se alcanza la mitad del total de los casos registrados en un período en particular, es decir, la semana a la que apunta la dirección mediana muestral, será igual o la inmediata posterior a la semana alrededor de la cual se concentran las grandes frecuencias de dengue, es decir, la semana a la cual apunta la dirección media muestral. Lo anterior también sucede con la semana modal, o sea, la semana a la cual apunta la dirección modal estimada por el modelo ajustado, será igual o la inmediata posterior a la semana media.

En la Tabla 7.6 también se tiene que, de acuerdo al historial global del comportamiento de los datos en estos tres períodos, el resumen 2 estima que alrededor de la semana 38, la tercera de septiembre, se han concentrado las mayores frecuencias de dengue y, además, en ella se ha alcanzado la mitad del total de los casos registrados en el resumen. Sin embargo, la mayor frecuencia de dengue se habrá dado entre la semana 39, la cuarta de septiembre, y la 40, la primera de octubre.

Período	Distribución (en sentido antihorario)			
	De la antimoda a la moda		Del cero a la moda	
	Probabilidad	Complemento	Probabilidad	Complemento
2	0.5512	0.4488	0.5461	0.4539
4	0.5469	0.4531	0.5317	0.4683
6	0.5	0.5	0.4705	0.5295
Resumen 2	0.5510	0.449	0.5382	0.4618

Tabla 7.7: Distribución de probabilidad calculada de la antimoda a la moda, y de la cero a la moda, recorriendo el círculo en sentido antihorario.

La Tabla 7.7 ofrece la probabilidad acumulada que se calcula de la dirección antimoda a la dirección modal, y de la dirección cero a la dirección modal, ambas recorriendo el círculo en sentido antihorario. Se observa que para los períodos 2 y 4 (inclusive para el período resumen 2), los resultados obtenidos de los cálculos con y otro inicio llevan a conclusiones que van en el mismo sentido, y esto es porque el arco corto que separa al cero de la antimoda no acumula mucha masa de probabilidad.

Calcular la probabilidad acumulada de la dirección cero a la moda, equivale a calcular la probabilidad de registrar al menos un caso de dengue entre el principio de mayo hasta la semana modal. Para los períodos 2 y 4, se concluye que había más posibilidad de incidencia del principio del período hasta la semana modal. Esto refleja la característica de sesgo en sentido horario de estos dos períodos.

En el período 2, se esperaba que de la primera semana de mayo hasta la semana modal (semana 40) se hubiese registrado casi un 55 % del total real para ese período. En el período 4, se esperaba que de la primera semana de mayo hasta la semana modal (semana 39) se hubiese registrado hasta un 53 % del total. Sin embargo, para el período 6, se llega a una conclusión en el sentido inverso al que se había llegado en los dos períodos anteriores, y es que aquí, se esperaba que de la semana primera semana de mayo hasta la semana modal (38) se hubiese registrado casi un 47 % del total, es de decir, hay más posibilidad de incidencia a partir de la semana modal o después de ella hasta el final del período.

Hay que notar que, en el caso del período 5, los resultados que se obtienen al medir la distribución de la antimoda a la moda, están de acorde a la característica de simetría reflexiva que se contempla en el modelo ajustado para los datos de ese período, sin embargo, estos resultados cambian cuando se calcula la distribución del cero a la moda, pues el arco corto que separa a la antimoda del cero ahora aporta un masa de probabilidad pequeña, pero considerable.

De acuerdo con el historial global del comportamiento de los datos para los períodos 2, 4 y 6, el resumen 2 estima que se espera hasta un 54 % del total del total, de la

primera semana de mayo hasta la moda que podría darse alrededor de las semanas 39 y 40.

7.3. Sobre la temporada de lluvias

Las lluvias y la migración son los principales factores de incidencia de dengue¹. Aquí, se analizará el comportamiento de las precipitaciones que se dan mensualmente en el rango de mayo 2009 a mayo 2015, y su posible efecto sobre el comportamiento de los casos de dengue registrados en los períodos de estudio que se han considerado. Los valores de precipitación mensual que se han utilizado para dicho análisis, se han obtenido de acuerdo con la información histórica recabada por la Comisión Nacional del Agua (CONAGUA), y que es presentada a través del Servicio Meteorológico Nacional (SMN)².

La Tabla A.8 del Apéndice A muestra los niveles de precipitación mensual en milímetros (mm) y también los porcentajes mensuales del total real de incidencia de dengue, obtenidos para los seis períodos de estudio mayo - abril anuales. Se ha calculado la media global de los niveles de precipitación que se han dado de enero del 2004 a la fecha, junio del 2015, y ha dado como resultado el valor de 83.5326 mm al mes. Se han considerado como “grandes niveles” o “valores considerables” de precipitación a los valores que están arriba de la media global, o los que están relativamente cerca de ella, y se han marcado con rojo en esta Tabla. En abril de 2010 se ha dado un precipitación de 53.8 mm, pero también se ha marcado con rojo porque hay un salto considerable respecto a lo que se tenía un mes antes, es decir, en marzo de 2010 con 11.4 mm; un argumento análogo se ha utilizado para el valor de mayo de 2013 (72.5 mm) con respecto a abril de 2013 (37 mm). También se ha marcado con rojo el valor de 71 mm obtenido en mayo de 2012 pues se presenta entre meses en los que se registran valores de precipitación muy por arriba de la media; un argumento análogo se ha utilizado para el valor de diciembre de 2013 (69.1 mm).

La Figura 7.2 muestra graficados los valores de precipitación de enero de 2009 a junio de 2015. Es evidente la existencia de ciclos con periodicidad de 12 meses, de hecho, en general, se puede observar que precisamente mayo se encuentra entre los primeros dos meses en los que se registra una precipitación arriba o muy cercana

¹“Lluvias 'riegan' el dengue” <http://sipse.com/milenio/precipitaciones-empeoran-el-panorama-del-dengue-en-yucatan-54134.html>

²CONAGUA. Temperaturas y Lluvia. http://smn.cna.gob.mx/index.php?option=com_content&view=article&id=12&Itemid=77

a la media, o bien, en los que se empieza a percibir un franco crecimiento de los niveles de precipitación.

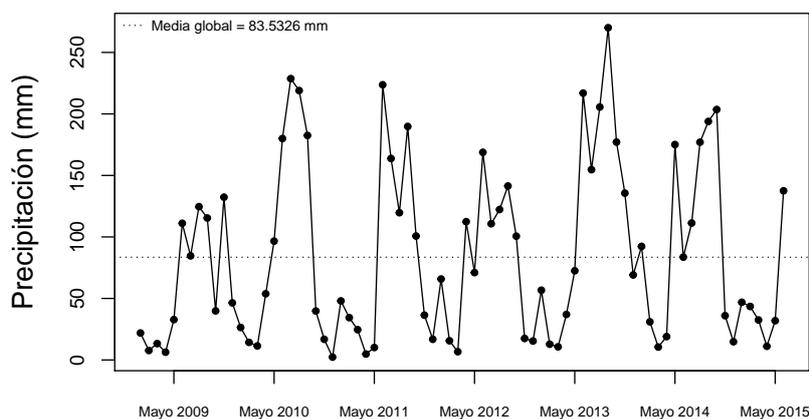


Figura 7.2: Valores de precipitación pluvial (mm) entre enero de 2004 y junio de 2015. La media global que se obtiene en este lapso es de 83.5326 mm.

A fin de tener una forma de comparación, los niveles de precipitación pluvial que se han observado en la Figura 7.2 y los casos de dengue mensuales que se han observado en la Figura 2 de la *Introducción*, se han estandarizado, es decir, a cada valor de un respectivo conjunto de datos se le ha restado su media y a esta diferencia se le ha dividido entre su desviación estándar. El resultado de este procedimiento se resume en la gráfica de valores estandarizados de la Figura 7.3. En esta gráfica podemos observar que hay un desfase entre los ciclos de precipitación pluvial y los de casos de dengue registrados. Por ende, hay un desfase entre los grandes valores de precipitación y las grandes frecuencias de incidencia de dengue.

La Tabla A.8, además de los grandes niveles de precipitación marcados en rojo, ofrece los porcentajes mensuales del total de incidencia arriba del 5% marcados de amarillo. Es evidente un desfase entre las grandes precipitaciones y los grandes porcentajes de incidencia. De manera global, hay un desfase de entre 1 y 3 meses hacia adelante de los grandes porcentajes con respecto a los grandes niveles de precipitación. Este desfase es razonable, pues obedece a la proliferación de criaderos de mosquito, al tiempo de incubación del vector *Aedes aegypti* y al tiempo de propagación de la enfermedad que tiene que pasar, una vez iniciadas las grandes precipitaciones, y antes de registrar numerosos casos de dengue.

Al observar la Figura 7.3 y el comportamiento observado en la Tabla A.8, se puede inferir que, si bien la forma de la distribución de los casos de dengue no está determinada totalmente por la forma de la distribución de los niveles de precipitación, los grandes niveles de precipitación sí influyen en la ocurrencia de las grandes

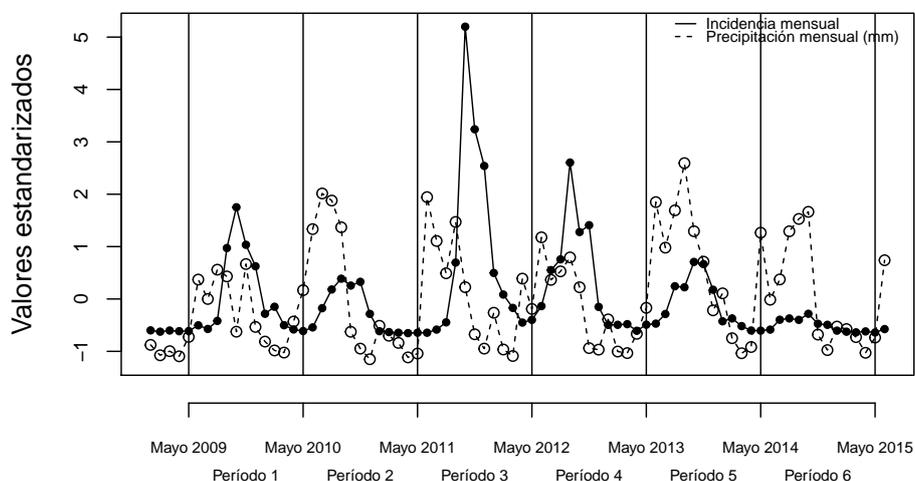


Figura 7.3: Comparación entre los valores estandarizados de la precipitación pluvial mensual y los casos mensuales de dengue, que se presentan entre enero de 2009 a junio de 2015.

frecuencias de casos registrados en un período en particular. Las precipitaciones arriba de la media global no influyen en el número de casos de lo que es considerado una frecuencia elevada, pero sí en terminar de manera relativa qué es una frecuencia elevada y qué es una frecuencia baja en un período particular.

Anual			Períodos de estudio					
Año	Suma	Promedio	Período	Suma	Promedio	Rango de meses altos	Suma meses altos	Promedio meses altos
2009	736.3	61.3583	1	793	66.0833	5	568	113.600
2010	1071.6	89.3000	2	1077.5	89.7917	6	960.6	160.100
2011	972.8	81.0667	3	1061.5	88.4583	5	797.7	159.540
2012	948.2	79.0167	4	864.9	72.0750	7	827.2	118.171
2013	1418.8	118.2333	5	1454.4	121.2000	9	1393.9	154.878
2014	1148.1	95.6750	6	1129.3	94.1083	6	944.5	157.417

Tabla 7.8: Valores de precipitación, en milímetros (mm), obtenidos de manera anual y por período de estudio.

A continuación se analizará el efecto de las precipitaciones en la incidencia de dengue en cada período de estudio. Se empezará con los períodos impares y luego los períodos pares. En todo momento se estará tomando en cuenta las Tablas 7.8 y A.8.

En los períodos impares

En el período de estudio 1 hubo un rango de 6 meses, junio - noviembre del 2009, en el que se presentaron grandes niveles de precipitación; sólo en un mes intermedio hubo un caso atípico de nivel bajo de precipitación (octubre del 2009 con 39 mm). A partir de junio del 2009, tuvieron que pasar dos meses más para que en agosto del 2009 se tuviera un porcentaje esperado arriba del 5 % del total de incidencia de dengue; de hecho, de manera paralela, hubo un rango de 6 meses, agosto 2009 - enero 2010, arriba del 5 % de incidencia.

