

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS

CARRERA PROFESIONAL DE MATEMÁTICAS



ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS EN DATOS CIRCULARES

TESIS PARA OPTAR AL TÍTULO PROFESIONAL DE LICENCIADA
EN MATEMÁTICAS, MENCIÓN ESTADÍSTICA

PRESENTADA POR

BACHILLER : OLGA MILUSKA MAMANI MUÑOZ
ASESOR : Mgt. EDGAR CENTENO HUAMANI

AUSPICIO DEL CONCEJO DE INVESTIGACIÓN

CUSCO-PERÚ

2013

DEDICATORIA

Dedico el trabajo de tesis y mi Título Profesional

A mis padres Justiniano Mamani y Felicitas Muñoz

Que con su amor incondicional y ejemplo, me demostraron

El significado real de la vida; a mis hermanos Odilón y Julia

Por su comprensión, en momentos difíciles y afecto desde el

Inicio de mis estudios. Hasta concluir.

A mi hijo, Leonardo Gabriel, y una dedicatoria

Muy especial a Jehová Dios.

Olga Miluska.

AGRADECIMIENTO

Quiero expresar mi agradecimiento,

A mi asesora de tesis Dra. Katia García Alfaro,

Por su generosidad y predisposición al brindarme

La oportunidad de recurrir a su capacidad y experiencia

En un marco de confianza, afecto y amistad, fundamentales

Para la concreción de este trabajo.

A Mgt. Domingo Walter Kehuarucho C. quien me

Brindo todas las facilidades para la aplicación,

de mi trabajo.

A todos los amigos, amigas en especial

A Janet Romero y Jean Paul Latorre que me

Brindaron su apoyo incondicional.

Agradezco infinitamente a Dios

Por brindarme el camino correcto,

Por darme la inspiración, valor y

Pasión en Todo momento para

Luchar y Alcanzar mi objetivo.

Olga Miluska

PRESENTACIÓN

Sr. Decano de la Facultad de Ciencias Químicas, Físicas y Matemáticas.

Sr. Coordinador de la carrera profesional de matemáticas.

Sres. Miembros del Jurado.

Conforme a las disposiciones vigentes de grados y títulos de nuestra facultad, pongo a vuestra disposición el presente trabajo de tesis intitulado "ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS EN DATOS CIRCULARES".

En este trabajo de investigación se presenta la estimación de parámetros y pruebas de hipótesis en datos circulares, información muy poco conocida.

En el primer capítulo se desarrolla el planteamiento metodológico del trabajo de investigación, en el segundo capítulo se desarrolla el marco teórico para datos lineales, en el tercer y cuarto capítulos se desarrolla toda la teoría correspondiente a la estadística descriptiva e inferencial de datos circulares y por último en el quinto capítulo se explica mediante aplicaciones de prácticas la teoría de datos circulares.

Atentamente:

Br. OLGA MILUSKA MAMANI MUÑOZ

ÍNDICE

CAPÍTULO I

1.1 PLANTEAMIENTO DEL PROBLEMA.....	1
1.2 FORMULACIÓN DEL PROBLEMA.....	2
1.3 JUSTIFICACIÓN DEL PROBLEMA.....	2
1.4 DELIMITACIÓN DEL PROBLEMA.....	2
1.5 OBJETIVOS	
1.5.1 Objetivo General.....	3
1.5.2 Objetivos Específicos.....	3
1.6 HIPÓTESIS.....	3
1.7 LIMITACIONES.....	3
1.8 METODOLOGIA.....	4
1.8.1 Tipo de investigación.....	4
1.8.2 Nivel y diseño de investigación.....	4

CAPÍTULO II

ANÁLISIS DESCRIPTIVO DE DATOS CIRCULARES

2.1 ANTECEDENTES DEL PROBLEMA.....	5
2.2 DATOS CIRCULARES.....	6
2.3 REPRESENTACIÓN GRÁFICA DE DATOS CIRCULARES.....	8
2.4 MEDIDAS DE POSICIÓN.....	10
2.4.1 DIRECCIÓN MEDIA CIRCULAR.....	10
2.4.2 MEDIANA CIRCULAR.....	13
2.4.3 DIRECCIÓN MODAL.....	14
2.5 MEDIDAS DE DISPERSION EN LA MUESTRA.....	14
2.5.1 VARIANZA DE LA MUESTRA.....	14
2.5.2 DESVIACIÓN ESTÁNDAR CIRCULAR.....	15
2.6 DISPERSIÓN CIRCULAR DE UNA MUESTRA Y ERROR ESTÁNDAR CIRCULAR.....	16

2.7	DISTRIBUCIONES CIRCULARES.....	16
2.7.1	DISTRIBUCIÓN UNIFORME.....	17
2.7.2	DISTRIBUCIÓN NORMAL ESPIRAL.....	18
2.7.3	DISTRIBUCIÓN VON MISES.....	18
2.7.4	DISTRIBUCIÓN CARDIOIDE.....	20
2.7.5	DISTRIBUCIÓN WRAPPED.....	20

CAPÍTULO III

ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS EN DATOS CIRCULARES

3.1	INTRODUCCIÓN.....	22
3.2	VERIFICACIÓN DE UNIFORMIDAD.....	22
3.3	ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS PARA UNA MUESTRA DE DATOS CON DISTRIBUCIÓN VON MISES.....	23
3.3.1	PRUEBA DE BONDAD DE AJUSTE.....	23
3.3.2	PRUEBA PARA OUTLIERS.....	24
3.3.3	ESTIMACIÓN DE MÁXIMA VEROSIMILITUD DE LOS PARÁMETROS μ Y K	25
3.3.4	ESTIMACIÓN POR INTERVALOS DE LOS PARÁMETROS $\hat{\mu}$ Y \hat{k}	28
3.3.5	PRUEBAS DE HIPÓTESIS SOBRE μ Y k	30
3.4	ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS DE DOS O MÁS MUESTRAS DE LA DISTRIBUCIÓN VON MISES.....	32
3.5	PRUEBAS DE IGUALDAD DE DIRECCIONES MEDIAS.....	32
3.5.1	SI LOS PARÁMETROS DE CONCENTRACIÓN NO SON TODOS IGUALES.....	33
3.5.2	SI LOS PARÁMETROS DE CONCENTRACIÓN SON IGUALES..	33
3.6	INFERENCIAS DESPUES DE LA COMPARACIÓN DE LAS DIRECCIONES MEDIAS.....	37
3.6.1	ESTIMACIÓN DE LA DIRECCIÓN MEDIA COMÚN DE DOS O	

MÁS DISTRIBUCIONES DE VON MISES.....	37
3.6.2 CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE DOS DIRECCIONES MEDIAS.....	39
3.7 PRUEBA DE IGUALDAD DE LOS PARÁMETROS DE CONCENTRACIÓN DE DOS O MÁS DISTRIBUCIONES DE VON MISES.....	41
3.8 ESTIMACIÓN DEL PARÁMETRO DE CONCENTRACIÓN COMÚN DE DOS O MÁS DISTRIBUCIONES DE VON MISES.....	42

CAPÍTULO V

APLICACIÓN DE DATOS CIRCULARES

4.1 CASO: DIRECCIÓN DEL VIENTO EN LA CIUDAD DEL CUSCO.....	44
4.2 SELECCIÓN DE LA MUESTRA.....	46
4.3 REPRESENTACIÓN GRÁFICA Y ANÁLISIS DESCRIPTIVO.....	46
4.4 INTERVALO DE CONFIANZA (Bootstrap).....	51
4.5 PRUEBA DE HIPÓTESIS PARA LA IGUALDAD DE PARÁMETROS DE CONCENTRACIÓN DE DISTRIBUCIONES DE VON MISES.....	52
4.6 PRUEBA DE HIPÓTESIS PARA LA IGUALDAD DE LAS DIRECCIONES MEDIAS DE DISTRIBUCIONES DE VON MISES.....	56
CONCLUSIONES.....	58
RECOMENDACIONES.....	59
BIBLIOGRAFIA.....	60
ANEXO 1.....	64
ANEXO 2.....	65
ANEXO 3.....	66
ANEXO 4.....	67
ANEXO 5.....	69
ANEXO 6.....	71

INTRODUCCIÓN

La estadística como una ciencia que está en constante crecimiento tiene una infinidad de aplicaciones en todos los campos de la ciencia, es decir es una ciencia que utiliza el análisis de datos para la toma de decisiones a todo nivel. Pero no todos los datos pueden ser analizados con las técnicas estadísticas clásicas, existen datos que por su naturaleza misma como los “datos circulares” o “datos direccionales” necesitan de otro tratamiento, debido a las limitaciones de la estadística clásica. La motivación del desarrollo del presente trabajo surge principalmente por la escasa información existente en la literatura para tratamiento de datos circulares. En el presente trabajo se sistematiza la teoría para el análisis estadístico de datos circulares referidos a las medidas de resumen como la dirección media, dirección mediana, dirección modal, varianza, etc., del mismo se presenta la teoría necesaria para la estimación de parámetros y las pruebas de hipótesis para datos circulares. La teoría desarrollada se aplica a algunos ejemplos cuyos datos son de carácter eminentemente circular. En la primera y segunda aplicación se generó datos circulares con la librería **CircStats** del software estadístico R, en la tercera aplicación se utilizó datos reales sobre la dirección del viento de la ciudad del Cusco.

CAPÍTULO I

1.1. PLANTEAMIENTO DEL PROBLEMA

La estadística es una de las herramientas más ampliamente utilizadas en la investigación científica y en la toma de decisiones. Su aplicación en instituciones gubernamentales, educativas, en los negocios, en la industria, en la banca y otros quehaceres diarios hacen que sea una herramienta indispensable.

En la estadística descriptiva se desarrolla toda una teoría para la obtención, organización, presentación y descripción de datos, mientras que en la estadística inferencial obtenemos generalizaciones a partir de una muestra, dividiéndose primero en un problema de estimación, donde se determina el valor de un parámetro o parámetros en base a un grupo de opciones y luego en un problema de prueba de hipótesis donde se decide aceptar o rechazar un conjunto de valores específicos de parámetros, todo ello con una validez para datos lineales.

Pero en los diferentes campos de la ciencia como la biología, geología, medicina, ecología, meteorología y otros, existen datos direccionales, como las direcciones de vuelo de aves, movimientos direccionales de animales, direcciones del viento, etc., a los cuales se les conoce como datos circulares. Este tipo de datos tienen características que lo hacen diferentes a los datos de naturaleza lineal con los cuales estamos más familiarizados. Esta diferencia de los datos circulares con los datos lineales hizo que en las últimas décadas se desarrolle una teoría estadística diferente a lo que conocemos, porque la estadística clásica tiene limitaciones cuando estudiamos datos direccionales.

1.2. FORMULACION DEL PROBLEMA

¿Cuál es la metodología para la estimación de parámetros y pruebas de hipótesis sobre la dirección media (μ) y concentración (κ) de la distribución von Mises en datos circulares?

1.3. JUSTIFICACIÓN DEL PROBLEMA

El presente trabajo de investigación tiene por finalidad sistematizar los procedimientos estadísticos para el análisis de datos circulares, debido a la escasa información existente sobre el tema. La información sobre el análisis de datos lineales es muy abundante, desde los primeros semestres en la universidad, inclusive a nivel de post grado, todos, estudiantes, profesionales e investigadores son capacitados en técnicas estadísticas descriptivas e inferenciales para datos lineales, y no así en técnicas estadísticas para datos circulares. Posiblemente esto se deba a que la mayoría de las aplicaciones existentes en el mundo utiliza datos lineales en su análisis. Pero nosotros como profesionales de la estadística también debemos pensar en ese grupo de personas que necesitan manejar las técnicas estadísticas para el análisis de datos circulares.

El presente trabajo será de gran utilidad tanto a estudiantes, investigadores y profesionales afines a la biología, meteorología, ecología, etc., debido a que gracias a este pequeño aporte, tendrán un acceso a la teoría sistematizada de la estimación y pruebas de hipótesis de datos circulares, y no tendrán las dificultades que usualmente uno encuentra en la búsqueda de información al respecto.

1.4. DELIMITACIÓN DEL PROBLEMA

El presente trabajo de investigación es de tipo descriptivo y la teoría de la estimación de parámetros y prueba de hipótesis para datos circulares es muy amplia, por lo que el presente trabajo se limita sólo al estudio de la teoría de estimación y prueba de hipótesis para datos circulares que tienen distribución von Mises, esto, porque esta distribución es la más utilizada en datos circulares, el cual es análoga a la distribución normal para datos lineales. Cabe

mencionar que se analizan los parámetros de la dirección media μ y la de concentración κ , de von Mises.

1.5. OBJETIVOS

1.5.1. Objetivo General:

Desarrollar los procedimientos para la estimación de parámetros y pruebas de hipótesis sobre medidas de posición (μ) y concentración (κ) de la distribución von Mises en datos circulares.

1.5.2 Objetivos Específicos:

- ✓ Sistematizar los procedimientos para la estimación de parámetros y pruebas de hipótesis de datos circulares que se distribuyen como una von Mises.
- ✓ Utilizar las diferentes rutinas computacionales del software R para el análisis de los datos de direcciones del viento de la ciudad del cusco, año 2005.

1.6. HIPÓTESIS

La metodología existente para la estimación y prueba de hipótesis de datos circulares cuya distribución es von Mises, es la que mejor se adecúa para el estudio de los datos circulares, en particular para el estudio de las direcciones del viento de la ciudad de Cusco del año 2005.

1.7. LIMITACIONES

En el presente trabajo de investigación se presentaron las siguientes limitaciones:

- ✓ La poca información existente de bibliografía en castellano, hizo que se busque información en inglés y en otros idiomas presentándose dificultades en la traducción, el cual no fue muy sencillo.
- ✓ La no existencia de profesionales en la región Cusco que dominen el tema y por consiguiente la imposibilidad de realizar consultas al respecto.

1.8. METODOLOGIA

1.8.1 Tipo de investigación

En el presente trabajo de investigación se encuentra dentro de la investigación básica debido a que nuestro propósito es organizar y sistematizar un conocimiento vigente dentro de la ciencia estadística. Una vez sistematizado este conocimiento realizamos la aplicación respectiva al estudio de las direcciones del viento de la ciudad del Cusco, durante el año 2005.

1.8.2 Nivel y diseño de investigación

El objetivo del presente trabajo de investigación fue desarrollar y sistematizar la teoría estadística para el análisis de datos circulares y su correspondiente aplicación en el estudio de las direcciones del viento de la ciudad del Cusco, por consiguiente el presente trabajo tiene un nivel de investigación del tipo descriptivo. En la aplicación a datos reales sobre la dirección del viento de la ciudad del Cusco; el respectivo análisis se realizó mediante las librerías *VecStatGraphs2D* y *Bootstrap* de *R 3.0.0*. del software estadístico R; por todo lo anterior se considera que el método de investigación utilizado tiene un enfoque no experimental.

CAPÍTULO II

MARCO TEÓRICO

2.1. ANTECEDENTES DEL PROBLEMA

Para este trabajo se ha tomado los siguientes antecedentes:

1.- PEDRO NEL JAIMES JAIMES Y SILVIA JOHANNA PINEDA GARAVITO (2012) en su trabajo “**Una introducción al análisis de datos circulares con algunas aplicaciones**”, afirma que las técnicas de análisis de los datos circulares son de gran importancia en diferentes ciencias. Así mismo, se pudo comprobar que el modelamiento de este tipo de datos no depende del origen y el sentido de rotación, como se pudo ver con los datos de dengue, donde la dirección era en el Oriente y el sentido de rotación era contrario a las manecillas de reloj, en cambio, en las asistencias a urgencias la dirección era el Norte y el sentido de rotación era en forma horaria, es decir, en sentido a las manecillas del reloj.

2.- RAFAEL IZBICKI (2008) en su trabajo “**Análisis de datos circulares**”, señala que las nuevas técnicas de análisis estadístico, así como los nuevos modelos probabilísticos, son necesarios con el propósito de analizar datos circulares, además indica que estas técnicas presentan una dificultad mayor que las técnicas lineales.

3.- GLADYS DOROTEA CACSIRE BARRIGA (1997) en su trabajo “**Inferencia sobre medidas de posición e dispersión en datos circulares**”, describe de forma clara y correcta el análisis de una muestra de datos circulares, así también como la estimación y comparación de medidas de dispersión y posición en experimentos completamente aleatorizados con un factor fijo.

4.- GABRIEL NUÑEZ ANTONIO (2010) en su trabajo “**Análisis Bayesiano de Modelos Lineales para Datos Direccionales considerando la Distribución Normal bajo Proyección**”, propone una metodología Bayesiana para realizar inferencias para todos los parámetros involucrados en un modelo lineal y concluye que los modelos basados en una distribución normal bivariada bajo proyección se pueden extender de manera más natural al estudio de datos direccionales en dimensiones mayores.

5.- MERCEDES FERNÁNDEZ SAU (2012) en su trabajo “**Estimadores de mínima distancia para datos circulares**” obtuvo estimadores de mínima distancia para datos circulares. Los estimadores de mínima distancia ocupan un lugar destacado entre las alternativas robustas al método clásico de estimación puntual por máxima verosimilitud.

2.2. DATOS CIRCULARES

En diferentes aplicaciones donde se utilizan datos es usual trabajar con datos lineales, pero, en muchos casos se observan datos en forma de direcciones o tiempos de ocurrencias, este tipo de datos se denomina datos direccionales. Para trabajar con este tipo de datos se necesita construir nuevos estadísticos puesto que los estadísticos clásicos para datos lineales son inapropiados, debido a que no toman en cuenta la periodicidad de esta clase de datos.

Los datos circulares se obtienen de diferentes maneras, por ejemplo con la brújula podemos medir las direcciones del viento y las direcciones migratorias de las aves, mientras que con un reloj podemos medir el tiempo de llegada de

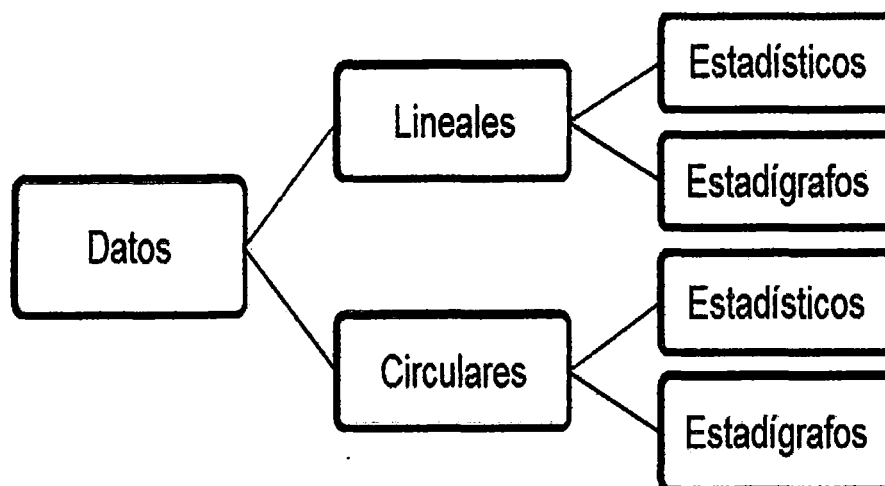
los pacientes a un servicio de un hospital o el número de veces en un periodo de tiempo de un evento.

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

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

En diversos campos surgen problemas estadísticos donde la medida de los datos no es escalar sino angular, dando una orientación en el plano (datos circulares).

Los datos circulares se encuentran presentes en muchos estudios científicos y requieren para su análisis el uso de herramientas y técnicas estadísticas especiales, que no son ampliamente conocidas, ni apreciadas debido a la falta de familiaridad con ellas.

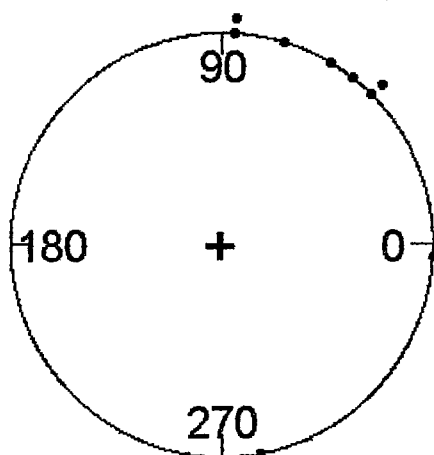


2.3. REPRESENTACIÓN GRÁFICA DE DATOS CIRCULARES.

1) DIAGRAMA DE DATOS CIRCULARES EN RAMAS

Una forma de representar las características de los datos circulares es por puntos en una circunferencia del círculo unitario, un punto a cada observación, de modo que se observa un diagrama de dispersión de los datos en la circunferencia de un círculo. Este gráfico provee información importante acerca de características de los datos tales como la tendencia central, dispersión o concentración. En el ejemplo de la ruleta descrito por Mardía (1972) se gráfica las posiciones de parada ver figura 2.1.

FIGURA 2.1
DIAGRAMA DE DATOS CIRCULARES EN RAMAS

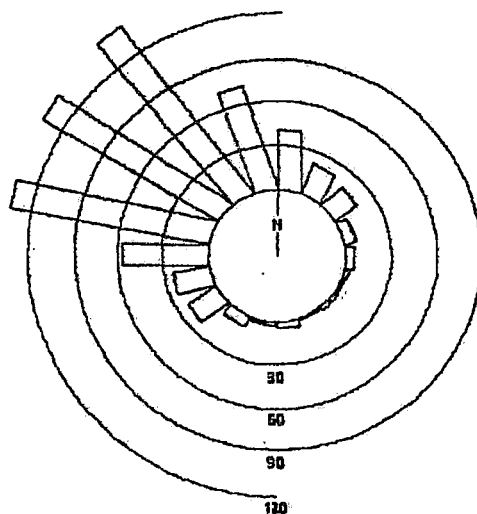


2) HISTOGRAMAS CIRCULARES

Si se tiene una gran cantidad de datos circulares, éstos se muestran mejor en forma de histograma donde los círculos concéntricos caracterizan los atributos de frecuencia de los datos. En este caso los datos pueden ser agrupados adoptando el mismo procedimiento que en los datos lineales. El rango de 0° a 360° se divide en un cierto número de intervalos de clase, y encima de ellos se dibujan rectángulos, donde su longitud y el área representan las frecuencias de las medidas observadas. La elección de los límites y la longitud de los intervalos de clase requieren la misma consideración que los datos lineales, y no puede haber superposición de intervalos. En la Figura 2.2

se muestra un ejemplo de la dirección del vuelo de pato salvaje descrito en Mardia (1972) [20].