En el período 3 se tuvo un rango de 5 meses en que se registraron niveles de precipitación arriba de la media, junio - octubre de 2011; de junio de 2011 tuvieron que pasar 3 meses para tener un mes arriba del 5 % de porcentaje esperado del total de incidencia, en septiembre de 2011. Paralelamente, hubo un rango de 5 meses con arriba del 5 % de incidencia, septiembre 2011 - enero de 2012. Sin embargo, desde febrero de 2011 hasta julio del mismo año se tuvieron porcentajes abajo del 1 %, para luego en agosto de 2011 tener un porcentaje de 1.1533 %, lo cual es más del doble de lo que se tenía un mes antes, en julio (0.4228 %); dos meses antes de agosto, ya se tenían registros de grandes niveles de precipitación en junio, por ende se podría decir que el efecto de las lluvias se da en realidad en 6 meses, agosto - enero, sólo que en 5 de ellos se tienen porcentajes relativamente altos de incidencia. La aparente compactación de los altas frecuencias en cinco meses obedece a causas externas a las lluvias. Aunque los grandes niveles de precipitación se dan en 5 meses, el total de precipitación ocurrida en ellos, de 797.7 mm, es suficiente para tener un efecto de seis meses en la incidencia de dengue, y este valor representa un incremento de cerca del 40.44 % respecto a la cantidad de precipitación acumulada en los meses de grandes valores en el período 1 (6 meses con 568 mm).

De acuerdo con lo observado en los períodos 1 y 3, cuando se esperaba registrar el primer valor alto de precipitación en junio, en el período 5 esto no es así, se adelanta un mes y sucede en mayo de 2013; y no sólo eso, sino que el rango de meses con altas precipitaciones se incrementa a 9. Sin embargo, el efecto de las lluvias en la incidencia de dengue sigue sucediendo dos meses después del mes de inicio del rango de grandes precipitaciones; julio es el primer con porcentaje arriba del 5 % de incidencia y, por ende, el rango de meses con porcentajes altos de incidencia también se adelanta, un mes, y aumenta a 7 meses, julio 2013 - enero 2014.

De hecho, el período 5 es el más lluvioso con un acumulado de 1454.4 mm de precipitación, según la información de la Tabla 7.8. Uno podría pensar que a mayor lluvia, mayor casos de dengue, pero en el caso del estado de Yucatán esto no se cumple y este período es un claro ejemplo. Consideremos el período inmediato

anterior, el período 4, se puede observar que está entre los períodos menos lluviosos, con un acumulado de 864.9 mm, y que también está entre los períodos con más casos de reportados, con un total de 4833; en el período 5 se tuvieron 2822 casos registrados, es decir, del período 4 al 5 hubo una reducción del 41.61 % de los casos registrados pero un incremento del 68.16 % en la precipitación. La reducción de los casos de dengue en el período 5 con respecto a los anteriores (a pesar de ser un período muy lluvioso), podría deberse al posible incremento de programas de prevención y lucha antivectorial que se dan en este período, y posiblemente meses antes, o bien, a una forma más eficaz de llevar a cabo los programas que ya se tienen establecidos.

Durante el período 5, el efecto de baja concentración, cercano a un comportamiento de dispersión, podría explicarse por dos sucesos: la ampliación del rango de meses con grandes porcentajes esperados del total, y el descenso en los números de casos de dengue que conforman las grandes frecuencias. En cuanto a la ampliación del rango de meses con porcentajes significativos de incidencia, esto es producto de que el rango de meses con grandes precipitaciones también se amplía. En ambos casos, la ampliación se presenta al contemplar meses próximos al inicio del período, que no se habían contemplado como meses significativos de precipitación o de incidencia en los períodos impares anteriores, por ejemplo, el mes de mayo en el caso de las precipitaciones o el mes de julio en el caso de la incidencia (ver Tabla A.8, y observar los períodos impares); al adelantarse las grandes precipitaciones, entonces se empiezan a tener numerosos casos de dengue antes, y esto contribuye a la dispersión de los datos. Las grandes frecuencias en este período están conformadas por números de casos de dengue mucho menores a los números de casos que conforman las grandes frecuencias en los períodos 1 y 3; el descenso de los números de casos de dengue que conforman las grandes frecuencias provoca que la masa de probabilidad, considerablemente concentrada alrededor de la dirección media en los períodos impares anteriores, escape en ambas direcciones, hacia los extremos, y ello hace que se le asigne un porcentaje mayor a los últimos y primeros meses del período, sin que eso signifique necesariamente un aumento extraordinario del número de casos de dengue en ellos. La dispersión de los datos es lo que provoca la pérdida de punta del modelo ajustado en su representación polar, y que el valor máximo de la densidad se reduzca considerablemente al observar la representación lineal del modelo.

Sin embargo, la forma simétrica de la distribución en el período 5 obedece principalmente a la reducción de casos de dengue en las grandes frecuencias; esta baja en el número de casos registrados ejerce un efecto de reescalamiento de las frecuencias, lo que provoca a su vez un efecto de simetría en los datos del período 5. La reducción de los casos no es atribuible a las lluvias, es decir, depende de factores

externos a ellas, pues se sabe que este período ha sido el más lluvioso.

En los períodos pares

En el período 2, sucede que las grandes precipitaciones empiezan un mes antes del inicio de este período de estudio, es decir, en abril del 2010 se registra un valor considerable de precipitación. Hubo un rango de 6 meses, abril - septiembre del 2010, con altos valores de precipitación. A partir de abril de 2010, tuvieron que pasar tres meses más para que en julio del 2010 se tuviera un porcentaje arriba del 5 % del total de incidencia. De manera paralela, hubo un rango de 6 meses, julio - diciembre del 2010, arriba del 5 % de incidencia.

En el período 4, sucede un comportamiento similar al visto en el período 2. Las grandes precipitaciones empiezan en abril de 2012, es decir, un mes antes del inicio de este período de estudio. Hay un rango de 7 meses, abril - octubre del 2012, con altos valores de precipitación. A partir de abril de 2012, tuvieron que pasar tres meses más para que en julio del 2012 se tuviera un porcentaje arriba del 5 % del total de incidencia. De manera paralela, hubo un rango de 6 meses, julio - diciembre del 2012, arriba del 5 % de incidencia. Aunque el rango de meses en este período es de 7, el acumulado de precipitación en estos meses, 827.2 mm, es menor al acumulado en el rango de 6 meses en el período 2, de 960.6 mm; sin embargo, este valor de precipitación es suficiente para tener a la par un rango de 6 meses de casos de altas frecuencias de dengue.

De acuerdo con lo observado en los períodos 2 y 4, cuando se esperaba registrar el primer valor alto de precipitación en abril, en el período 6 esto no es así, sucede un mes después, en mayo de 2014; y no sólo eso, sino que el rango de meses con alta precipitación se reduce a 5 meses, además de que el efecto de las lluvias en la incidencia de dengue se da el mes inmediatamente después del inicio del rango de altas precipitaciones, en junio, el cual es el primero con porcentaje arriba del 5 % de incidencia. El rango de meses con porcentajes altos de incidencia aumenta a 7 meses, junio - diciembre 2014.

Durante el período 6, la cantidad numerosa de meses con grandes porcentajes de incidencia no es explicada por el rango de grandes precipitaciones, pues éste inclusive se reduce. Tanto el efecto de dispersión como la forma un tanto simétrica de la distribución obedece principalmente a que las grandes frecuencias en este período están conformadas por números de casos de dengue mucho menores a los números de casos que conforman las grandes frecuencias en los períodos 2 y 4, inclusive al compararlas con las de los períodos impares; esta baja del número de casos ejerce un efecto de reescalamiento que provoca a su vez un efecto de simetría

en los datos de este período. Al mismo tiempo, el efecto de reescalamiento provoca que la masa de probabilidad, concentrada alrededor de la dirección media en los períodos pares anteriores, escape hacia los extremos, y ello hace que se le asigne un porcentaje del total mayor a los últimos y primeros meses del período, sin que eso signifique necesariamente un aumento extraordinario del número de casos de dengue en ellos. Todo ello conlleva a que la cima plana del modelo ajustado en su representación polar se prolongue aún más, y que el valor máximo de la densidad se reduzca considerablemente al observar la representación lineal del modelo.

7.4. Conclusiones generales

En cuanto a los aspectos técnicos del presente trabajo, se concluye que la familia de distribuciones flexibles propuesta por Jones y Pewsey (2012), las distribuciones Inversas Batschelet, se ajustan razonablemente bien a los datos de incidencia en cada período de estudio. Modelos completos y submodelos (entre ellos la distribución von Mises) pertenecientes a esta familia han resultado ser los mejores al modelar la incidencia de dengue en un período mayo-abril anual.

También, si se consideran períodos de estudio mayo-abril anuales, definiéndolos a partir de año 2009, se tendrá que en los llamados períodos impares se observarán características tales como que los datos de incidencia exhibirán sesgo en sentido antihorario o bien simetría reflexiva con respecto a la dirección media, además de contar con un mes “punta” que concentrará el mayor porcentaje de incidencia. En los llamados períodos pares, se observarán características tales como que los datos de incidencia exhibirán sesgo en sentido horario o bien simetría reflexiva con respecto a la dirección media, además de contar con dos meses “punta” cuyos respectivos porcentajes de incidencia diferirán entre sí de a lo más un punto porcentual, y que juntos concentrarán el mayor porcentaje de incidencia en el período.

Con base en los pasados resultados, y en muchos otros presentados a lo largo de este trabajo de investigación, se han encontrado ciertas similitudes en el comportamiento de la incidencia de dengue, que se exhiben si se estudia este fenómeno considerando períodos mayo-abril anuales, y que sugieren la existencia de un aparente dualidad en el comportamiento de la incidencia de dengue en Yucatán, de acuerdo con los registros de la Dirección General de Epidemiología. Es decir, pareciera que existiesen dos estructuras de comportamiento de la incidencia que se presentan cada dos períodos mayo-abril anuales; se tiene uno en los períodos impares, y otro en los períodos pares. Al menos, este aparente comportamiento dual se exhibe en los seis períodos mayo-abril anuales que se han considerado en es-

te trabajo de investigación. Sin embargo, sería muy aventurado afirmar que este comportamiento dual sea definitivo, es decir, que se espere se presente en todos los períodos mayo-abril anuales venideros, pues estas características de similitud entre unos y otros períodos podrían depender de manera directa del efecto de la acción del hombre sobre la incidencia de dengue, al implementar o dejar de implementar programas de prevención y lucha antivectorial en lugares y períodos determinados, al incremento o disminución de estos programas, o bien, a la mejora o empeoramiento en la forma de llevar a cabo estos programas de manera eficaz (aunque es de esperarse que en la actualidad estos programas no disminuyan o empeoren sino todo lo contrario).

De la Figura 7.1 de la Sección 7.2, se observa que un período impar determinado registra más casos de dengue que el período par inmediato (observar los valores de los períodos 1 y 2, 3 y 4, y 5 y 6). Inclusive, al comparar las tablas descriptivas 6.9 y 6.12 (ver Capítulo 6) de los períodos resumen 1 y 2, respectivamente, el resumen de los períodos impares, el 1, registra más casos de dengue que el resumen de los períodos pares, el 2. De hecho, de la Tabla A.7 del Apéndice A, si se observa los porcentajes de incidencia que se tienen para el mes de octubre, uno de los dos meses con mayor porcentaje de incidencia en general, se encontrará que dicho porcentaje en este mes es mayor en los períodos impares que en los períodos pares, y es que, en efecto, octubre es el mes “punta” en los períodos impares, es decir, el de mayor porcentaje de incidencia. Con estas nuevas observaciones, una vez más, se exhibe que hay un claro comportamiento dual de la incidencia de dengue que se presenta, como ya se ha dicho, cada dos períodos mayo-abril anuales y que se exhibe, al menos, para los seis períodos considerados en este estudio. Más aún, hay indicios de que esta aparente dualidad ya se ha observado con anterioridad pues, en el caso del dengue, “esporádicamente, cada dos o tres años, vamos a ver un repunte de casos”³ según asegura el infectólogo Carlos Machaín Williams, profesor investigador del Centro de Investigaciones Regionales “Dr. Hideyo Noguchi” de la Universidad Autónoma de Yucatán. Lo anterior está de acuerdo con lo observado en los períodos impares, con respecto a que uno en particular presenta un mayor registro de los casos de dengue, si se le compara con el período par inmediato. En resumen, si estas estructuras de comportamiento no son definitivas, cuando se presenten, se espera que lo hagan de manera ocasionalmente cada dos o tres años, según asegura Machaín Williams.