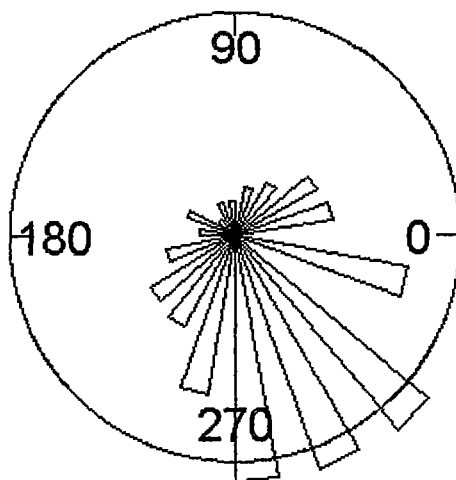
FIGURA 2.2
HISTOGRAMA DE DATOS CIRCULARES



3) DIAGRAMA DE ROSA

Otra representación es el diagrama de rosa, donde se hace corresponder a cada intervalo un sector con un vértice en el origen, radio proporcional a la frecuencia de clase y arco subtendido al intervalo de clase, descrito en Mardia (1972) [20].

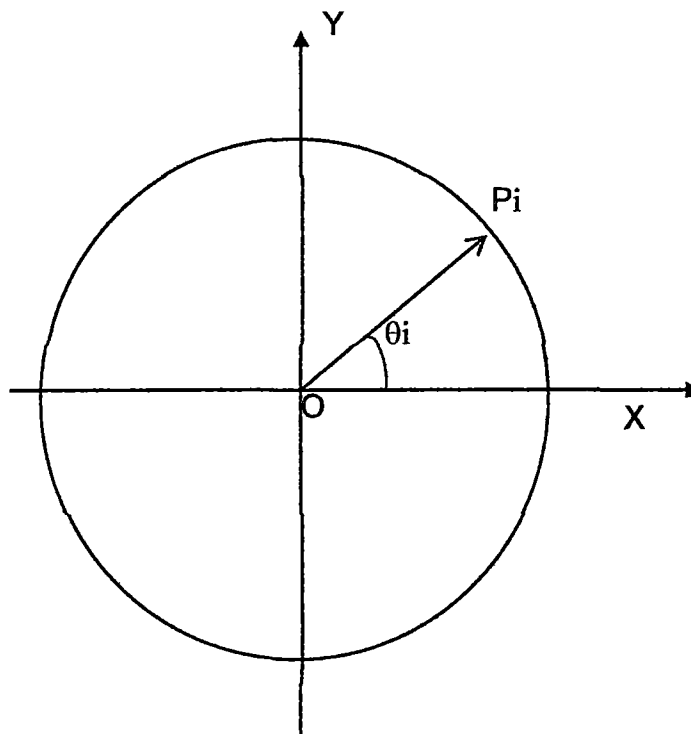
FIGURA 2.3
DIAGRAMA DE ROSA DE DATOS CIRCULARES



2.4. MEDIDAS DE POSICIÓN

Sea P_i el punto correspondiente al ángulo θ_i en la circunferencia centrada en el origen O . A cada punto P_i le corresponde un vector unitario \overline{OP}_i , $i = 1, \dots, n$, donde $P_i = (\cos\theta_i, \text{sen}\theta_i)$.

FIGURA 2.4
REPRESENTACIÓN GRÁFICA DEL i -ésimo PUNTO MUESTRAL θ_i

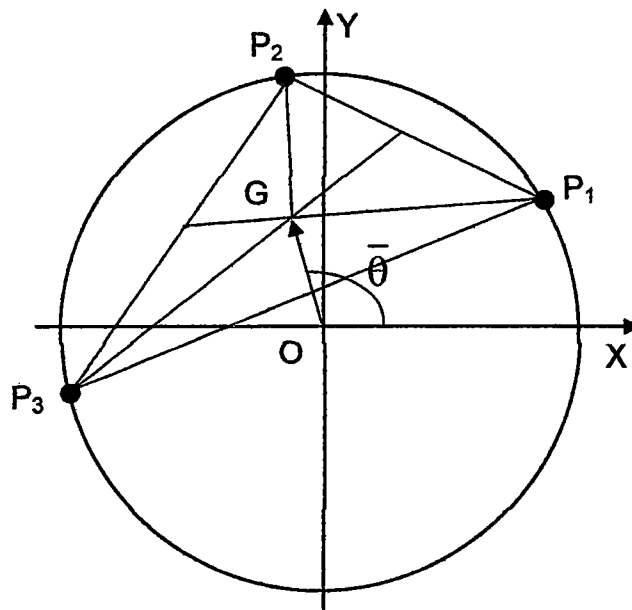


2.4.1. DIRECCIÓN MEDIA CIRCULAR

La definición de la dirección media se define como:

Si a cada punto asociamos una masa de igual valor M , entonces obtenemos el centro de masa G , también llamado centro de gravedad. Si este centro es diferente del origen O , entonces el vector que apunta para el centro de gravedad \overline{OG} es definido como el vector media muestral y determina una dirección llamada dirección media muestral o ángulo medio.

FIGURA 2.5
REPRESENTACIÓN GRÁFICA DEL CENTRO DE GRAVEDAD DE UNA MUESTRA.



Consideremos dos formas de determinación del vector medio: uno que utiliza el álgebra vectorial y la otra que utiliza funciones trigonométricas.

1) VECTOR MEDIO UTILIZANDO EL ÁLGEBRA VECTORIAL

Sean $\overrightarrow{OP_1}, \overrightarrow{OP_2}, \dots, \overrightarrow{OP_n}$, los vectores unitarios correspondientes a la muestra $\theta_1, \theta_2, \dots, \theta_n$. Si asignamos una masa M a los puntos P_1, P_2, \dots, P_n , se tiene que el vector:

$$\overrightarrow{OG} = \frac{1}{n} \sum_{i=1}^n \overrightarrow{OP_i}$$

apunta hacia el centro de masa de P_1, P_2, \dots, P_n , entonces \overrightarrow{OG} es el vector medio de la muestra.

Si denotamos por R a la longitud del vector resultante,

$$R = \left\| \sum_{i=1}^n \overrightarrow{OP_i} \right\|$$

entonces, r es la longitud del vector medio:

$$r = \|\overline{OG}\| = \frac{R}{n}$$

Observamos que el centro de masa G se localiza, en general, en el interior de la circunferencia de radio unitario. Éste será un punto de la circunferencia sólo en el caso especial de que todos los ángulos de una muestra sean iguales, o sea, las masas concurren en un único punto. Al considerar vectores de longitud unitario, se tiene:

$$0 \leq R \leq n \quad \text{y} \quad 0 \leq r \leq 1$$

2) VECTOR MEDIO UTILIZANDO FUNCIONES TRIGONOMÉTRICAS

Denotemos por (\bar{x}, \bar{y}) las coordenadas cartesianas del centro de masa G de P_1, \dots, P_n ; $P_i = (x_i, y_i)$, $i = 1, 2, \dots, n$.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \text{y} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

También podemos escribir \bar{x} y \bar{y} , por:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i = \bar{C} \quad \text{y} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n \text{sen} \theta_i = \bar{S}. \quad (2.1)$$

Entonces, las coordenadas cartesianas del centro de masa G pueden ser escritas como (\bar{C}, \bar{S}) , donde r es la longitud del vector medio \overline{OG} ,

$$r = [\bar{C}^2 + \bar{S}^2]^{\frac{1}{2}} = \frac{1}{n} \left[\left(\sum_{i=1}^n \cos \theta_i \right)^2 + \left(\sum_{i=1}^n \text{sen} \theta_i \right)^2 \right]^{\frac{1}{2}}.$$

Luego:

$$R = \left[\left(\sum_{i=1}^n \cos \theta_i \right)^2 + \left(\sum_{i=1}^n \text{sen} \theta_i \right)^2 \right]^{\frac{1}{2}}.$$

De la definición de dirección media y de (2.1), la media muestral $\bar{\theta}$ se define por:

$$\bar{\theta} = \begin{cases} \arctag\left(\frac{\bar{S}}{\bar{C}}\right) & \text{si } \bar{S} \geq 0, \bar{C} > 0; \\ 180^\circ + \arctag\left(\frac{\bar{S}}{\bar{C}}\right) & \text{si } \bar{C} < 0; \\ 360^\circ + \arctag\left(\frac{\bar{S}}{\bar{C}}\right) & \text{si } \bar{S} < 0, \bar{C} > 0. \end{cases} \quad (2.2)$$

Para situaciones particulares se tiene:

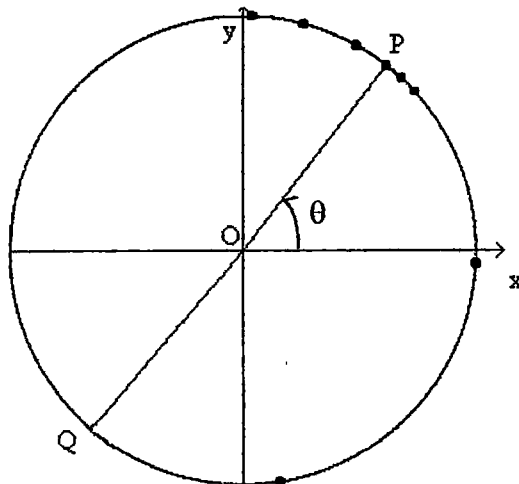
$$\bar{\theta} = \begin{cases} 90^\circ & , \text{ si } \bar{S} > 0, \bar{C} = 0 \\ 270^\circ & , \text{ si } \bar{C} = 0, \bar{S} < 0 \\ \text{no determinado} & , \text{ si } \bar{S} = 0, \bar{C} = 0 \end{cases}$$

2.4.2. MEDIANA CIRCULAR

Teniendo en cuenta la representación de la muestra $\theta_1, \dots, \theta_n$, en el círculo unitario de centro O. Un segmento OP, donde P es un punto de la circunferencia de tal manera que la mitad de los puntos muestrales están a ambos lados del diámetro PQ y la mayoría de los puntos muestrales que son más próximos a P que a Q define la dirección mediana de la muestra.

Si el tamaño de la muestra n es impar, P es uno de los puntos de la muestra; cuando n es par, P está entre dos puntos muestrales.

FIGURA 2.6
REPRESENTACIÓN DE LA DIRECCIÓN MEDIANA



La mediana se puede determinar también mediante la minimización de la expresión:

$$d(\theta) = \pi - \frac{1}{n} \sum_{i=1}^n |\pi - |\theta_i - \theta|| \quad (2.3)$$

2.4.3 DIRECCIÓN MODAL

La dirección modal $\check{\theta}$ en la muestra es la dirección de concentración máxima de los datos. Su determinación es análoga al caso de datos lineales, es decir el dato que más se repite.

2.5. MEDIDAS DE DISPERSIÓN EN LA MUESTRA

2.5.1 VARIANZA DE LA MUESTRA

En la Figura 2.7 se muestra cuatro muestras de tamaño 5 con el mismo ángulo medio. Se observó que a medida que aumenta la dispersión de los puntos, disminuye la longitud del vector de la media (r). Para $0 \leq r \leq 1$, observamos que los valores extremos de r tienen propiedades interesantes: $r = 1$ implica que todos los puntos son coincidentes $r = 0$ implica una dispersión uniforme por el círculo. Así para datos unimodales la longitud del vector medio puede ser considerada como una medida de concentración alrededor del ángulo medio.

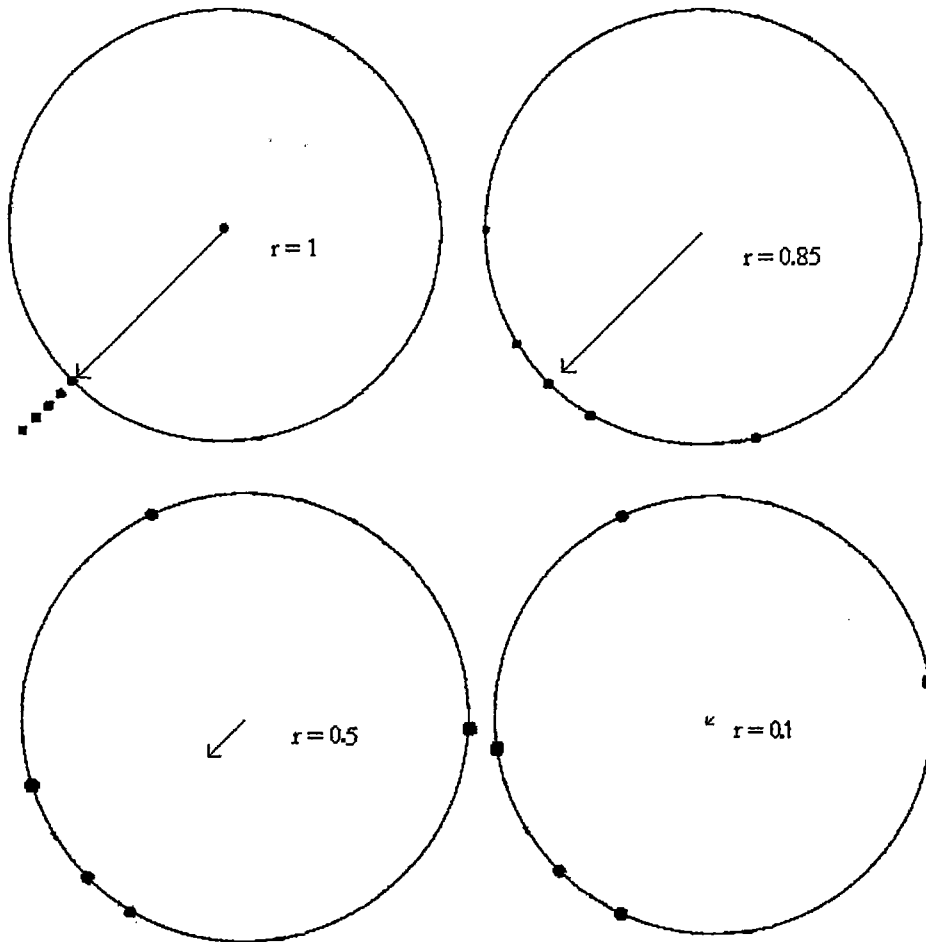
La varianza de la muestra circular se define entonces como:

$$V = 1 - r. \quad (2.4)$$

Tenga en cuenta que $0 \leq V \leq 1$, el cual no ocurre para datos lineales. Cuanto más bajo sea el valor de V , es más homogénea la muestra.

FIGURA 2.7

VECTOR MEDIA DE CUATRO MUESTRAS DE DATOS CIRCULARES.



2.5.2. DESVIACIÓN ESTÁNDAR CIRCULAR

La desviación estándar circular de la muestra es definido como:

$$v = [-2\ln(1 - v)]^{1/2} \quad (2.5)$$

donde V es la varianza circular. Tenga en cuenta que mientras que $0 \leq V \leq 1$, v asume valores en el intervalo $[0, \infty)$. Entiéndase que la desviación estándar circular no se define como la raíz cuadrada de la varianza circular.

Una aproximación para v es dada por: $v \approx (2V)^{1/2}$, pues $V \approx \ln(1 - V)$ para valores pequeños de V .

El error en esta aproximación es menos de 5% para $V < 0.18$, es decir $r > 0.82$, como comenta Fisher (1993).

2.6 DISPERSIÓN CIRCULAR DE UNA MUESTRA Y ERROR ESTÁNDAR CIRCULAR

Teniendo en cuenta el primero y segundo momento trigonométrico centrado en la media,

$$r = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) \quad y \quad (2.6)$$

$$r_2 = \left[\left(\frac{1}{n} \sum_{i=1}^n \cos 2(\theta_i - \bar{\theta}) \right)^2 + \left(\frac{1}{n} \sum_{i=1}^n \text{sen} 2(\theta_i - \bar{\theta}) \right)^2 \right]^{1/2}$$

Es posible definir otra medida de dispersión denominada dispersión circular, dada por:

$$\hat{\delta} = \frac{1 - r_2}{2r^2}, \quad (2.7)$$

Utilizando la dispersión circular de una muestra, definimos el error estándar circular como:

$$\hat{\sigma} = \left(\frac{\hat{\delta}}{n} \right)^{1/2}. \quad (2.8)$$

2.7. DISTRIBUCIONES CIRCULARES

Describimos algunos modelos de probabilidades continuas utilizadas con mayor frecuencia en el análisis de datos circulares: la distribución circular uniforme, la normal espiral y la distribución de von Mises. Denotaremos por $f(\theta)$ la función de densidad de probabilidad de la variable aleatoria circular continua Θ , el cual cumple las condiciones:

- i) $f(\theta) \geq 0, \quad -\infty < \theta < \infty$
- ii) $f(\theta + 2\pi) = f(\theta), \quad -\infty < \theta < \infty$

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

La función de distribución $F(\theta)$ que correspondiente a $f(\theta)$ se puede definir en cualquier intervalo (θ_1, θ_2) , como:

$$F(\theta_2) - F(\theta_1) = \int_{\theta_1}^{\theta_2} f(\theta) d\theta. \quad (2.9)$$

Sin embargo, es conveniente definir un punto de partida para la distribución:

$$F(\theta) = \int_0^{\theta} f(\varphi) d\varphi. \quad (2.10)$$

En particular se cumple que: $F(2\pi) = 1$

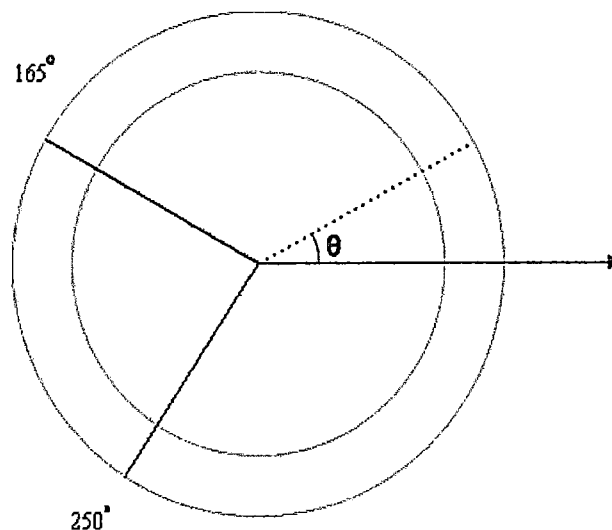
2.7.1. DISTRIBUCIÓN UNIFORME

En este modelo todas las direcciones entre 0 y 2π se eligen con igual probabilidad, en otras palabras, no existe dirección preferencial.

La densidad de probabilidad es constante sobre la circunferencia de radio unitario, es decir:

$$f(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta \leq 2\pi \quad (2.11)$$

FIGURA 2.8
REPRESENTACIÓN GRÁFICA DE LA DISTRIBUCIÓN UNIFORME
CIRCULAR



Su función de distribución está dada por

$$F(\theta) = \frac{\theta}{2\pi}, \quad 0 \leq \theta \leq 2\pi. \quad (2.12)$$

La dirección media μ no es definida. La longitud del vector medio ρ es cero. Por consiguiente, tenemos:

$$V = 1, \quad (2.13)$$

2.7.2. DISTRIBUCIÓN NORMAL ESPIRAL

Es una distribución simétrica y unimodal obtenido envolviendo el círculo de radio unitario por una distribución normal. Esa función de densidad es construida de la siguiente forma: X una variable aleatoria en la recta con función de densidad $g(x)$ dada por $N(\mu, \sigma^2)$. Podemos definir una variable circular Θ como:

$$\Theta = X \pmod{2\pi}$$

La función densidad de probabilidad $f(\theta)$ de Θ es dada por:

$$\begin{aligned} f(\theta) &= \sum_{k=-\infty}^{\infty} g(\theta + 2k\pi) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{1}{2} \frac{((\theta - \mu) + 2k\pi)^2}{\sigma^2}\right), \quad 0 < \theta \leq 2\pi. \end{aligned} \quad (2.15)$$

Mardia (1972) demostró que la expresión anterior puede ser escrita como:

$$f(\theta) = \frac{1}{2\pi} \left[1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu) \right], \quad 0 \leq \theta \leq 2\pi, \quad 0 \leq \rho \leq 1, \quad (2.16)$$

donde μ es la dirección media y ρ es la longitud del vector medio.

2.7.3. DISTRIBUCIÓN VON MISES

La distribución Von Mises $VM(\mu, \kappa)$, es una de las distribuciones de probabilidad más utilizada para datos circulares, es análoga a la distribución normal para datos lineales, por eso a la distribución de von Mises es conocida también como "Distribución Circular Normal". Su función de densidad está dada por.

$$f(\theta) = [2\pi I_0(\kappa)]^{-1} \exp(\kappa \cos(\theta - \mu)), \quad 0 \leq \theta \leq 2\pi, \quad 0 \leq \kappa \leq \infty, \quad (2.20)$$

donde

μ : Parámetro que representa a la dirección media, $\mu \in [0, \pi)$

κ : Parámetro de concentración, $\kappa \geq 0$

I_0 : denota la función de Bessel modificada de primer tipo y orden cero 0, definida por

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos \theta} d\theta \quad (2.21)$$

Esta distribución es unimodal y simétrica con respecto a $\theta = \mu$. La moda se encuentra en $\theta = \mu$ y la antimoda en $\theta = \mu + \pi$. El cociente entre el valor de la densidad en la moda y el valor de la densidad en la antimoda es $e^{2\kappa}$, así que cuanto mayor sea el valor de κ , mayor será la concentración alrededor de la moda.

FIGURA 2.9

REPRESENTACIÓN POLAR DE LA DISTRIBUCIÓN DE VON MISES (μ, κ).