Habrá que tener muy en cuenta los demás períodos mayo-abril anuales venideros, para poder observar si esta aparente dualidad en el comportamiento de incidencia de dengue se mantiene o bien, se observan cosas distintas producto de la acción

³En guardia por el zika: file:///C:/Users/Miguel/Downloads/En%20guardia%20por%20el%20zika%20-%20Dengue_El%20Diario%20de%20Yucat%C3%A1n19nov2015.html

del hombre sobre la incidencia.

Por lo pronto, en el período mayo-abril anual más próximo, es decir, el período 7 definido como mayo de 2015-abril de 2016, de acuerdo con el resumen de la información que se tiene para los períodos impares a través del período resumen 1 (ver Sección 6.4.1 del Capítulo 6), se espera: que sea el rango de meses entre septiembre y diciembre el que concentre los mayores números semanales de casos de dengue, que los datos semanales exhiban sesgo en sentido antihorario o simetría reflexiva con respecto a la dirección media, que los datos se concentren alrededor de una semana próxima a la 45 (la primera de noviembre), que la semana que registre el mayor número de casos sea una que incluya los últimos días de octubre o los primeros días de noviembre, y que sea octubre el mes con el mayor número de casos de dengue registrados, con hasta un 25.71 % del total de incidencia en el período (ver Tabla A.7 del Apéndice A).

7.5. Trabajos futuros

Debido a las similitudes observadas entre los denominados períodos impares y los pares, es menester seguir muy de cerca lo que acontezca en los futuros períodos mayo-abril anuales considerados durante y después del 2015. Esto a fin de confirmar la aparición de dos estructuras de dengue subyacentes en los registros de los casos de dengue que se tienen para el estado de Yucatán, según la Dirección General de Epidemiología.

Además, sería interesante observar si estas estructuras pudiesen ser modeladas siguiendo una misma familia de distribuciones circulares (como la familia de distribuciones flexibles Inversa Batschelet). Si es así, entonces analizar los parámetros que gobiernan la distribución para un período mayo-abril anual y compararlos con los correspondientes a los otros períodos, podría dar pie a exhibir posibles relaciones funcionales entre estos parámetros, y así poder hacer predicciones a futuro acerca del comportamiento de la incidencia de dengue.

También, hay una invitación para repetir los cálculos de los estimadores de verosimilitud de los parámetros de los modelos von Mises, Jones-Pewsey e Inversa Batschelet (con base von Mises), tomando las consideraciones necesarias para datos agrupados en las funciones computacionales que hacen estos cálculos (Pewsey *et al.*, 2013), a fin de confirmar que los resultados arrojados serán cercanos a los aquí obtenidos, esto derivado de la gran cantidad de datos y de intervalos de clase con los que se ha trabajado.

Además, existen pruebas de uniformidad que consideran una alternativa general

y no solo una unimodal como la prueba de Rayleigh. Estas pruebas reciben el nombre de *omnibus tests* y algunas de ellas, como la prueba V_n de Kuiper y la U^2 de Watson, poseen versiones modificadas para su uso con datos agrupados (Mardia y Jupp, 1999). Sería interesante implementar estas pruebas junto con sus respectivas modificaciones para datos agrupados y comparar sus resultados con los aquí obtenidos utilizando solo la prueba de Rayleigh.

Apéndice A

Tablas

Numeración	Orden de aparición en el calendario gregoriano	Número de días acumulados	Semana ubicada en los meses de un <i>año normal</i> (no bisiesto)	Meses que abarca
1	18	7	30 abril - 6 mayo	abril - mayo
2	19	14	7 - 13	mayo
3	20	21	14 - 20	mayo
4	21	28	21 - 27	mayo
5	22	35	28 mayo - 3 junio	mayo - junio
6	23	42	4 - 10	junio
7	24	49	11 - 17	junio
8	25	56	18 - 24	junio
9	26	63	25 junio - 1 julio	junio - julio
10	27	70	2 - 8	julio
11	28	77	9 - 15	julio
12	29	84	16 - 22	julio
13	30	91	23 - 29	julio
14	31	98	30 julio - 5 agosto	julio - agosto
15	32	105	6 - 12	agosto
16	33	112	13 - 19	agosto
17	34	119	20 - 26	agosto
18	35	126	27 agosto - 2 septiembre	agosto - septiembre
19	36	133	3 - 9	septiembre
20	37	140	10 - 16	septiembre
21	38	147	17 - 23	septiembre
22	39	154	24 - 30 septiembre	septiembre
23	40	161	1 - 7 octubre	octubre
24	41	168	8 - 14	octubre
25	42	175	15 - 21	octubre
26	43	182	22 - 28	octubre
27	44	189	29 octubre - 4 noviembre	octubre - noviembre
28	45	196	5 - 11	noviembre
29	46	203	12 - 18	noviembre
30	47	210	19 - 25	noviembre
31	48	217	26 noviembre - 2 diciembre	noviembre - diciembre
32	49	224	3 - 9	diciembre
33	50	231	10 - 16	diciembre
34	51	238	17 - 23	diciembre
35	52	246	24 - 31	diciembre
36	1	253	1 - 7 enero	enero
37	2	260	8 - 14	enero
38	3	267	15 - 21	enero
39	4	274	22 - 28	enero
40	5	281	29 enero - 4 febrero	enero - febrero
41	6	288	5 - 11	febrero
42	7	295	12 - 18	febrero
43	8	302	19 - 25	febrero
44	9	309	26 febrero - 4 marzo	febrero - marzo
45	10	316	5 - 11	marzo
46	11	323	12 - 18	marzo
47	12	330	19 - 25	marzo
48	13	337	26 marzo - 1 abril	marzo - abril
49	14	344	2 - 8	abril
50	15	351	9 - 15	abril
51	16	358	16 - 22	abril
52	17	365	23 - 29	abril

Tabla A.1: Ubicación de las semanas de un período de estudio mayo-abril anual en el calendario gregoriano para una año normal (no bisiesto).

Orden de aparición en el calendario gregoriano	Representante semanal (grados)	Representante semanal (radianes)	Mayo 2009 - Abril 2010	Mayo 2010 - Abril 2011	Mayo 2011 - Abril 2012	Mayo 2012 - Abril 2013	Mayo 2013 - Abril 2014	Mayo 2014 - Abril 2015	Mayo Impar - Abril Par	Mayo Par - Abril Impar
18	2.95890411°	0.051642619	5	7	3	29	9	4	17	40
19	9.863013699°	0.172142063	6	8	4	29	13	11	23	48
20	16.76712329°	0.292641507	8	7	3	19	10	6	21	32
21	23.67123288°	0.413140952	8	5	4	36	21	0	33	41
22	30.57534247°	0.533640396	12	10	3	38	23	9	38	57
23	37.47945205°	0.65413984	15	9	4	36	11	6	30	51
24	44.38356164°	0.774639284	12	6	3	67	26	9	41	82
25	51.28767123°	0.895138729	17	10	4	42	18	8	39	60
26	58.19178082°	1.015638173	15	20	3	36	29	15	47	71
27	65.09589041°	1.136137617	13	30	4	107	38	22	55	159
28	72°	1.256637061	12	39	5	180	18	42	35	261
29	78.90410959°	1.377136506	11	49	8	56	50	22	69	127
30	85.80821918°	1.49763595	7	85	15	153	52	28	74	266
31	92.71232877°	1.618135394	18	95	6	123	54	22	78	240
32	99.61643836°	1.738634838	18	88	19	165	55	27	92	280
33	106.5205479°	1.859134283	37	80	13	130	82	25	132	235
34	113.4246575°	1.979633727	32	84	35	161	84	32	151	277
35	120.3287671°	2.100133171	62	114	27	286	97	18	186	418
36	127.2328767°	2.220632615	95	85	57	191	96	25	248	301
37	134.1369863°	2.34113206	125	0	144	357	107	33	376	390
38	141.0410959°	2.461631504	166	125	134	261	74	17	374	403
39	147.9452055°	2.582130948	218	106	219	229	87	38	524	373
40	154.8493151°	2.702630392	304	114	286	182	129	25	719	321
41	161.7534247°	2.823129837	227	111	393	174	117	22	737	307
42	168.6575342°	2.943629281	231	69	490	227	177	35	898	331
43	175.5616438°	3.064128725	218	83	515	206	136	40	869	329
44	182.4657534°	3.184628169	218	153	686	206	87	38	991	397
45	189.369863°	3.305127614	191	86	326	238	141	18	658	342
46	196.2739726°	3.425627058	158	91	449	177	129	31	736	299
47	203.1780822°	3.546126502	124	76	426	191	82	20	632	287
48	210.0821918°	3.666625946	163	59	379	30	103	14	645	103
49	216.9863014°	3.787125391	125	60	308	30	131	37	564	127
50	223.890411°	3.907624835	113	15	431	87	85	11	629	113
51	230.7945205°	4.028124279	62	16	233	48	67	17	362	81
52	238.1917808°	4.157230827	63	9	326	48	62	10	451	67
1	245.5890411°	4.286337374	12	1	49	1	1	4	62	6
2	252.4931507°	4.406836818	56	11	169	23	14	1	239	35
3	259.3972603°	4.527336262	44	2	120	25	14	4	178	31
4	266.3013699°	4.647835707	48	12	136	25	46	13	230	50
5	273.2054795°	4.768335151	85	7	97	18	27	8	209	33
6	280.109589°	4.888834595	54	6	92	20	37	5	183	31
7	287.0136986°	5.009334039	45	1	76	12	26	6	147	19
8	293.9178082°	5.129833484	30	5	43	25	36	5	109	35
9	300.8219178°	5.250332928	22	4	57	19	25	6	104	29
10	307.7260274°	5.370832372	21	2	52	20	25	2	98	24
11	314.630137°	5.491331816	14	0	42	16	20	4	76	20
12	321.5342466°	5.611831261	2	4	25	20	15	6	42	30
13	328.4383562°	5.732330705	13	4	29	5	5	5	47	14
14	335.3424658°	5.852830149	8	3	11	11	6	1	25	15
15	342.2465753°	5.973329593	7	4	34	6	8	11	49	21
16	349.1506849°	6.093829038	8	1	27	8	6	3	41	12
17	356.0547945°	6.214328482	16	0	20	4	11	9	47	13
			3594	2071	7044	4833	2822	830	13460	7734

Tabla A.2: Representantes angulares para las semanas de un período de estudio mayo - abril anual. Para un período particular, se exhiben las frecuencias de casos de dengue semanales, según la DGE y contemplando las consideraciones hechas en la Sección 5.4. También se exhiben las frecuencias de los dos resúmenes de datos Mayo Año Impar - Abril Año Par, y Mayo Año Par - Abril Año Impar.