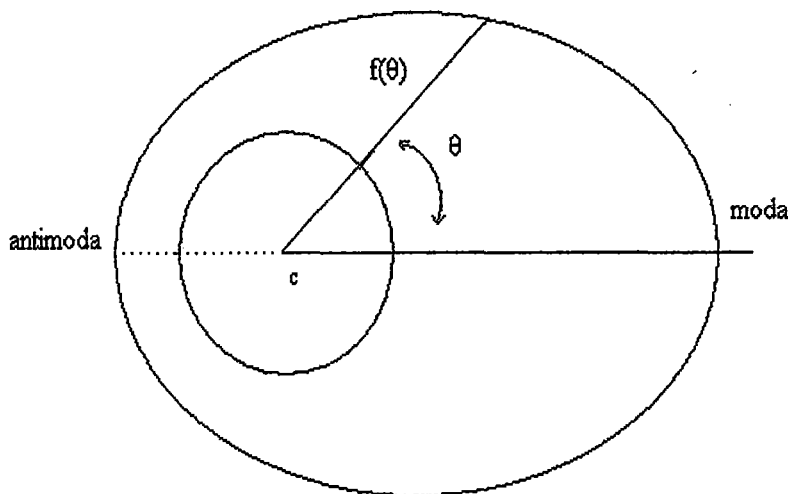
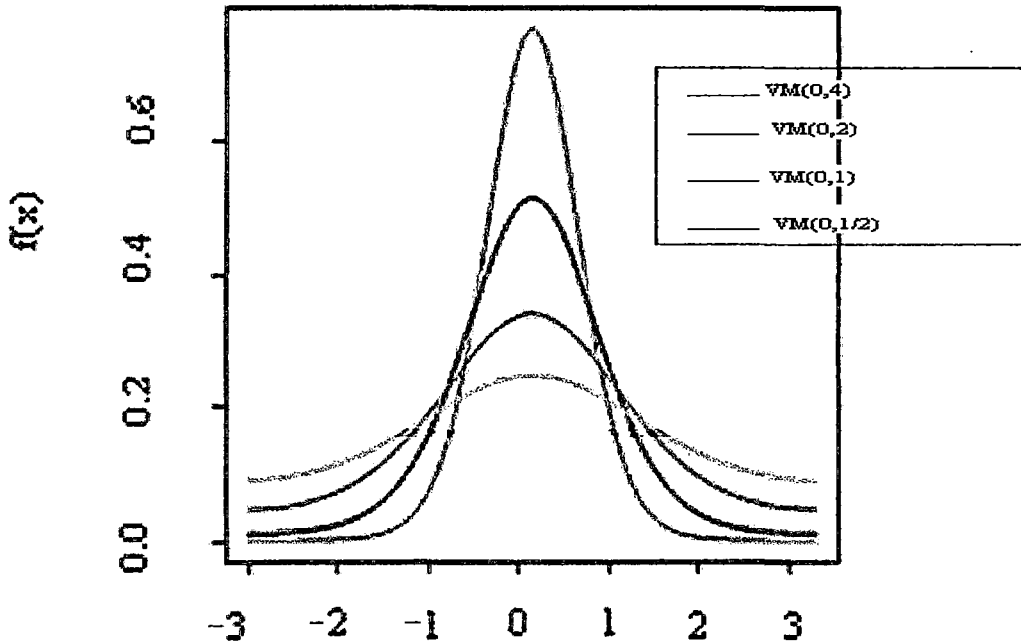


FIGURA 2.10
REPRESENTACIÓN DE LA DISTRIBUCIÓN VON MISES PARA: $\mu = 0$ Y
 $\kappa = 1/2, 1, 2, 4.$



2.7.4. DISTRIBUCIÓN CARDIOIDE

La perturbación de la densidad uniforme por la función coseno da lugar a una distribución llamada cardioide $C(\mu, \rho)$, cuya función de densidad está dado por

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

La longitud media resultante de una distribución $C(\mu, \rho)$ es ρ y si $\rho > 0$, la dirección media es μ . Esta es una distribución simétrica y unimodal con moda en μ cuando $\rho > 0$. Si $\rho = 0$, la distribución cardioide se reduce a la uniforme.

2.7.5. DISTRIBUCIÓN WRAPPED

Dada una distribución cualquiera en la recta real, es posible envolverla alrededor de la circunferencia del círculo unitario. Es decir, si X es una variable aleatoria en la recta, la correspondiente aleatoria X_w de la distribución wrapped viene dada por

$$X_w = X \pmod{2\pi}$$

Si el círculo es identificado con el conjunto de números complejos de módulo 1 entonces la función que envuelve se puede escribir como

$$X \rightarrow e^{iX}.$$

Si X tiene función de distribución F , la función de distribución F_w de X_w está dada por

$$F_w(\theta) = \sum_{\kappa=-\infty}^{\infty} [F(\theta + 2\pi\kappa) - F(2\pi\kappa)], \quad 0 \leq \theta \leq 2\pi$$

En particular, si la variable aleatoria X tiene función de densidad f , la correspondiente función de densidad f_w de X_w es

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

CAPÍTULO III

ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS EN DATOS CIRCULARES

3.1. INTRODUCCIÓN

En este capítulo presentamos métodos para verificar la uniformidad de datos circulares, luego explicamos la prueba de bondad de ajuste para comprobar si los datos provienen de una distribución von Mises, ya que la distribución de von Mises es el más utilizado en técnicas paramétricas en el análisis de datos circulares. También se explica las técnicas de estimación y pruebas de hipótesis de datos provenientes de una distribución von Mises referidas a sus parámetros μ (posición) y κ (concentración). Todos procedimientos se desarrollan para el caso de una muestra y para el caso de dos o más muestras.

3.2. VERIFICACIÓN DE UNIFORMIDAD (PRUEBA DE RAYLEIGH)

Para verificar la uniformidad de datos circulares existen varias pruebas (método gráfico, prueba de bondad de ajuste, prueba de Rayleigh). Pero nos interesa que los datos correspondan a una muestra proveniente de una población unimodal; en este caso conviene utilizar la prueba de Rayleigh [3], para probar la hipótesis:

$$H_0: \text{La distribución es uniforme} \quad (3.1)$$

$$H_1: \text{La distribución es unimodal}$$

Este procedimiento se basa en el hecho de que la longitud del vector medio muestral, r , da una idea de la concentración respecto a una sola dirección. Para muestras de la distribución uniforme circular, el valor de r , en general, es pequeño. Bajo la hipótesis nula de que los datos siguen una distribución uniforme, la distribución muestral de r es conocida, entonces r es una estadística natural para la prueba. La hipótesis de uniformidad es rechazada para valores grandes de r .

Mardia (1972) [20], proporcionó los niveles descriptivos asociados a valores observados de la estadística de la prueba para diferentes tamaños de la muestra.

Cuando el tamaño de la muestra es demasiado grande la estadística

$$Z = nr^2 \quad (3.2)$$

puede ser usada para probar la hipótesis (3.1). En este caso se rechaza H_0 si el valor observado de Z es grande.

3.3. ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS PARA UNA MUESTRA DE DATOS CON DISTRIBUCIÓN VON MISES

3.3.1. PRUEBA DE BONDAD DE AJUSTE

Presentamos dos métodos, un método gráfico y un método formal para evaluar si los datos siguen una distribución von Mises.

1) MÉTODO GRÁFICO

Vamos a construir un gráfico del tipo pp.plot con el *Software Estadístico R*, con la librería "*CircStats*" que se indica en el Anexo 1.

Si los datos corresponden a una muestra de una distribución de von Mises, los puntos deben estar dispersos sobre una recta que pasa por el origen con una inclinación de 45° .

2) PRUEBA DE BONDAD DE AJUSTE

Cuando $n \geq 20$, podemos utilizar un procedimiento formal para probar la hipótesis:

H_0 : Los datos siguen una distribución von Mises

H_1 : Los datos no siguen una distribución von Mises

La prueba de H_0 se basa en la función de distribución acumulada de von Mises. El estadístico de prueba está dado por:

$$U^2 = \sum_{i=1}^n [z_{(i)}^2] - \frac{1}{n} \sum_{i=1}^n (2i-1)z_i + \frac{n}{3} - n\left(\bar{z} - \frac{1}{2}\right)^2 \quad (3.3)$$

Donde

- $z_i = F_{\hat{\kappa}}(\theta_i - \hat{\mu})$, $i = 1, \dots, n$,
- $\hat{\mu}$ y $\hat{\kappa}$ son obtenidos de las expresiones (3.8) y (3.12)
- $\bar{z} = \sum_{i=1}^n z_{(i)} / n$ se obtiene reordenando los z_i en orden creciente, $z_{(1)} < \dots < z_{(n)}$

La distribución de U^2 esta tabulada en Fisher (1993) y rechazamos H_0 para valores grandes de U^2 .

3.3.2. PRUEBA PARA OUTLIERS

Ocasionalmente encontramos conjuntos de datos en los que una o más observaciones se destacan de las demás. Esos valores discrepantes (outliers ó datos atípicos) pueden ocurrir debido a un error en la codificación de los datos, o la variación muestral haber ocasionado un valor aislado. Exploratoriamente, podemos juzgar si un punto se destaca de la masa de datos representando gráficamente la muestra en un diagrama circular (Ver capítulo 2 representación gráfica). Sin embargo la ocurrencia de outliers no presenta serios problemas al análisis de datos circulares, consideramos en esta sub sección un procedimiento formal de prueba que permite verificar si un punto se aparta de forma significativa de la masa de datos, cuando la muestra proviene de una distribución de von Mises.

Suponga que una observación θ_k se aleja de la masa de datos, y sean R y R_k las longitudes de los resultados en la muestra completa y en la muestra obtenida excluyéndose θ_k , respectivamente. La estadística de la prueba es dada por:

$$M_n = \frac{R_k - R + 1}{n - R}, \quad k = 1, \dots, n. \quad (3.4)$$

Si θ_k es un valor discrepante la diferencia entre R y R_k es grande; así θ_k es clasificado como un "outliers" si M_n es grande.

3.3.3. ESTIMACIÓN DE MÁXIMA VEROSIMILITUD DE LOS PARÁMETROS μ Y κ

Sea $\theta_1, \dots, \theta_n$ una muestra de tamaño n de una distribución von Mises (μ, κ) . La función de verosimilitud de las observaciones está dada por:

$$L(\mu, \kappa; \boldsymbol{\theta}) = [2\pi I_0(\kappa)]^{-n} \exp\left\{\kappa \sum_{i=1}^n \cos(\theta_i - \mu)\right\}. \quad (3.5)$$

Maximizar (3.5) es equivalente a maximizar el logaritmo de la función de verosimilitud, es decir,

$$l(\mu, \kappa) = -n \ln 2\pi - n \ln I_0(\kappa) + \kappa \sum_{i=1}^n \cos(\theta_i - \mu). \quad (3.6)$$

El estimador de máxima verosimilitud de μ , $\hat{\mu}$, satisface a la ecuación:

$$\frac{\partial l}{\partial \mu} = \kappa \sum_{i=1}^n \text{sen}(\theta_i - \hat{\mu}) = 0 \Rightarrow \sum_{i=1}^n \text{sen}(\theta_i - \hat{\mu}) = 0. \quad (3.7)$$

De la sub sección (2.4.1) sabemos que esta ecuación se satisface cuando $\hat{\mu}$ es el ángulo medio de la muestra. Así, el estimador de máxima verosimilitud de μ está dado por:

$$\hat{\mu} = \bar{\theta}. \quad (3.8)$$

El estimador de κ es tal que:

$$\frac{\partial l}{\partial \kappa} = -n \frac{I'_0(\hat{\kappa})}{I_0(\hat{\kappa})} + \sum_{i=1}^n \cos(\theta_i - \hat{\mu}) = 0. \quad (3.9)$$

Considerando la propiedad de la función Bessel: $I'_0(\hat{\kappa}) = I_0(\hat{\kappa})$, tenemos:

$$\frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \hat{\mu}). \quad (3.10)$$

De la sub sección (2.4.1) sabemos que $\sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = nr$. Luego, el estimador de máxima verosimilitud de κ es la solución de:

$$A(\hat{\kappa}) = \frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = r. \quad (3.11)$$

Pero esta ecuación no tiene una solución explícita y debe resolverse numéricamente. Fisher (1993), usando propiedades de la función Bessel, presenta una aproximación para la solución de (3.11) que es dada por:

$$\hat{\kappa} = \begin{cases} 2r + r^3 + \frac{5r^5}{6} & \text{si } r < 0.53; \\ -0.4 + 1.39r + \frac{0.43}{1-r} & \text{si } 0.53 \leq r < 0.85; \\ \frac{1}{r^3 - 4r^2 + 3r} & \text{si } r \geq 0.85. \end{cases} \quad (3.12)$$

Desafortunadamente, cuando el tamaño de la muestra y r son pequeños ($r < 0.45$), $\hat{\kappa}$ tiende a sobrestimar el parámetro κ . Fisher propuso el siguiente estimador para κ cuando $n \leq 15$:

$$\hat{\kappa}^* = \begin{cases} \max(\hat{\kappa} - 2(n\hat{\kappa})^{-1}, 0) & \text{si } \hat{\kappa} < 2; \\ \frac{(n-1)^3 \hat{\kappa}}{n^3 + n} & \text{si } \hat{\kappa} \geq 2. \end{cases} \quad (3.13)$$

PROPIEDADES DE LOS ESTIMADORES DE MÁXIMA VEROSIMILITUD $\hat{\mu}$ Y

$\hat{\kappa}$

Para n grande se tiene la esperanza y varianza del estimador de μ :

$$E(\hat{\mu}) = \mu \tag{3.14}$$

$$\text{var}(\hat{\mu}) \cong \frac{1}{n \kappa A(\kappa)} \tag{3.15}$$

Utilizando aproximación para $A(\kappa)$, Mardia (1972) muestra que la expresión (3.15) puede ser escrita como:

$$\text{var}(\hat{\mu}) = \begin{cases} \frac{2}{n\kappa^2}, & \text{si } \kappa \leq 10; \\ \frac{1}{n\kappa}, & \text{si } \kappa > 10. \end{cases} \tag{3.16}$$

Así que μ puede estimarse con menos precisión para pequeñas κ que para valores grandes de κ lo que se esperaría, puesto que cuando $\kappa \rightarrow 0$, la distribución de von Mises se aproxima a la distribución uniforme circular.

Con relación a $\hat{\kappa}$, la esperanza y la varianza es:

$$E(\hat{\kappa}) \approx \kappa + \frac{1}{2nB^2} (2\kappa^2 A(\kappa)(1 - A(\kappa)) - 3A^2(\kappa)\kappa + \kappa - 2A(\kappa)),$$

donde

$$B = \kappa(1 - A^2(\kappa)) - A(\kappa),$$

y

$$\text{var}(\hat{\kappa}) = \frac{\kappa}{n(\kappa - \kappa A^2(\kappa) - A(\kappa))}. \tag{3.17}$$

Tomamos nota de que el sesgo de $\hat{\kappa}$ no es insignificante, a menos que n sea grande. Todavía tenemos que

$$\text{cov}(\hat{\mu}, \hat{\kappa}) = 0.$$

Por lo tanto, para n grande, $\hat{\mu}$ y $\hat{\kappa}$ son variables aleatorias independientes y

$$\hat{\mu} \underline{D}N(\mu, \text{var}(\hat{\mu})) \text{ y}$$

$$\hat{\kappa} \underline{D}N(\kappa, \text{var}(\hat{\kappa})),$$

donde $\text{var}(\hat{\mu})$ y $\text{var}(\hat{\kappa})$ son dadas por (3.16) y (3.17), respectivamente.

Recordemos que los estimadores de máxima verosimilitud son asintóticamente eficientes.

3.3.4 ESTIMACIÓN POR INTERVALOS DE LOS PARÁMETROS μ Y κ

1) ESTIMACIÓN POR INTERVALO DE μ

El modo de obtener los límites de Confianza para μ , fijar un coeficiente de confianza $\gamma = 1 - \alpha$, dependerá del tamaño de la muestra y de la magnitud $\hat{\kappa}$.

Una guía para la selección del método a ser utilizada sobre la base de los valores de n y $\hat{\kappa}$ es presentado en la siguiente tabla.

TABLA 3.1
GUÍA PARA SELECCIÓN DEL MÉTODO DE ESTIMACIÓN

	Método bootstrap	Método asintótico
$\hat{\kappa} < 0.4$	$\forall n$	
$0.4 \leq \hat{\kappa} < 1.0$	$n < 25$	$n \geq 25$
$1.0 \leq \hat{\kappa} < 1.5$	$n < 15$	$n \geq 15$
$1.5 \leq \hat{\kappa} < 2$	$n < 10$	$n \geq 10$
$\hat{\kappa} \geq 2.0$		$\forall n$

Para valores muy pequeños de $\hat{\kappa}$ ($\hat{\kappa} < 0.4$), se sugiere que el intervalo sea construido a través del procedimiento bootstrap paramétrico que sea descrito en la sección 5) del anexo 6.

Para valores grandes de $\hat{\kappa}$ ($\hat{\kappa} > 2$) o tamaños de muestras grandes, pueden ser utilizados los resultados asintóticos presentados en Fisher y Lewis (1983).

Estos autores mostraron que, para n o κ grandes un intervalo de confianza para μ es:

$$\hat{\mu} \pm \text{sen}^{-1}(z_{\alpha/2} \hat{\sigma}_{VM}) \tag{3.18}$$

donde $z_{\alpha/2}$ denota el percentil de orden $1 - \frac{\alpha}{2}$ de la distribución $N(0, 1)$ y,

$$\hat{\sigma}_{VM} = \frac{1}{(nr\hat{k})^{1/2}}. \quad (3.19)$$

Considerando las propiedades asintóticas de $\hat{\mu}$ y \hat{k} presentadas en la sección anterior, tenemos que, para n grande, un intervalo de confianza para μ es dado por:

$$\hat{\mu} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\mu})} \Rightarrow \hat{\mu} \pm z_{\alpha/2} (nr\hat{k})^{-1/2}, \quad (3.20)$$

donde $z_{\alpha/2}$ es el cuantil de orden $1 - \frac{\alpha}{2}$ de la distribución $N(0, 1)$. Como $\text{sen}^{-1}x = x + \frac{1}{2.3}x^3 + \frac{1.3}{2.4.5}x^5 + \frac{1.3.5}{2.4.6.7}x^7 + \dots$ ($-1 < x < 1$), los intervalos (3.18) y (3.20) son esencialmente equivalentes para pequeños valores de $(nr\hat{k})^{-1}$, lo que ocurre cuando n es grande.

2) ESTIMACIÓN POR INTERVALO DE κ

Cuando $\hat{k} < 2$ utilizamos el método bootstrap paramétrico descrito en 7) del anexo 6.

Obteniendo, a través del remuestreo, B valores de $C = \sum \cos \theta_i$ y $S = \sum \text{sen} \theta_i$:

$$\begin{bmatrix} C_1 \\ S_1 \end{bmatrix} \dots \begin{bmatrix} C_B \\ S_B \end{bmatrix}.$$

Calculamos $r_i^* = (C_i^2 + S_i^2)^{1/2}$, $\hat{\kappa}_i^*$ utilizando las expresiones (3.12) y (3.13) con $r = r_i^*$ para $i=1, \dots, B$. Ordenamos las estimaciones bootstraps $\hat{\kappa}_1^*, \dots, \hat{\kappa}_B^*$ en orden creciente, obteniendo $\hat{\kappa}_{(1)}^* \leq \dots \leq \hat{\kappa}_{(B)}^*$. Un intervalo de confianza para κ con coeficiente de confianza $\gamma = 1 - \alpha$ viene dado por:

$$\left(\hat{\kappa}_{(l+1)}^*, \hat{\kappa}_{(m)}^* \right),$$

donde, $l = \left[\frac{1}{2} B\alpha + \frac{1}{2} \right]$, $[\cdot]$ indica la parte entera y $m = B - l$.

Para $\hat{\kappa} \geq 2$, se utiliza la siguiente aproximación demostrada por Stephen (1969):

$$2\psi(n-R) \approx \chi_{n-1}^2, \quad (3.21)$$

donde

$$2\psi^{-1} = \kappa^{-1} + \frac{3}{8} \kappa^{-2}.$$

Un intervalo de confianza para κ con coeficiente de confianza $\gamma = (1 - \alpha)$ es dado por:

$$\left[\frac{1 + (1 + 3a)^{1/2}}{4a}, \frac{1 + (1 + 3b)^{1/2}}{4b} \right],$$

donde

$$a = \frac{(n-r)}{\chi_{n-1}^2(1-\frac{1}{2}\alpha)}, b = \frac{(n-r)}{\chi_{n-1}^2(\frac{1}{2}\alpha)} \text{ y } \chi_{n-1}^2(\alpha) \text{ es el percentil de orden } 1-\alpha \text{ de una } \chi_{n-1}^2.$$

La elección entre los dos métodos de construcción del intervalo de confianza para κ descrito anteriormente depende solo del valor $\hat{\kappa}$ observado, que es un indicativo de la magnitud del verdadero valor del parámetro. Para muestras grandes, podemos construir un intervalo de confianza utilizando los resultados asintóticos presentados en la sección 3.3.3. Fijando un coeficiente de confianza $\gamma = 1 - \alpha$, tenemos que un intervalo de confianza para κ es dado por:

$$\hat{\kappa} \pm z_{\frac{\alpha}{2}} \sqrt{\text{var}(\hat{\kappa})}, \quad (3.22)$$

donde $z_{\frac{\alpha}{2}}$ es el cuantil de orden $1 - \frac{\alpha}{2}$ de la distribución $N(0,1)$ y $\text{var}(\hat{\kappa})$ es la varianza estimada de κ obtenida a partir del (3.17) sustituyéndose κ por $\hat{\kappa}$.

3.3.5. PRUEBAS DE HIPÓTESIS SOBRE μ Y κ

1) PRUEBA DE HIPÓTESIS SOBRE μ

Supongamos que deseamos probar las hipótesis:

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_1 : \mu &\neq \mu_0 \end{aligned} \quad (3.23)$$

Donde μ_0 es una constante conocida. Supongamos que el parámetro de concentración κ es desconocida. Una forma de probar H_0 es a través de la construcción del intervalo de confianza para μ (ver subsección 3.3.4 1)); si μ_0 esta en este intervalo se acepta la hipótesis nula.

Otra forma de probar H_0 (3.23) es a través de la estadística:

$$T_n = \frac{[\text{sen}(\hat{\mu} - \mu_0)]}{(nr\hat{\kappa})^{-1/2}}. \quad (3.24)$$

Para valores grandes de n o κ , (Fisher y Lewis (1983)) T_n tiene distribución próxima a una $N(0,1)$. Así, rechazamos H_0 a un nivel aproximadamente igual α si $|T_n| > z_{\frac{\alpha}{2}}$, donde $z_{\frac{\alpha}{2}}$ denota el cuantil de orden $1 - \frac{\alpha}{2}$ de la distribución normal estándar. Para valores pequeños de n y $\hat{\kappa}$ la distribución de la estadística es generada por el método bootstrap (6) del anexo 6. Tabla 3.1 sirve como una guía para la selección del método de prueba que se adoptará.