Período	Parámetros	Modelos para el período mayo - abril				
		Inversa Batschelet	von Mises sesgada	Inversa Batschelet Simétrica	Jones-Pewsey	von Mises
1	ν	0.06834066	0.0365793	0		0
	λ	0.3231239	0	0.2927446		0
	ξ	3.141593	3.141593	3.074718		μ
	κ	1.633167	1.750218	1.631359	1.55327	1.744845
	μ				3.071628	3.140074
	ψ				-0.4776594	0
	l_{\max}	-4807.958	-4857.68	-4827.102	-4818.764	-4864.515
	AIC	9623.916	9721.36	9660.204	9643.528	9733.029
	BIC	9648.664	9739.921	9678.765	9662.089	9745.403
2	ν	-0.1938533	-0.3333831	0		0
	λ	-0.2857464	0	-0.3051598		0
	ξ	2.29223	2.119792	2.532583		μ
	κ	2.088345	1.784457	2.104757	2.076395	1.769665
	μ				2.543854	2.553416
	ψ				0.243294	0
	l_{\max}	-2748.52	-2774.396	-2752.45	-2768.181	-2784.977
	AIC	5505.04	5554.791	5510.901	5542.362	5573.953
	BIC	5527.583	5571.699	5527.808	5559.27	5585.225
3	ν	0.5074749	0.4991516	0		0
	λ	0.1031728	0	0.1027011		0
	ξ	4.217897	4.206759	3.430241		μ
	κ	2.218277	2.337744	2.117545	2.037716	2.221853
	μ				3.426034	3.451379
	ψ				-0.1554874	0
	l_{\max}	-8172.019	-8183.845	-8414.455	-8406.852	-8426.114
	AIC	16352.04	16373.69	16834.91	16819.7	16856.23
	BIC	16379.48	16394.27	16855.49	16840.28	16869.95
4	ν	-0.2054774	-0.2036836	0		0
	λ	-0.03759723	0	-0.03434725		0
	ξ	2.19966	2.201661	2.474367		μ
	κ	1.540041	1.522607	1.531182	2.037716	1.515572
	μ				3.426034	2.47632
	ψ				-0.1554874	0
	l_{\max}	-6930.155	-6930.971	-6942.955	-8406.852	-6943.552
	AIC	13868.31	13867.94	13891.91	13892.89	13891.1
	BIC	13894.24	13887.39	13911.36	13912.34	13904.07
5	ν	-0.04474859	-0.032749	0		0
	λ	0.0434742	0	0.0314053		0
	ξ	2.865702	2.879542	2.924136		μ
	κ	1.184241	1.193051	1.186378	1.180146	1.19285
	μ				2.922824	2.922138
	ψ				-0.06048432	0
	l_{\max}	-4393.796	-4394.138	-4394.156	-4394.13	-4394.348
	AIC	8795.592	8794.276	8794.312	8794.26	8792.695
	BIC	8819.372	8812.112	8812.148	8812.095	8804.585
6	ν	-0.1293235	-0.2867141	0		0
	λ	-0.274895	0	-0.2890412		0
	ξ	2.324011	2.158657	2.462044		μ
	κ	1.027045	0.9665705	1.029367	1.028588	0.9620456
	μ				2.472272	2.479446
	ψ				0.3812923	0
	l_{\max}	-1356.715	-1360.511	-1356.948	-1359.508	-1361.825
	AIC	2721.429	2727.022	2719.896	2725.017	2727.649
	BIC	2740.315	2741.186	2734.06	2739.181	2737.092
Resumen año impar - año par	ν	0.2971027	0.274622	0		0
	λ	0.2241965	0	0.2078794		0
	ξ	3.720231	3.690173	3.25632		μ
	κ	1.659224	1.759545	1.644426	1.586268	1.736479
	μ				3.258377	3.285394
	ψ				-0.314453	0
	l_{\max}	-18063.79	-18148.32	-18186.37	-18180.48	-18258.16
	AIC	36135.57	36302.64	36378.75	36366.96	36520.32
	BIC	36165.6	36325.16	36401.27	36389.48	36535.33
Resumen año par - año impar	ν	-0.2096247	-0.2235167	0		0
	λ	-0.1187337	0	-0.1317068		0
	ξ	2.223296	2.205966	2.49067		μ
	κ	1.57123	1.509348	1.571206	1.563596	1.502219
	μ				2.494843	2.499243
	ψ				0.1137855	0
	l_{\max}	-11115.78	-11129.23	-11133.99	-11143.18	-11149.66
	AIC	22239.55	22264.46	22273.99	22292.36	22303.32
	BIC	22267.37	22285.32	22294.85	22313.22	22317.23

Tabla A.3: Resumen de los cálculos hechos para cada modelo circular ajustado en un período de estudio mayo-abril anual.

Período	Estadísticos descriptivos circulares			
	$\bar{\theta}$	\bar{R}	\bar{b}_2	\bar{a}_2
(1) Mayo 2009 - Abril 2010	(3.1396,3.1406) 3.1401	(0.6351,0.6670) 0.6510	(-0.134,-0.0968) -0.1154	(0.2999,0.3467) 0.3233
(2) Mayo 2010 - Abril 2011	(2.5527,2.5541) 2.5534	(0.6387,0.6732) 0.6559	(0.0106,0.056) 0.0333	(0.1529,0.2081) 0.1805
(3) Mayo 2011 - Abril 2012	(3.4511,3.4517) 3.4514	(0.7219,0.7401) 0.7310	(-0.1266,-0.1052) -0.1159	(0.3559,0.3869) 0.3714
(4) Mayo 2012 - Abril 2013	(2.4758,2.4768) 2.4763	(0.5860,0.6138) 0.5999	(0.0196,0.0503) 0.0349	(0.1852,0.2235) 0.2043
(5) Mayo 2013 - Abril 2014	(2.9213,2.9230) 2.9221	(0.4897,0.5311) 0.5104	(-0.0157,0.031) 0.0076	(0.1262,0.1758) 0.151
(6) Mayo 2014 - Abril 2015	(2.4775,2.4813) 2.4794	(0.3929,0.4713) 0.4321	(-0.0244 0.0662) 0.0209	(-0.0086,0.0855) 0.0384
Mayo año impar - Abril año par	(3.2851,3.2857) 3.2854	(0.6413,0.6575) 0.6494	(-0.0774,-0.0588) -0.0681	(0.2907,0.3136) 0.3022
Mayo año par - Abril año impar	(2.4988,2.4997) 2.4992	(0.5859,0.6074) 0.5966	(0.0199,0.0443) 0.0321	(0.1651,0.195) 0.1801

Tabla A.4: Intervalos del 95 % de confianza para los estadísticos descriptivos circulares $\bar{\theta}$, \bar{R} , \bar{b} , y \bar{a}_2 para cada período de estudio. Abajo de cada intervalo, está el estimado puntual del respectivo estadístico.

Período	Parámetros			
	ξ	κ	ν	λ
(1) Mayo 2009 - Abril 2010	(3.0067,3.2765) 3.1416	(1.5655,1.7008) 1.6332	(-0.0108,0.1475) 0.0683	(0.2598,0.3864) 0.3231
(2) Mayo 2010 - Abril 2011	(2.1168,2.4677) 2.2922	(1.9333,2.2434) 2.0883	(-0.3305, -0.0572) -0.1939	(-0.3601,-0.2113) -0.2857
(3) Mayo 2011 - Abril 2012	(-2.1311,-1.9995) -2.0653	(2.1431,2.2934) 2.2183	(0.4659,0.5491) 0.5075	(0.0616,0.1447) 0.1032
(4) Mayo 2012 - Abril 2013	(2.0860,2.3133) 2.1997	(1.4757,1.6044) 1.5400	(-0.2872,-0.1237) -0.2055	(-0.0952,0.0200) -0.0376
(5) Mayo 2013 - Abril 2014	(2.7211,3.0103) 2.8657	(1.1160,1.2525) 1.1842	(-0.1485,0.0590) -0.0447	(-0.0596,0.1466) 0.0435
(6) Mayo 2014 - Abril 2015	(1.9205,2.7275) 2.3240	(0.8964,1.1577) 1.0270	(-0.4936,0.2349) -0.1293	(-0.4596,-0.0902) -0.2749
Mayo año impar - Abril año par	(-2.6223,-2.5036) -2.5630	(1.6232,1.6952) 1.6592	(0.2604,0.3338) 0.2971	(0.1904,0.2580) 0.2242
Mayo año par - Abril año impar	(2.1315,2.3151) 2.2233	(1.5174,1.6251) 1.5712	(-0.2784,-0.1408) -0.2096	(-0.1633,-0.0741) -0.1187

Tabla A.5: Intervalos del 95 % de confianza para los parámetros ξ , κ , ν y λ del modelo Inversa Batschelet Completo (base von Mises) que se ajusta para cada período de estudio. Abajo de cada intervalo está el estimado puntual del respectivo parámetro.

Período	Parámetros			
	ξ	κ	ν	λ
(1) Mayo 2009 - Abril 2010	3.1416	1.6332	0.0683	0.3231
(2) Mayo 2010 - Abril 2011	2.2922	2.0883	-0.1939	-0.2857
(3) Mayo 2011 - Abril 2012	4.2179	2.2183	0.5075	0.1032
(4) Mayo 2012 - Abril 2013	2.2017	1.5226	-0.2037	0
(5) Mayo 2013 - Abril 2014	2.9221	1.1929	0	0
(6) Mayo 2014 - Abril 2015	2.4620	1.0293	0	-0.2890
Mayo año impar - Abril año par	3.7202	1.6592	0.2971	0.2242
Mayo año par - Abril año impar	2.2233	1.5712	-0.2096	-0.1187

Tabla A.6: Estimados de máxima verosimilitud de los parámetros ξ , κ , ν y λ del modelo Inversa Batschelet (base von Mises). Estos parámetros han dado el mejor ajuste para cada período de estudio.

Mes	Períodos							
	1: Mayo 2009 - Abril 2010		2: Mayo 2010 - Abril 2011		3: Mayo 2011 - Abril 2012		4: Mayo 2012 - Abril 2013	
	%	% acumulado	%	% acumulado	%	% acumulado	%	% acumulado
Mayo	1.2654	1.2654	1.4365	1.4365	0.6091	0.6091	2.5897	2.5897
Junio	1.6689	2.9343	3.8600	5.2965	0.4020	1.0111	4.7495	7.3393
Julio	3.1354	6.0697	9.5236	14.8201	0.4228	1.4339	9.3508	16.6900
Agosto	7.0214	13.0911	16.9167	31.7368	1.1533	2.5872	16.1467	32.8368
Septiembre	15.6471	28.7382	20.9287	52.6655	8.1298	10.7170	21.4156	54.2524
Octubre	28.9416	57.6798	21.8615	74.5270	26.3034	37.0204	21.4005	75.6529
Noviembre	21.3631	79.0429	15.8042	90.3312	28.4501	65.4705	12.8511	88.5040
Diciembre	10.6839	89.7268	6.7086	97.0398	18.2267	83.6972	5.5636	94.0676
Enero	5.0083	94.7351	1.5977	98.6374	9.1014	92.7986	2.1734	96.2410
Febrero	2.3956	97.1307	0.4511	99.0885	3.9844	96.7830	1.1312	97.3722
Marzo	1.6520	98.7827	0.3600	99.4485	2.1656	98.9486	1.1495	98.5217
Abril	1.2173	100.0000	0.5515	100.0000	1.0514	100.0000	1.4783	100.0000
Mes	Períodos							
	5: Mayo 2013 - Abril 2014		6: Mayo 2014 - Abril 2015		R1: Mayo Impar - Abril Par		R2: Mayo Par - Abril Impar	
	%	% acumulado	%	% acumulado	%	% acumulado	%	% acumulado
Mayo	2.1653	2.1653	3.8192	3.8192	1.1345	1.1345	2.4819	2.4819
Junio	3.2005	5.3658	6.6785	10.4977	1.0765	2.2111	4.8438	7.3257
Julio	5.9710	11.3368	11.2967	21.7944	1.6766	3.8876	9.7023	17.0280
Agosto	10.8912	22.2280	14.8721	36.6665	4.2903	8.1779	16.1416	33.1697
Septiembre	16.2835	38.5114	15.5760	52.2424	12.1879	20.3658	20.2918	53.4615
Octubre	19.8759	58.3873	15.5127	67.7551	25.7122	46.0780	20.8394	74.3009
Noviembre	16.6941	75.0814	12.3375	80.0926	24.1901	70.2681	13.9524	88.2533
Diciembre	11.3749	86.4563	8.3565	88.4491	14.5615	84.8296	6.2535	94.5069
Enero	6.2820	92.7382	4.6335	93.0826	7.5921	92.4217	2.1923	96.6992
Febrero	3.1714	95.9096	2.4791	95.5617	3.7228	96.1445	1.0055	97.7048
Marzo	2.2693	98.1788	2.1190	97.6807	2.3924	98.5369	0.9814	98.6862
Abril	1.8212	100.0000	2.3193	100.0000	1.4631	100.0000	1.3138	100.0000

Tabla A.7: Porcentajes mensuales de registrar al menos un caso de incidencia de dengue, calculados usando los modelos circulares de probabilidad ajustados en cada período de estudio.