2) PRUEBA DE HIPÓTESIS PARA κ

Supongamos que deseamos probar las siguientes hipótesis:

$$\begin{aligned} H_0 : \kappa &= \kappa_0 \\ H_1 : \kappa &\neq \kappa_0 \end{aligned} \quad (3.25)$$

Considerando que μ es desconocido. De la misma forma como en 1) podemos utilizar el intervalo de confianza para κ para probar las hipótesis de anterior. Cuando κ es grande ($\hat{\kappa} > 2$) podemos también realizar la prueba directamente, considerando la estadística:

$$R = nA(\hat{\kappa}).$$

Rechazamos H_0 al nivel α cuando:

$$R < n - X_{n-1}^2 \left(\frac{\alpha}{2} \right) / 2\psi_0$$

o

$$R < n - X_{n-1}^2 \left(1 - \frac{\alpha}{2} \right) / 2\psi_0,$$

donde, $\psi_0 = \frac{1}{k_0 + \frac{3}{8k_0^2}}$.

3.4. ESTIMACIÓN DE PARÁMETROS Y PRUEBAS DE HIPÓTESIS DE DOS O MÁS MUESTRAS DE LA DISTRIBUCIÓN VON MISES

Consideraremos dos o más muestras independientes de datos provenientes de distribuciones de von Mises, y comparamos sus direcciones medias y dispersiones a través de prueba de hipótesis. Si la suposición de que las poblaciones son de von Mises es verdadera. En las secciones 3.5 y 3.6 presentamos inferencias sobre direcciones medias, y en las secciones 3.7 y 3.8, sobre parámetros de concentración.

Adoptamos la misma notación que las secciones anteriores, considerando que las q muestras de tamaño n_i siguen una distribución de von Mises, $VM(\mu_i, k_i)$.

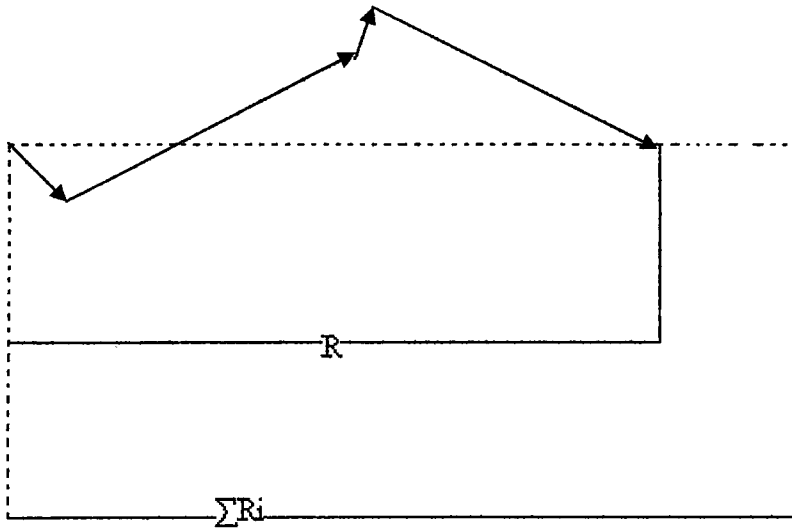
3.5. PRUEBAS DE IGUALDAD DE DIRECCIONES MEDIAS

Nuestro objetivo es probar las hipótesis:

$$\begin{aligned} H_0 : \mu_1 = \dots = \mu_q = \mu, \\ H_1 : \text{por lo menos una igualdad no es válida.} \end{aligned} \tag{3.26}$$

Si la hipótesis nula es correcta, la dirección de la resultante en cada muestra debe ser próxima al valor de la dirección media común μ . En el caso de las muestras serán agrupados, entonces la resultante en la muestra conjunta también debe tener dirección próxima a μ , y su magnitud no debe ser muy menor del que la suma de las magnitudes de los resultantes en cada muestra (ver Figura 3.1).

FIGURA 3.1
LA BASE DE UNA PRUEBA PARA UNA DIRECCION MEDIA COMUN



Una estadística natural para probar la hipótesis (3.26) es, entonces:

$$A = R_1 + R_2 + \dots + R_q - R \quad (3.27)$$

y valores grandes de A sugieren que H_0 no es verdadera. Las pruebas propuestos para probar H_0 son en general, funciones de A , y pueden ser divididos en dos subclases, dependiendo de los q parámetros de concentración $\kappa_1, \dots, \kappa_q$ serán o no todos iguales.

3.5.1. SI LOS PARÁMETROS DE CONCENTRACIÓN NO SON TODOS IGUALES

La conducta que se adopte depende del tamaño de la muestra y los valores observados de los estimadores de máxima verosimilitud de κ_i .

1) Existe por lo menos un $n_i < 25$ o $\hat{\kappa}_i < 2, i = 1, \dots, q$.

Se utiliza el método bootstrap, descrito en la sección 9) del anexo 6, con remuestreo de la distribución de von Mises $VM(0, \hat{\kappa}_i), i = 1, 2, \dots, q$. La estadística a ser considerada es:

$$Y_q = 2\left(\sum_{i=1}^q \frac{1}{\hat{\sigma}_i^2} - R_M\right) \quad (3.28)$$

Siendo ahora:

$$\frac{1}{\hat{\sigma}_i^2} = n_i r_i \hat{\kappa}_i,$$

$$R_M = (\hat{C}_M^2 + \hat{S}_M^2)^{1/2}, \quad \hat{C}_M = \sum_{i=1}^q \cos \hat{\mu}_i / \hat{\sigma}_i^2, \quad \hat{S}_M = \sum_{i=1}^q \sin \hat{\mu}_i / \hat{\sigma}_i^2,$$

$\hat{\kappa}_i$ es el valor observado del estimador de máxima verosimilitud de κ_i ,

r_i es la resultante media en la i -ésima muestra, y

$\hat{\sigma}_i = \frac{1}{r_i \hat{\kappa}_i}$ es la dispersión muestral en la i -ésima muestra, $i = 1, \dots, q$.

La hipótesis de igualdad de las q direcciones medias es rechazada para valores grandes de Y_q .

2) $n_i \geq 25, \hat{\kappa}_i \geq 2, i = 1, \dots, q$.

En este caso, Watson (1983) muestra que la distribución de Y_q es próxima de una χ^2 con $(q-1)$ grados de libertad, esto es,

$$Y_q \rightarrow \chi_{q-1}^2.$$

Rechazamos la hipótesis nula para valores grandes de Y_q .

3.5.2 SI LOS PARÁMETROS DE CONCENTRACIÓN SON IGUALES

Varias pruebas fueron propuestas para el caso en que $\kappa_1 = \dots = \kappa_q = \kappa$, desconocido. Describiremos inicialmente la prueba de Watson-Williams (1956) adecuado para comparar las direcciones medias de dos o más poblaciones de von Mises cuando el valor del parámetro de concentración κ es grande. Como el valor de κ es desconocido, la decisión sobre que procedimiento debe ser adoptado en un dado problema debe ser tomada con base en una estimación de este parámetro. Una posible estimación puede ser obtenida calculándose el valor del estimador de máxima verosimilitud de κ en la muestra combinada, o sea,

$$\hat{\kappa} = A^{-1}(r) \tag{3.29}$$

1) Prueba de Watson-Williams

Watson y Williams (1956) mostró que, cuando κ es grande,

$$U = 2\kappa \left(\sum_{j=1}^q R_j - R \right) \rightarrow X_{q-1}^2, \quad \text{bajo la } H_0 \quad (3.30)$$

$$V = 2k \left(n - \sum_{j=1}^q R_j \right) \rightarrow X_{q-1}^2, \quad (3.31)$$

U y V son independientes. Los autores propusieron la siguiente estadística para la prueba:

$$F = \frac{(n-q) \left(\sum_{i=1}^q R_i - R \right)}{(q-1) \left(n - \sum_{i=1}^q R_i \right)}. \quad (3.32)$$

Para valores grandes de κ , la distribución de F , bajo la H_0 , se aproxima a una F-Snedecor con $(q-1)$ y $(n-q)$ grados de libertad. Rechazamos H_0 si $F > F_{q-1, n-q, \alpha}$.

Observemos que los cálculos necesarios para la obtención de la estadística F pueden estar dispuestos en un cuadro de análisis de la varianza, que se muestra en la Tabla (3.2).

TABLA 3.2
ANÁLISIS DE VARIANZA.

fuelle de variación	de	g.l	Suma de cuadrados	Cuadrados medios	F
entre muestras		$q-1$	$\sum_{i=1}^q R_i - R$	$\left(\sum_{i=1}^q R_i - R \right) / (q-1) = a$	a/b
dentro de muestras		$n-q$	$n - \sum_{i=1}^q R_i$	$\left(n - \sum_{i=1}^q R_i \right) / (n-q) = b$	
total		$n-1$	$n - R$		

Stephen (1972) propone la siguiente modificación en el estadístico F :

$$F^* = \left(1 - \frac{3}{8k} \right) F \quad (3.33)$$

y mostro, a través de simulación, que el factor de correlación $g = 1 - \frac{3}{8k}$ hace que la distribución de F^* sea más cercano de F-Snedecor con $(q-1)$ y $(n-q)$ grados de libertad. El autor ha demostrado que en el caso de ambas muestras es una buena aproximación para $\kappa \geq 2$, y por más de dos muestras, la aproximación es adecuada para $\kappa \geq 1$. En los casos en que la utilización de F^* no es recomendada, pueden ser adoptados los procedimientos descritos a seguir. Observemos que, para $\kappa > 10$, el factor de correlación g es insignificante, F y F^* se convierten en equivalentes.

2) Prueba exacto de Stephens

Stephens (1972) dedujo una prueba precisa para comparar las direcciones medias de dos poblaciones de von Mises con parámetros iguales de concentración. Esta prueba se basa en la estadística A dada en la expresión (3.27). Los valores críticos para la prueba son obtenidos en tablas especiales que no son particularmente fáciles de usar. Upton y Fingleton (1989) sugieren formulas aproximadas para la obtención de los valores críticos de la prueba (ver Tabla 3.3). Estas formulas proporcionan una buena aproximación para el caso en que $n_1 = n_2$, y son conservadoras en otros casos.

TABLA 3.3
FORMULA PARA OBTENER VALORES CRÍTICOS PARA LA PRUEBA DE
STEPHENS.

Nivel de significancia	Fórmula
5%	$n \cdot \exp(1.30 - 0.75 \ln(n) - 3.785R/n + 14.0R/n^2)$
1%	$n \cdot \exp(1.27 - 0.66 \ln(n) - 3.30R/n + 12.4R/n^2)$

3) Prueba de Mardia

Considerando el caso en que $0 < \kappa < 1$ y $q > 2$, ($r < 0.45$) Mardia (1972) mostró que la estadística de prueba de la razón de verosimilitud se puede escribir, aproximadamente, como:

$$-2 \ln \lambda \cong \frac{2}{n} \left(1 - \frac{\kappa^2}{8} + \frac{q}{2n\kappa^2} \right)^{-1} \left(\left(\sum_{i=1}^q R_i \right)^2 - R^2 \right) = U.$$

Cuando H_0 es verdadera, U tiene distribución próxima a una X_{q-1}^2 . En la práctica, sustituimos κ por su estimación de máxima verosimilitud $\hat{\kappa}$, y rechazamos H_0 para valores grandes de U . La aproximación anterior es buena incluso para valores pequeños de n , dado que κ no está muy próximo de cero o uno.