Período	Año	Mes	Precipitación (mm)	% de incidencia
1	2009	Mayo	32.8	1.2654
	2009	Junio	111.1	1.669
	2009	Julio	84.6	3.1354
	2009	Agosto	124.6	7.0214
	2009	Septiembre	115.4	15.6471
	2009	Octubre	39.9	28.9416
	2009	Noviembre	132.3	21.3631
	2009	Diciembre	46.4	10.6839
	2010	Enero	26.4	5.0083
	2010	Febrero	14.3	2.3956
	2010	Marzo	11.4	1.6520
	2010	Abril	53.8	1.2173
2	2010	Mayo	96.6	1.43648
	2010	Junio	180	3.8600
	2010	Julio	228.7	9.5236
	2010	Agosto	219	16.9167
	2010	Septiembre	182.5	20.9287
	2010	Octubre	39.8	21.8615
	2010	Noviembre	16.8	15.8042
	2010	Diciembre	2.3	6.7087
	2011	Enero	48	1.5977
	2011	Febrero	34.4	0.4511
	2011	Marzo	24.6	0.36
	2011	Abril	4.8	0.55147
3	2011	Mayo	10.1	0.6091
	2011	Junio	223.7	0.4020
	2011	Julio	163.8	0.4228
	2011	Agosto	119.7	1.1533
	2011	Septiembre	189.8	8.1298
	2011	Octubre	100.7	26.3034
	2011	Noviembre	36.4	28.4501
	2011	Diciembre	16.8	18.2267
	2012	Enero	65.8	9.1014
	2012	Febrero	15.6	3.9844
	2012	Marzo	6.7	2.1656
	2012	Abril	112.4	1.0514
4	2012	Mayo	71	2.5897
	2012	Junio	168.8	4.7495
	2012	Julio	110.7	9.3508
	2012	Agosto	122.3	16.1467
	2012	Septiembre	141.4	21.4156
	2012	Octubre	100.6	21.4005
	2012	Noviembre	17.5	12.8511
	2012	Diciembre	15.4	5.5636
	2013	Enero	56.7	2.1734
	2013	Febrero	12.8	1.1312
	2013	Marzo	10.7	1.1495
	2013	Abril	37	1.4783
5	2013	Mayo	72.5	2.1653
	2013	Junio	216.9	3.2005
	2013	Julio	154.7	5.9710
	2013	Agosto	205.6	10.8912
	2013	Septiembre	270.1	16.2835
	2013	Octubre	177.1	19.8759
	2013	Noviembre	135.6	16.6941
	2013	Diciembre	69.1	11.3749
	2014	Enero	92.3	6.2820
	2014	Febrero	31	3.1714
	2014	Marzo	10.5	2.2693
	2014	Abril	19	1.8212
6	2014	Mayo	175.1	3.8192
	2014	Junio	83.6	6.6785
	2014	Julio	111.3	11.2967
	2014	Agosto	177	14.8721
	2014	Septiembre	193.9	15.5760
	2014	Octubre	203.6	15.5127
	2014	Noviembre	36	12.3375
	2014	Diciembre	14.8	8.3565
	2015	Enero	46.9	4.6335
	2015	Febrero	43.5	2.4791
2015	Marzo	32.5	2.1190	
2015	Abril	11.1	2.3193	

Tabla A.8: Comparación entre los niveles de precipitación y los porcentajes de posibilidad de registrar al menos un caso de dengue.

Apéndice B

Gráficos de comparación entre los períodos de estudio

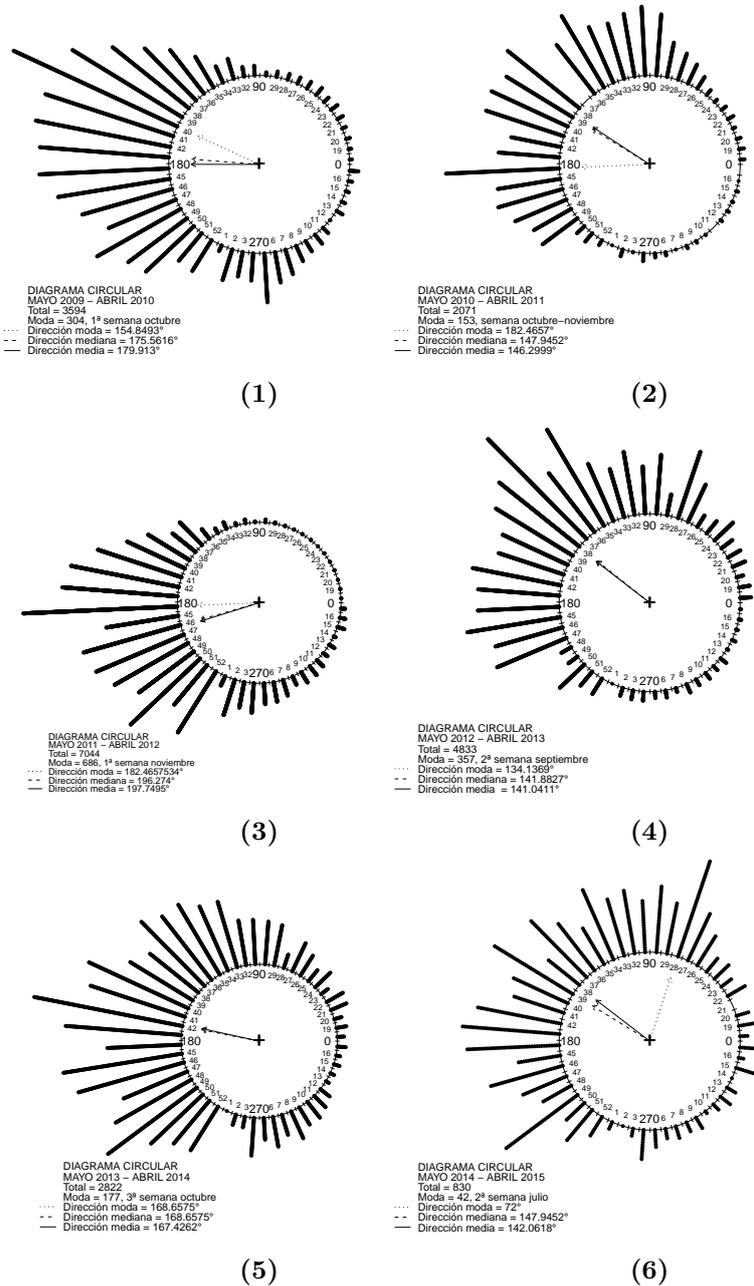


Figura B.1: Diagramas circulares que incluyen las direcciones moda, media y mediana muestrales de los 6 períodos de estudio mayo - abril anuales: (1) 2009 - 2010, (2) 2010 - 2011, (3) 2011 - 2012, (4) 2012 - 2013, (5) 2013 - 2014, (6) 2014 - 2015

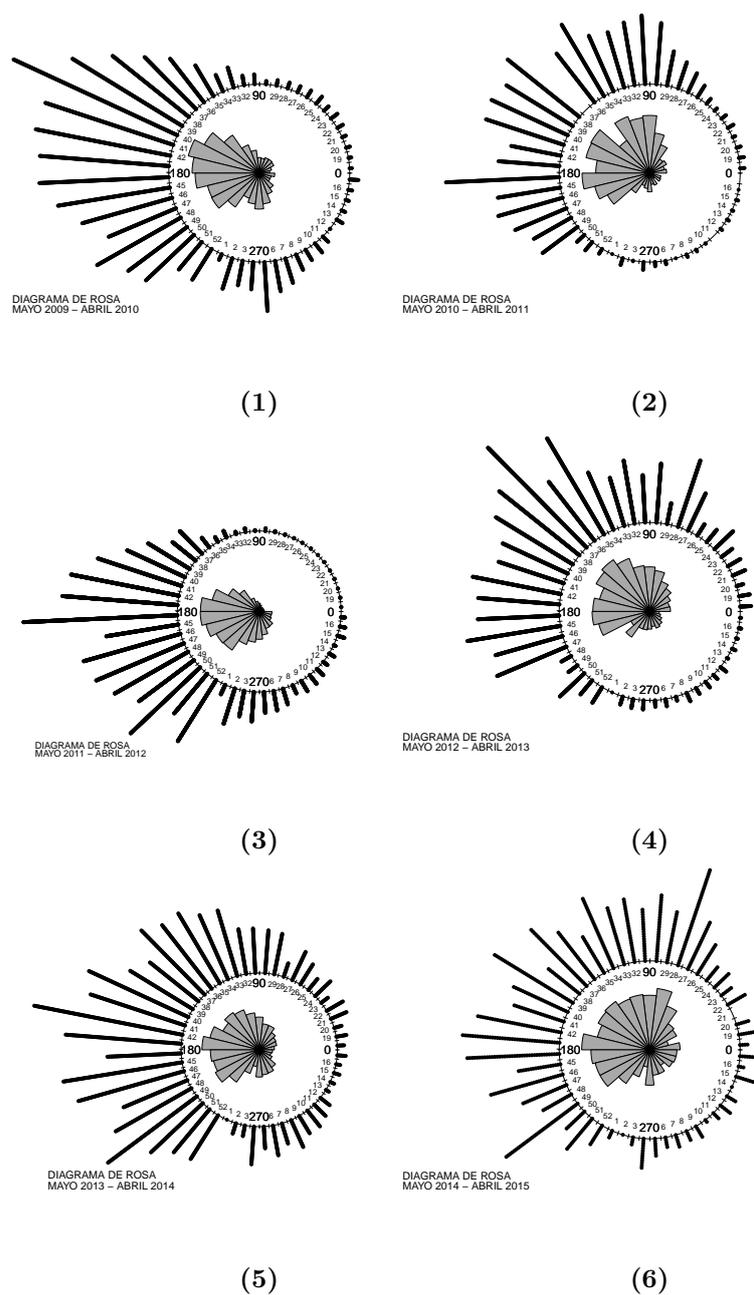


Figura B.2: Diagramas de rosa de los 6 períodos de estudio mayo - abril anuales: (1) 2009 - 2010, (2) 2010 - 2011, (3) 2011 - 2012, (4) 2012 - 2013, (5) 2013 - 2014, (6) 2014 - 2015

APÉNDICE B. GRÁFICOS DE COMPARACIÓN ENTRE LOS PERÍODOS DE ESTUDIO

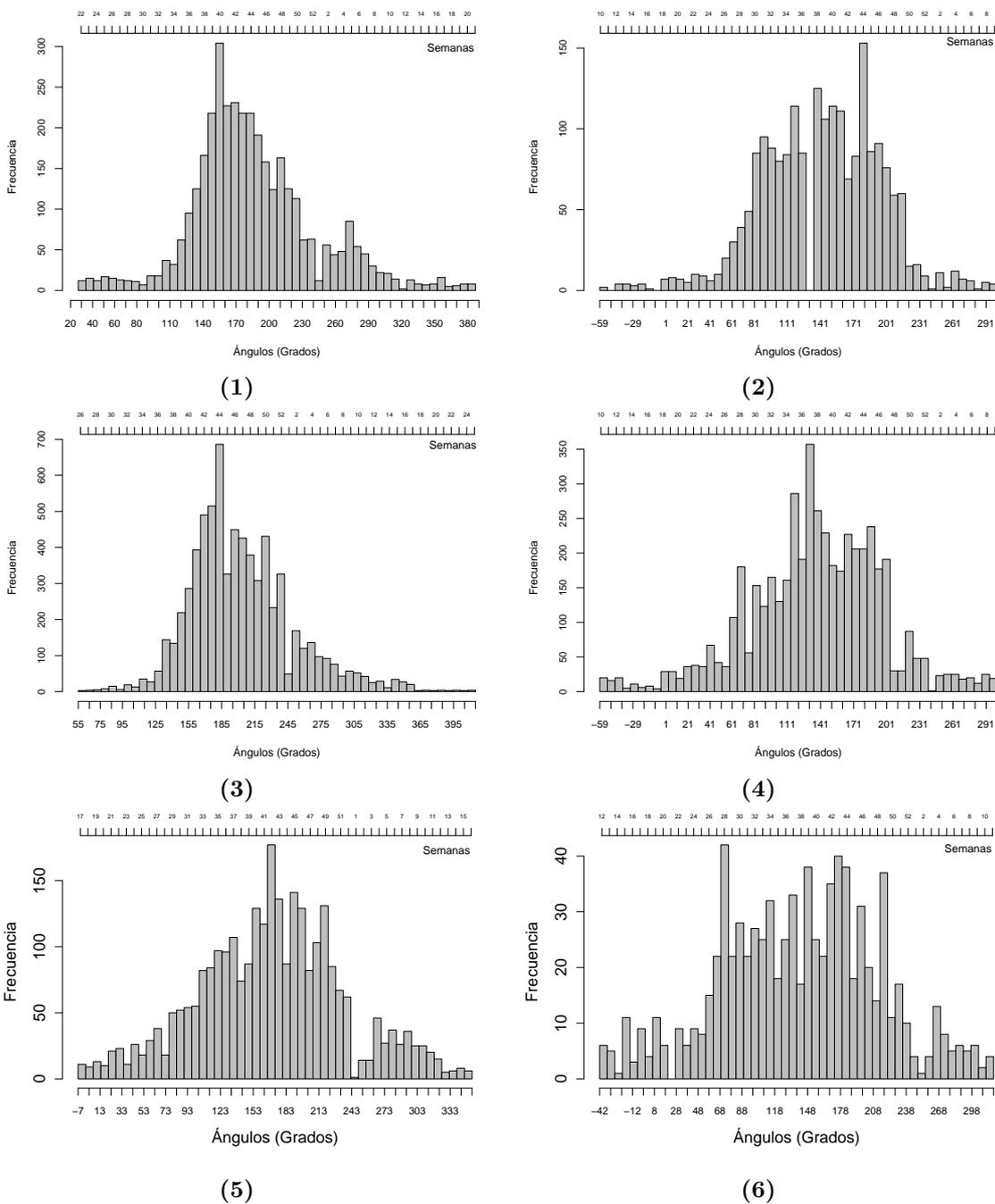


Figura B.3: Histogramas lineales de los 6 períodos de estudio mayo - abril anuales considerados: (1) 2009 - 2010, (2) 2010 - 2011, (3) 2011 - 2012, (4) 2012 - 2013, (5) 2013 - 2014, (6) 2014 - 2015.

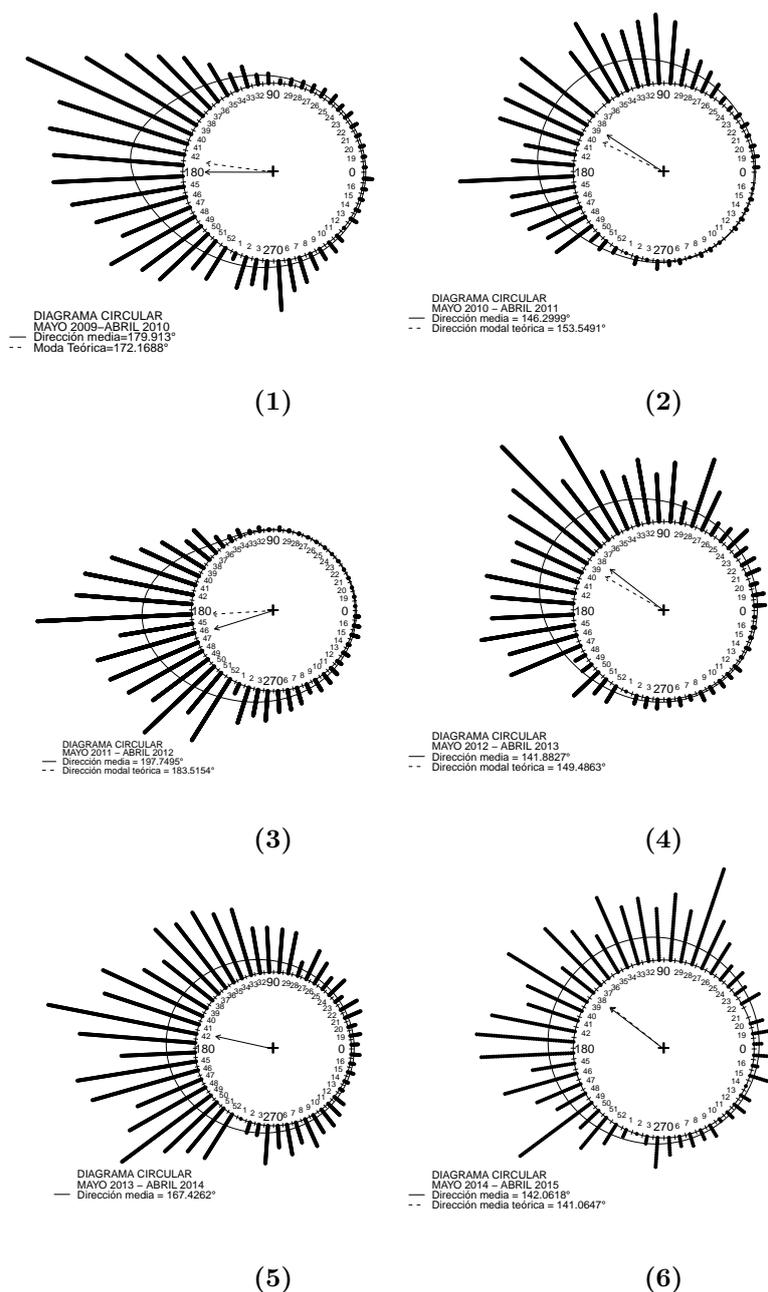


Figura B.4: Diagramas circulares que incluyen la forma polar de la distribución, la dirección media muestral y la dirección modal teórica de los 6 períodos de estudio mayo - abril anuales: (1) 2009 - 2010, (2) 2010 - 2011, (3) 2011 - 2012, (4) 2012 - 2013, (5) 2013 - 2014, (6) 2014 - 2015

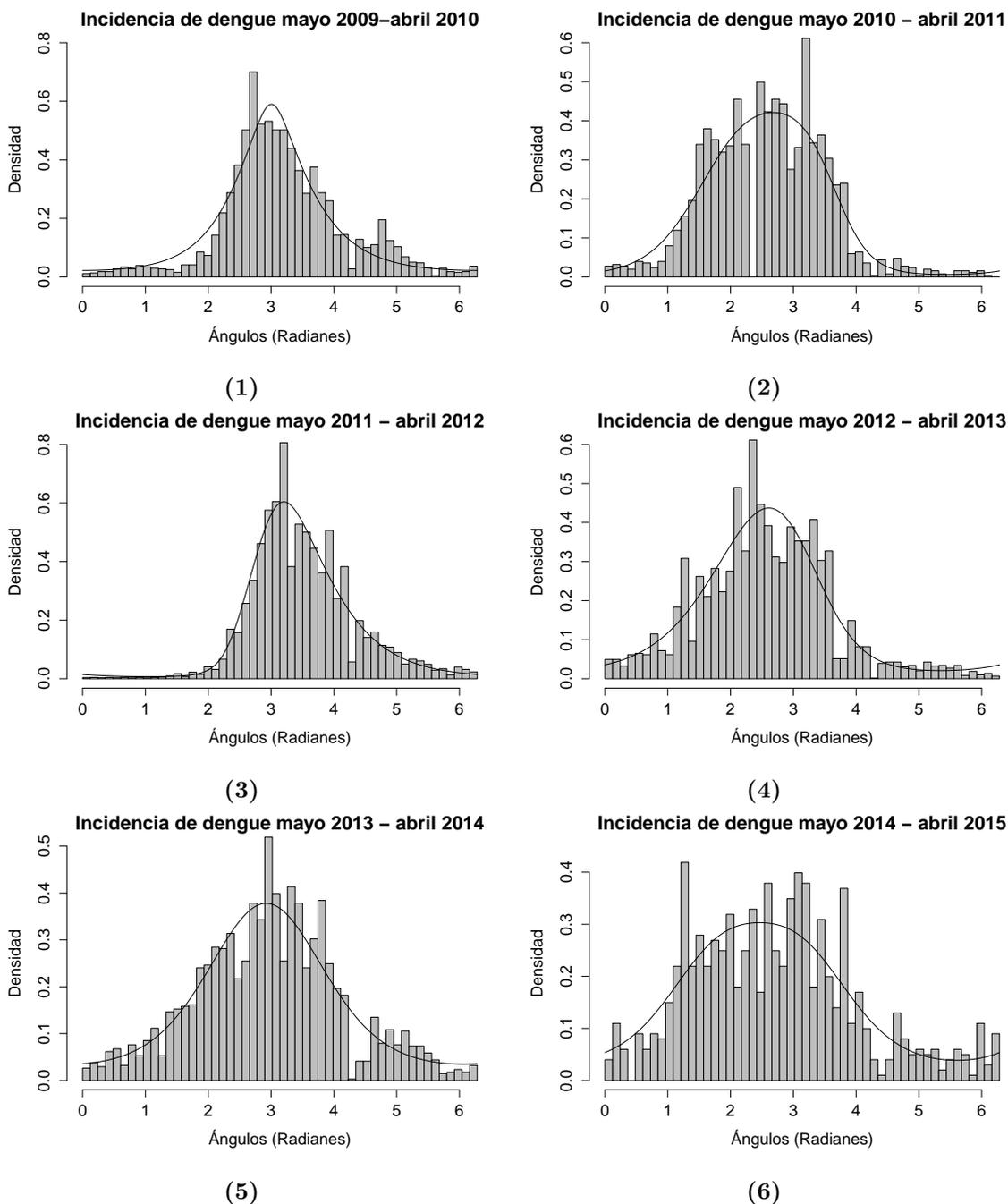


Figura B.5: Histogramas lineales con inicio en la dirección cero que incluyen las representaciones lineales de las distribuciones de los 6 períodos de estudio mayo - abril anuales: (1) 2009 - 2010, (2) 2010 - 2011, (3) 2011 - 2012, (4) 2012 - 2013, (5) 2013 - 2014, (6) 2014 - 2015.

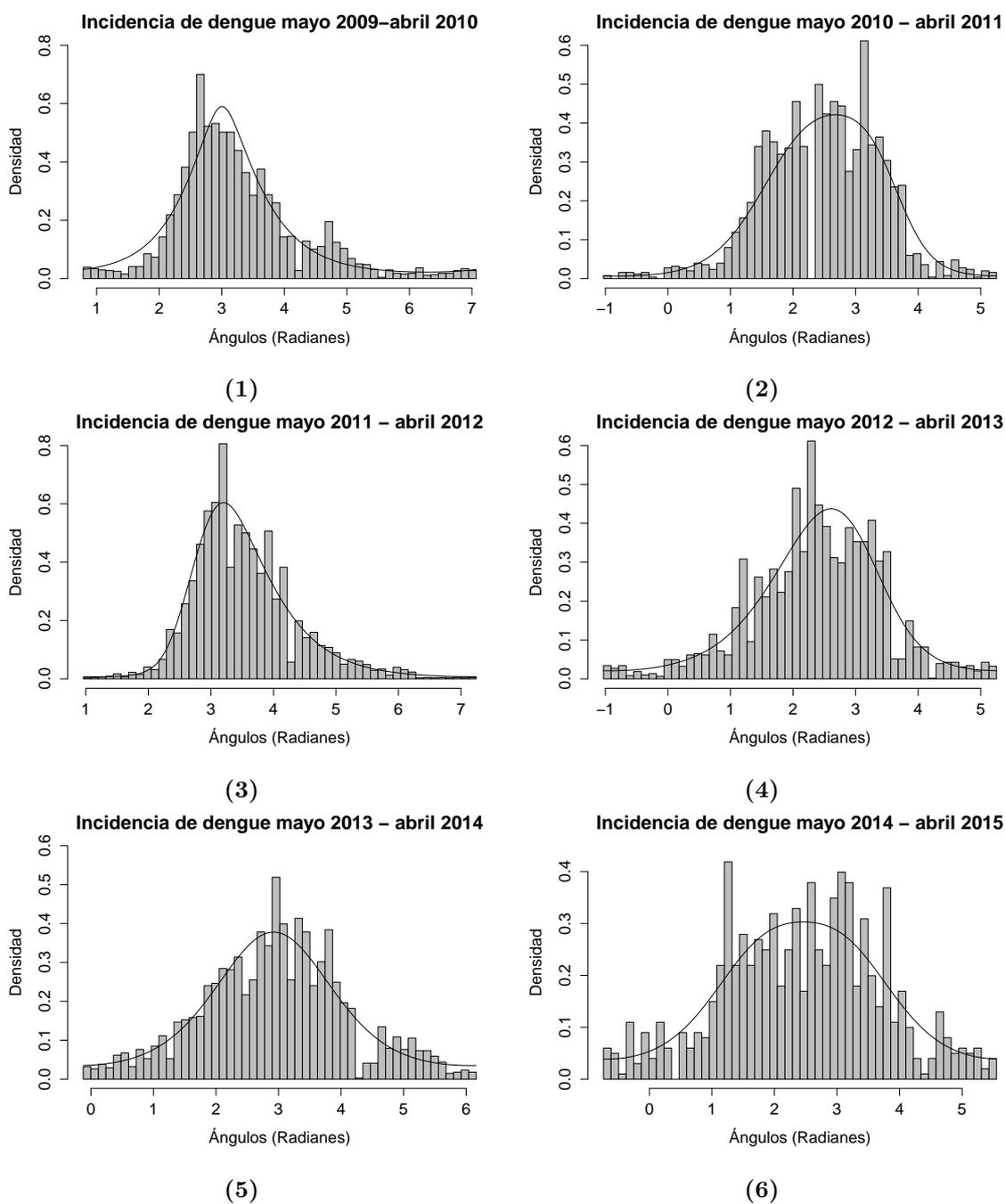


Figura B.6: Histogramas lineales con inicio en un punto conveniente que incluyen las representaciones lineales de las distribuciones de los 6 períodos de estudio mayo - abril anuales: (1) 2009 - 2010, (2) 2010 - 2011, (3) 2011 - 2012, (4) 2012 - 2013, (5) 2013 - 2014, (6) 2014 - 2015.

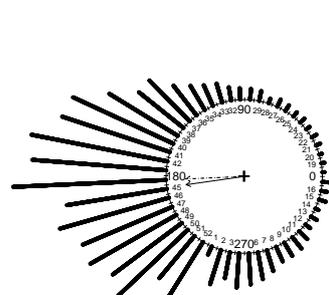


DIAGRAMA CIRCULAR RESUMEN
MAYO AÑO IMPAR - ABRIL AÑO PAR
Total = 13460
Moda = 291, semana octubre - noviembre
--- Dirección moda = 182.4657534°
- - Dirección mediana = 192.4656°
— Dirección media = 188.2392°

(1)

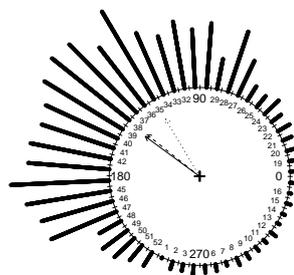


DIAGRAMA CIRCULAR RESUMEN
MAYO AÑO PAR - ABRIL AÑO IMPAR
Total = 7734
Moda = 418, semana agosto - septiembre
--- Dirección moda = 120.3288°
- - Dirección mediana = 141.04111°
— Dirección media = 143.1961°

(2)

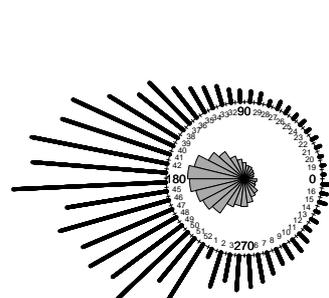


DIAGRAMA DE ROSA RESUMEN
MAYO AÑO IMPAR - ABRIL AÑO PAR

(3)

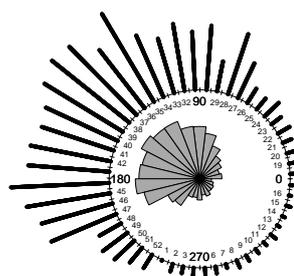
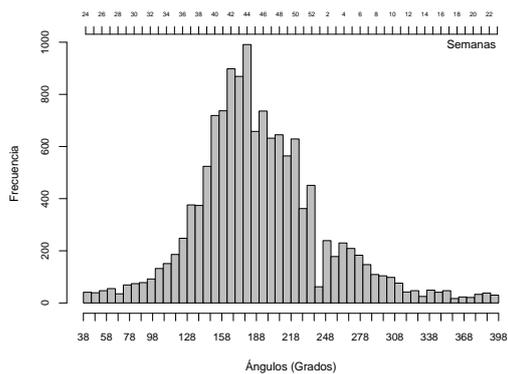
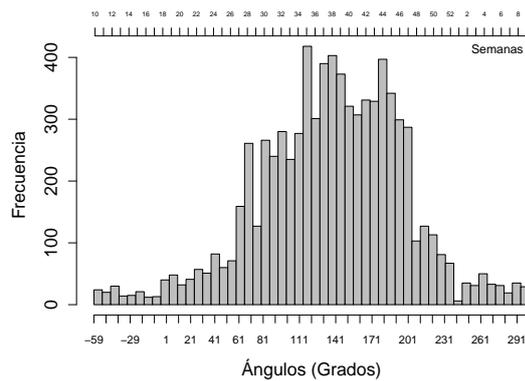


DIAGRAMA DE ROSA RESUMEN
MAYO AÑO PAR - ABRIL AÑO IMPAR

(4)

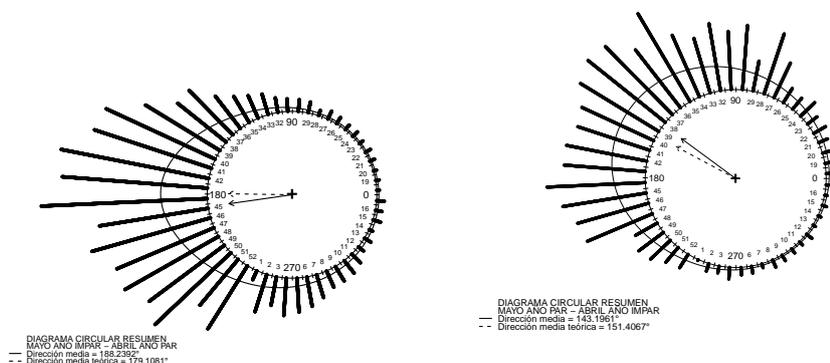


(5)



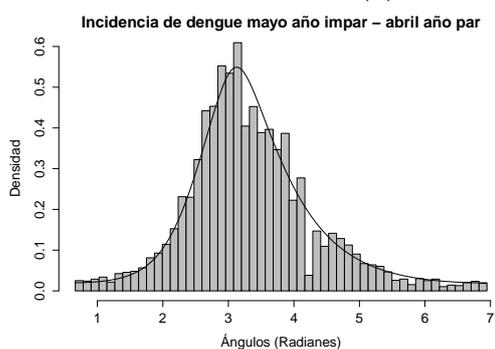
(6)

Figura B.7: Comparación entre los diagramas circulares, de rosa y los histogramas lineales del período resumen 1, a la izquierda, y el período resumen 2, a la derecha.

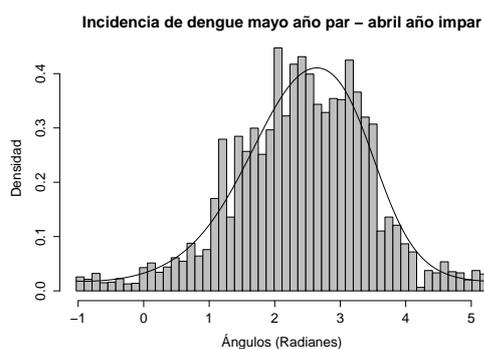


(1)

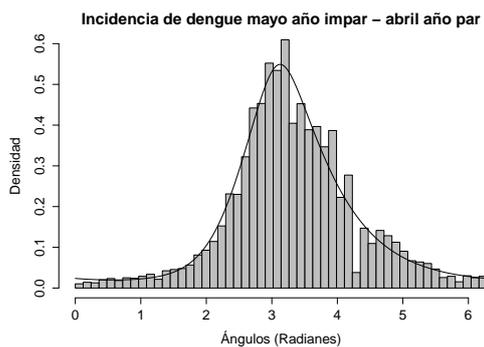
(2)



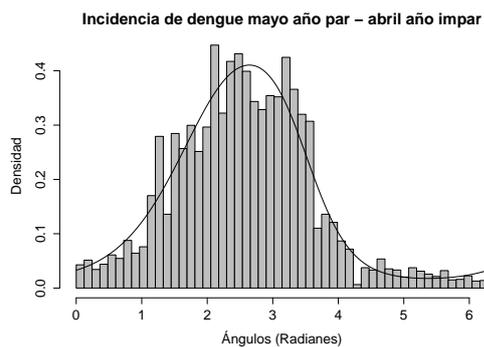
(3)



(4)



(5)



(6)

Figura B.8: Comparación entre las distribuciones en su forma polar y lineal (con histogramas cortados una en un punto ideal y la otra en cero) del período resumen 1, a la izquierda, y el período resumen 2, a la derecha.

APÉNDICE B. GRÁFICOS DE COMPARACIÓN ENTRE LOS PERÍODOS DE ESTUDIO

192

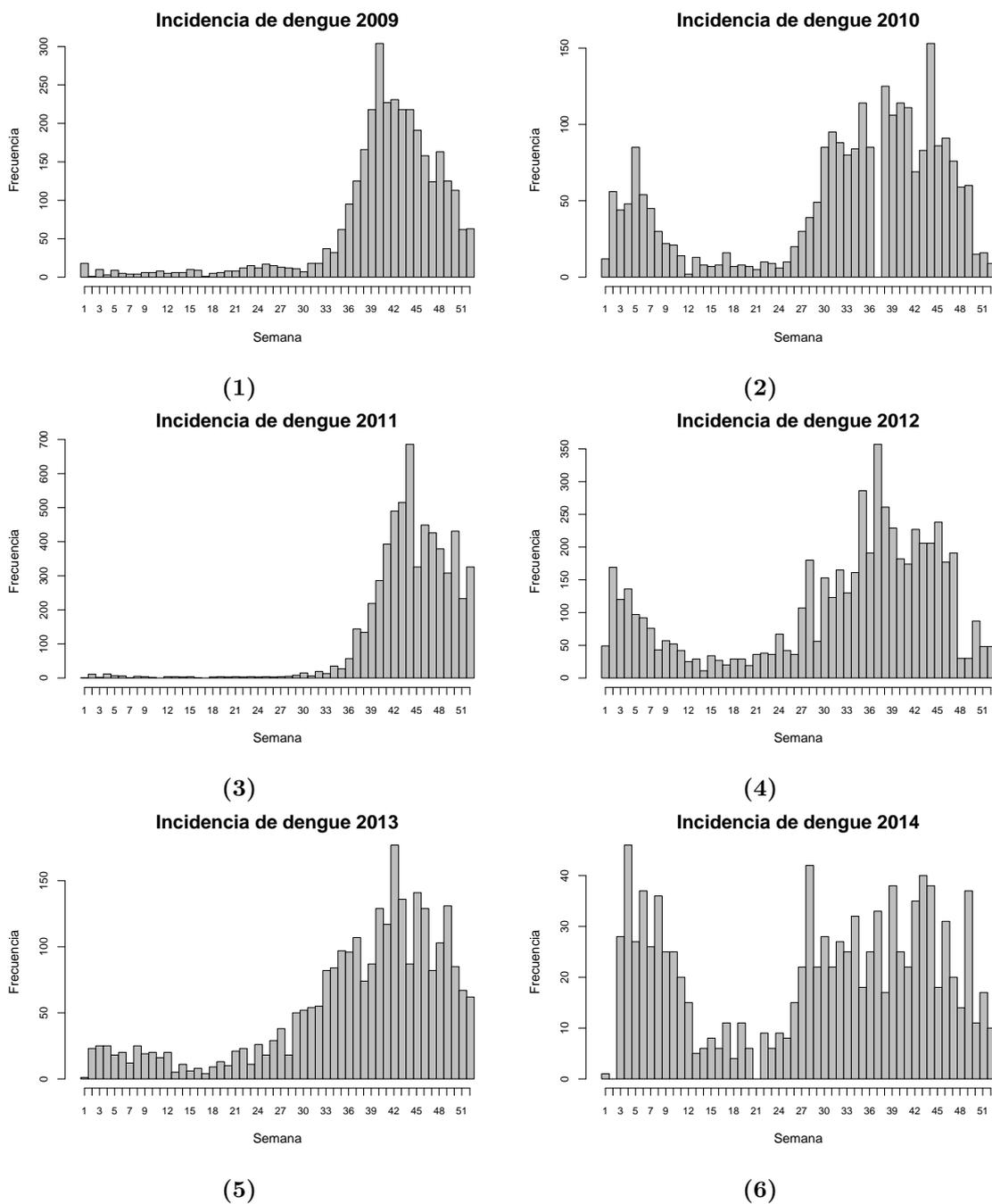


Figura B.9: Histogramas lineales anuales de casos de dengue semanales: (1) para 2009, (2) para 2010, (3) para 2011, (4) para 2012, (5) para 2013, (6) para 2014.

Apéndice C

Consideraciones con respecto a la base de datos

Sobre 2009

En el caso del año 2009, la semana epidemilógica 1 abarca del 3 al 9 de enero, y la semana 1 ubicada en el calendario gregoriano abarca de 1 al 7 de enero. A partir de aquí, se puede ver habrá un desfase hacia adelante de dos días, de una semana epidemilógica en este año con respecto a su semana análoga gregoriana. Todas las semanas epidemilógicas aquí tienen un lapso de 7 días.

Sobre 2010

El caso del año 2010, ocurre el mismo suceso que en el año anterior, o sea, la semana epidemiológica 1 abarca del 3 al 9 de diciembre, y por ende, se tiene un desfase de dos días hacia adelante de todas las demás semanas epidemiológicas.

Sobre 2011

En el año 2011 hay que tener en cuenta las siguientes consideraciones:

1. En las semanas epidemiológicas de la 1 hasta la 16, hay un desfase de un día hacia adelante con respecto a sus semanas análogas gregorianas; por ejemplo, la semana epidemilógica 1 abarca del 2 al 8 de enero, y la 16, abarca 17 al 23 de abril.

2. La semana epidemiológica 17 abarca diez días, del 24 abril al 3 de mayo.
3. Las semanas epidemiológicas de la 18 a la 25, tienen un desfase hacia adelante de 4 días.
4. La semana epidemiológica 26, del 30 de junio al 5 de julio, cuanta con 6 días.
5. Las semanas epidemiológicas de la 27 a la 49, tienen un desfase hacia adelante de 4 días.
6. La semana epidemiológica 50, del 14 al 22 de diciembre, tiene 9 días.
7. La semana epidemiológica 51, tiene un desfase hacia adelante de 6 días.
8. La semana epidemiológica 52 tiene 2 días, 30 y 31 de diciembre.

Con el fin de evitar un desfase drástico de las semanas epidemiológicas de la 17 a la 52, se han hecho las siguientes acciones:

1. Considerando que las semanas epidemiológicas de la 17 a la 27 forman parte del grupo de frecuencias bajas, pues el total de casos de dengue que estas semanas registran en conjunto es de 35 (de un total de 6 032), lo anterior permite redefinir éstas y modificar los casos de dengue ocurridos en ellas, con el fin de reducir el desfase en las semanas epidemiológicas con mucho mayor frecuencia. La semana epidemiológica 17, abarcará del 24 al 26 de abril (3 días) y se ha puesto tenga cero casos de dengue. Las semanas epidemiológicas posteriores, de la 18 a la 27, abarcarán 7 días (empezando con una del 27 de abril al 3 de mayo, y terminando con otra del 29 de junio al 5 julio). De acuerdo a la tendencia en los datos reales, a todas estas semanas se les ha puesto que los números 3 y 4, de forma alternada, sean los casos registrados en ellos, como una forma de compensar y equilibrar. Estas nuevas semanas epidemiológicas, con estas características, siguen sumando en total 35 casos.
2. Las semanas epidemiológicas de la 27 a la 50, se han renombrado y ahora pasan a ser las semanas epidemiológicas de la 28 a la 51.
3. Las semanas epidemiológicas 51 y 52, ahora formarán una sola semana epidemiológica, la 51. Aquí, se han sumado los casos de ambas semanas.

Con todas estas acciones se ha conseguido que el desfase de una semana epidemiológica en particular, respecto a su equivalente gregoriano, sea de a lo más tres días.

Sobre 2012

En el año 2012 hay que tener en cuenta las siguientes consideraciones:

1. Los primeros 8 informes semanales no tienen especificado el lapso de tiempo para el cual hace referencia la información que contienen. Estos reportes sólo se hacen llamar literalmente “semana epidemiológica 1”, “semana epidemiológica 2”, y así, hasta la “semana epidemiológica 8”.
2. La semana epidemiológica 9 presenta un desfase hacia atrás de 4 días con respecto a su equivalente gregoriano.
3. Las semanas epidemiológicas de la 10 a la 13, presentan todas desfases de 5 días hacia atrás.
4. La semana 14, del 28 de marzo al 2 de abril, solo tiene 6 días.
5. Las semanas epidemiológicas 15 y 16 presentan ambas desfases de 6 días hacia atrás.
6. Las semanas epidemiológicas 17, del 17 al 24 de abril, y 18, del 25 de abril al 1 de mayo, ambas cuentan con 8 días.
7. Las semanas epidemiológicas de la 19 a la 41, presentan todas un desfase de 5 días hacia atrás.
8. La semana 42, del 10 al 15 de octubre, solo tiene 6 días.
9. Las semanas epidemiológicas de la 43 a la 48, presentan todas un desfase de 6 días hacia atrás.
10. La semana epidemiológica 49, del 27 de noviembre al 10 de diciembre, tiene 14 días.
11. Las semanas epidemiológicas 50, 51 y 52, presentan todas un desfase de un día hacia adelante.

Con el fin de evitar un desfase drástico en la mayoría de las semanas epidemiológicas, se han tomado las siguientes acciones:

1. De acuerdo a la tendencia de valores semanales relativamente grandes que se traía desde finales de 2011, el informe de la semana epidemiológica 1, con cero casos reportados, no se ha tomado en cuenta en este estudio.
2. Las semanas epidemiológicas de la 2 a la 48, se han renombrado y ahora pasan a ser las semanas epidemiológicas de la 1 a la 47.

3. La semana epidemiológica 49, de catorce días y con 60 casos reportados, se ha dividido en dos y ha dado lugar a lo que serán las nuevas semanas epidemiológicas 48 y 49; la primera abarcará del 27 de noviembre al 3 de diciembre, y la segunda del 4 al 10 de diciembre. A ambas semanas se les ha asignado el mismo número de casos, es decir 30, que es la mitad de los casos reportados.

4. Las semanas epidemiológicas 50, 51 y 52 permanecen sin cambios.

Con todas estas acciones se ha conseguido que el desfase de una semana epidemiológica en particular, respecto a su equivalente gregoriano, sea de a lo más tres días (en la mayoría de los casos será de dos).

Sobre 2013

Todas las semanas epidemiológicas de este año presentan el mismo lapso de tiempo que sus equivalentes gregorianos. Excepto la semana epidemiológica 52, la cual sólo tiene 7 días y abarca del 24 al 30 de diciembre, es decir, no incluye al día 31.

Sobre 2014

La semana epidemiológica 1 abarca del 1 al 6 de enero, es decir, sólo cuenta con 6 días. Esto provoca que todas las demás semanas del año estén desfasadas un día, hacia atrás, de su equivalente gregoriano. La semana epidemiológica 52 abarca del 23 al 29 de diciembre. Sin embargo, para este año se incluye una semana “epidemiológica 53”, que abarca del 30 de diciembre de 2014 al 5 de enero de 2015, y que solo presenta 4 casos reportados de dengue.

Sobre 2015

Para este año, sólo han servido los informes hasta la semana epidemiológica 16.

La semana epidemiológica 1 abarca hasta el 12 de enero. Lo anterior hace que todas las demás semanas epidemiológicas estén desfasadas 5 días hacia adelante respecto de sus análogos gregorianos.

Para evitar un desfase drástico, se han tomado las siguientes acciones:

1. La semana epidemiológica 53 del 2014 (con 4 casos reportados), que contiene los últimos 2 días de dicho año, y los 5 primeros días de enero de 2015, al

tener más días de este último año, se ha tomado como la nueva semana epidemiológica 1.

2. La semana epidemiológica 2 se ha redefinido, como la que abarca del 6 al 12 de enero, y se ha impuesto tenga 1 caso de dengue.
3. Las semanas epidemiológicas de 2 al 16, se han renombrado y ahora pasan a ser las semanas epidemiológicas del 3 al 17.

Con todas estas acciones se ha conseguido que el desfase de una semana epidemiológica, respecto a su análogo gregoriano, sea de dos días.

Referencias

- Agostinelli, C. (2007). Robust estimation for circular data. *Computational Statistics & Data Analysis*, 51(12): 5867–5875.
- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6): 716–723.
- Åström, C., Rocklöv, J., Hales, S., Béguin, A., Louis, V., y Sauerborn, R. (2012). Potential distribution of dengue fever under scenarios of climate change and economic development. *EcoHealth*, 9(4): 448–454.
- Batschelet, E. (1981). *Circular statistics in biology*. Academic press London.
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., *et al.* (2013). The global distribution and burden of dengue. *Nature*, 496(7446): 504–507.
- Cartwright, D. (1963). The use of directional spectra in studying the output of a wave recorder on a moving ship. *Ocean wave spectra*, pages 203–218.
- Elith, J., Leathwick, J. R., y Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 802–813.
- Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge University Press, Cambridge.
- Fisher, N. I., Lewis, T., y Embleton, B. J. (1987). *Statistical analysis of spherical data*. Cambridge University Press, Cambridge.
- Fisher, R. (1953). Dispersion on a sphere. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 217, pages 295–305. The Royal Society.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4): 367–378.

- Gómez, H. y San Martín, J. (2011). Estrategia mesoamericana para la prevención y control integrado del dengue plan dengue. *Monografía en internet*. Instituto Nacional de Salud Pública, México, Organización Panamericana de la Salud.[Acceso 25 de marzo de 2012]. Disponible en: <http://www.iadb.org/wmsfiles/products/SM2015/Documents/Spanish/Plan-dengue.pdf>. [Links].
- Góngora-Biachi, R. A. (2004). La erradicación de la fiebre amarilla en mérida, yucatán: una historia de tenacidad y éxito. *Rev Biomed*, 15: 251–258.
- Gordon, A., Jupp, P., y Byrne, R. (1989). The construction and assessment of mental maps. *British Journal of Mathematical and Statistical Psychology*, 42(2): 169–182.
- Gradshteyn, I. S. y Ryzhik, I. (2007). Table of integrals, series, and products. translated from the russian. translation edited and with a preface by alan jeffrey and daniel zwillinger.
- Hales, S., De Wet, N., Maindonald, J., y Woodward, A. (2002). Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *The Lancet*, 360(9336): 830–834.
- Jaimes Jaimes, P. N. y Pineda Garavito, S. J. (2012). *Una introducción al análisis de datos circulares con algunas aplicaciones*. Universidad Industrial de Santander, Facultad de Ciencias, Escuela de Matemáticas.
- Jammalamadaka, S. R., Bhadra, N., Chaturvedi, D., Kutty, T., Majumdar, P., y Poduval, G. (1986). Functional assessment of knee and ankle during level walking. *Data Analysis in the Life Sciences*, pages 21–54.
- Jammalamadaka, S. R. y Sengupta, A. (2001). *Topics in circular statistics*. World Scientific, Singapore.
- Jeffreys, H. (1948). *Theory of Probability*. Clarendon Press, Oxford.
- Jones, M. y Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100(472): 1422–1428.
- Jones, M. y Pewsey, A. (2012). Inverse Batschelet Distributions for Circular Data. *Biometrics*, 68(1): 183–193.
- Jones, M. C. (2010). Distributions Generated by Transformation of Scale Using an Extended Cauchy-Schlömilch Transformation. *Sankhyā: The Indian Journal of Statistics*, 72-A, Part 2: pp. 359–575.
- Mardia, K. V. (1972). *Statistics of directional data*. Academic Press, London.

- Mardia, K. V. y Jupp, P. E. (1999). *Directional statistics*. John Wiley & Sons, London.
- Mena, M. (2004). *Aplicaciones de estadística circular a problemas de ciencias naturales*. Akadia, Buenos Aires.
- Muñoz Vicente, M. D. (2008). *Aportaciones al Estudio de la Anisotropía y Modelado Espacial de Información*. Universidad de Salamanca, Departamento de Informática y Automática.
- Nelder, J. A. y Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4): 308–313.
- Nuñez Antonio, G. (2010). *Análisis Bayesiano de Modelos Lineales para Datos Direccionales considerando la Distribución Normal bajo Proyección*. Universidad Autónoma Metropolitana, UAM-Iztapalapa, División de Ciencias Básicas e Ingeniería.
- Pewsey, A. (2002). Testing circular symmetry. *The Canadian Journal of Statistics*, 30(4): 591–600.
- Pewsey, A. (2004). The large-sample joint distribution of key circular statistics. *Metrika*, (60): 25–32.
- Pewsey, A., Neuhauser, M., y Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press, Great Britain.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. John Wiley & Sons, United States of America.
- Ross, H. E., Crickmar, S., Sills, N., y Owen, E. (1969). Orientation to the vertical in free divers. *Aerospace medicine*, 40(7): 728–732.
- RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- Schmidt-Koenig, K. (1965). Current problems in bird orientation. *Advances in the Study of Behavior*, 1: 217–278.
- Schwarz, G. *et al.* (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2): 461–464.

- Sengupta, S. y Rao, J. (1966). Statistical analysis of cross-bedding azimuths from the kamthi formation around bheemaram, pranhita: Godavari valley. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 165–174.
- Stuart, A. y Ord, J. (1987). Kendall's advanced theory of statistics, vol 1, charles griffin & co. *Ltd, London*.
- Subsistema de Notificación Semanal de Casos Nuevos de Enfermedades. (2015). *Información Epidemiológica de Morbilidad. Reporte Correspondiente a Marzo de 2015*.
- Von Mises, R. (1918). Über die "ganzzahligkeit" der atomgewichte und verwandte fragen. *Phys. z*, 19: 490–500.
- Watson, G. (1980). *Statistics on Spheres*. Wiley, Michigan.