3.6. INFERENCIAS DESPUES DE LA COMPARACION DE LAS DIRECCIONES MEDIAS

Las pruebas que se describen en las secciones anteriores pueden conducir al rechazo o no de la hipótesis de igualdad de las direcciones medias. En el caso de no rechazar H_0 , es interesante estimar la dirección media común μ por punto y por intervalo. Por otro lado, si H_0 es rechazada y tenemos tres o más muestras, continuamos el análisis con el objetivo de localizar las diferencias existentes. Procedimientos adecuados para cada situación se describirán a continuación.

3.6.1. ESTIMACIÓN DE LA DIRECCIÓN MEDIA COMÚN DE DOS O MÁS DISTRIBUCIONES DE VON MISES

Para estimar la dirección media común μ debemos considerar si los parámetros de concentración en las q poblaciones son o no iguales.

1) $k_1 = \dots = k_q = k$

Tenemos q muestras de una sola población de von Mises. Por lo tanto, todos los datos se pueden combinar en una sola muestra de n puntos y se aplican los métodos presentados en las secciones 3.3.3 y 3.3.4.

2) k_1, \dots, k_q no son todos iguales

Considerando las expresiones para los errores estándar circulares y dispersiones circulares muestrales propias a la distribución de von Mises,

$$\hat{\sigma}_i = \frac{1}{(n_i r_i \hat{\kappa}_i)^{1/2}}, \quad \hat{\delta}_i = \frac{1}{(r_i \hat{\kappa}_i)^{1/2}}, \quad i = 1, \dots, q. \quad (3.34)$$

Tomemos ahora que:

$$v_i = (r_i \hat{\sigma}_i^2), \quad v = \sum_{i=1}^q v_i, \quad w_i = \frac{v_i}{v} \quad y \quad \sigma_M^2 = \sum_{i=1}^q \frac{w_i^2 r_i^2 \hat{\sigma}_i^2}{\left(\sum_{i=1}^q w_i r_i \right)^2}. \quad (3.35)$$

La forma de construcción del intervalo de confianza para la media común μ depende de los tamaños n_i y de las estimaciones de los parámetros de concentración $\hat{\kappa}_i, i = 1, \dots, q$.

A) Por lo menos un $n_i < 25$ o $\hat{\kappa}_i < 2$

Deben ser utilizadas las técnicas bootstrap descritas en 10) en el anexo 6, con remuestreo a partir de la distribución von Mises $VM(0, \hat{\kappa}_i), i = 1, \dots, q$, para estimar la dirección media común y construir un intervalo de confianza.

B) $n_i \geq 25, \hat{\kappa}_i \geq 2$

En este caso un estimador combinado por puntos para la dirección media común μ es dado por:

$$\hat{\mu}_w = \begin{cases} \arctag\left(\frac{\bar{S}_w}{\bar{C}_w}\right) & \text{si } \bar{S}_w > 0, \bar{C}_w > 0; \\ 180^\circ + \arctag\left(\frac{\bar{S}_w}{\bar{C}_w}\right) & \text{si } \bar{C}_w < 0; \\ 360^\circ + \arctag\left(\frac{\bar{S}_w}{\bar{C}_w}\right) & \text{si } \bar{S}_w < 0, \bar{C}_w > 0, \end{cases} \quad (3.36)$$

Donde:

$$\bar{C}_w = \sum_{i=1}^q w_i r_i \cos \hat{\mu}_i = \sum_{i=1}^q \frac{w_i C_i}{n_i}$$

$$\bar{S}_w = \sum_{i=1}^q w_i r_i \sin \hat{\mu}_i = \sum_{i=1}^q \frac{w_i S_i}{n_i}$$

Utilizando para su cálculo w_i y v_i como en la expresión (3.35)

Un intervalo de confianza con coeficiente de confianza $\gamma = 1 - \alpha$ (Fisher, 1993) es dado por:

$$[\hat{\mu}_w \pm \arcsen(z_{\alpha/2} \hat{\sigma}_M)], \quad (3.37)$$

donde

$\hat{\mu}_w$ es dado en la expresión (3.36) y $\hat{\sigma}_M$ es obtenido a partir de (3.35).

3.6.2. CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE DOS DIRECCIONES MEDIAS

En el caso de los tres términos o más poblaciones de la muestra von Mises con el mismo parámetro κ , podemos encontrar las diferencias existentes entre las direcciones medias construyendo intervalos de confianza para las diferencias;

$$\delta_{ij} = (\mu_i - \mu_j) \pmod{2\pi}, \quad i, j = 1, \dots, q, \quad i > j. \quad (3.38)$$

Mardia mostro que $\bar{\theta}_i / R_i$ y $\bar{\theta}_j / R_j$, tiene distribución $VM(\mu_i, \kappa R_i)$ y $VM(\mu_j, \kappa R_j)$ respectivamente, y la diferencia:

$$d_{ij} = (\bar{\theta}_i - \bar{\theta}_j) \pmod{2\pi}$$

tiene distribución próxima a una $VM(\delta_{ij}, \kappa_{ij}^*)$, donde κ_{ij}^* es tal que:

$$A(\kappa_{ij}^*) = A(\kappa R_i) \cdot A(\kappa R_j). \quad (3.39)$$

En la práctica, sustituimos κ por su estimador de máxima verosimilitud $\hat{\kappa}$ dado por:

$$A(\hat{\kappa}) = \frac{R_1 + \dots + R_q}{n} \quad (3.40)$$

Así, la distribución de d_{ij} puede ser aproximada por una $VM(\delta_{ij}, \hat{\kappa}_{ij}^*)$, donde

$$A(\hat{\kappa}_{ij}^*) = A(\hat{\kappa}R_i) \cdot A(\hat{\kappa}R_j). \quad (3.41)$$

Sea θ una variable aleatoria distribuida según una $VM(0, \kappa_{ij}^*)$ y suponga que, para un determinado α , v es tal que:

$$P(\pi - v < \theta < \pi + v) = \alpha, \quad (3.42)$$

donde el arco $(\pi - v, \pi + v)$ incluyen π . Entonces un intervalo de confianza para δ_{ij} con coeficiente de confianza $\gamma = 1 - \alpha$ es dado por:

$$\text{arc}\{(d_{ij} + \pi - v) \pmod{2\pi}; (d_{ij} + \pi + v) \pmod{2\pi}\}, \quad (3.43)$$

donde el arco $(d_{ij} + \pi - v, d_{ij} + \pi + v)$ incluyen d_{ij} . Supongamos que estamos interesados en hacer c comparaciones de las q medias dos a dos a un nivel de significancia, en el máximo, igual a un valor α^* fijado. Debemos, entonces, en la construcción del intervalo (3.43), fijar $\alpha = \frac{\alpha^*}{c}$.

Alternativamente, considerando que $d_{ij} \sim VM(\delta_{ij}, \kappa_{ij}^*)$, donde κ_{ij}^* satisface la expresión (3.39), podemos construir intervalos de confianza para las diferencias δ_{ij} , $i, j = 1, \dots, q$, $i > j$, utilizando los procedimientos de estimación de la media de una distribución de von Mises presentados en la sección 3.3.

Proponemos que la elección entre el método bootstrap y el asintótico se haga de acuerdo con la tabla 3.1, considerándose $\hat{\kappa}^* = \min(\hat{\kappa}_{ij}^*)$ en el lugar de $\hat{\kappa}$ en la primera columna, y $n^* = \min(n_1, \dots, n_q)$ en tercera columna de la tabla.

En situaciones en que se haya adoptado el método bootstrap, debemos utilizar el procedimiento de remuestreo descrito en 5) del anexo 6.

Si $\hat{\kappa}^*$ y n^* son tales que los resultados para grandes muestras son adecuados, el intervalo de confianza para δ_{ij} , con coeficiente de confianza $\gamma = 1 - \alpha$ es:

$$(\bar{\theta}_i - \bar{\theta}_j) \pm \arcsen \left(z_{\frac{\alpha}{2}} (\hat{\kappa}_{ij}^* A(\hat{\kappa}_{ij}^*))^{-1/2} \right). \quad (3.44)$$

3.7 PRUEBA DE IGUALDAD DE LOS PARÁMETROS DE CONCENTRACIÓN DE DOS O MÁS DISTRIBUCIONES DE VON MISES

Consideremos el problema de probar la hipótesis:

$$H_0 : \kappa_1 = \dots = \kappa_q \quad (3.45)$$

Contra la alternativa de que por lo menos un κ_i es diferente de los demás.

Sean:

$$\begin{aligned} \bar{\theta}_i &= \hat{\mu}_i \text{ la dirección media en la } i\text{-ésima muestra,} \\ d_{ij} &= |\text{sen}(\theta_{ij} - \bar{\theta}_i)|, \quad j=1, \dots, n_i \\ \bar{d}_i &= \sum_{j=1}^{n_i} \frac{d_{ij}}{n_i}, \quad i=1, \dots, q \quad \text{y} \\ \bar{d} &= \sum_{i=1}^q \frac{n_i \bar{d}_i}{n} \end{aligned} \quad (3.46)$$

Observemos que d_{ij} puede ser considerado como el "residuo" de la j -ésima observación en el i -ésimo grupo. Si las concentraciones en las q poblaciones fueron iguales, debemos esperar que los valores medios de los "residuos" \bar{d}_i en las q muestras sean próximos a \bar{d} . Fisher (1986) sugirió, entonces, la siguiente estadística para probar la hipótesis (3.45):

$$f_q = \frac{(n-q) \sum_{i=1}^q (\bar{d}_i - \bar{d})^2}{(q-1) \sum_{i=1}^q \sum_{j=1}^{n_i} (d_{ij} - \bar{d}_i)^2} \quad (3.47)$$

Rechazamos H_0 para valores grandes de f_q . El procedimiento para la obtención de la región crítica de la prueba depende de los tamaños de las muestras y de las dispersiones muestrales. En primer lugar debemos obtener $\tilde{\kappa}$ a partir de la expresión:

$$\tilde{\kappa} = \text{mediana}(\hat{\kappa}_1, \dots, \hat{\kappa}_q) \quad (3.48)$$

es decir, $\tilde{\kappa}$ es la estimación mediana de κ en q muestras. Entonces, pueden ocurrir las siguientes situaciones:

1) $n_i \geq 10, \hat{\kappa} \geq 1, i = 1, \dots, q.$

El valor crítico de la prueba, a un nivel de significancia α es dado por el cuantil de orden $(1 - \alpha)$ de la distribución $F_{q-1, n-q}$.

2) **Al menos una de las desigualdades en 1) no está satisfecha**

En este caso, utilizamos una prueba aleatorizada (Fisher 1993). Cada muestra debe ser centrada en su media y procedemos a la prueba aleatorizada basada en la asignación de n valores centrados en las q muestras, calculando la estadística f_q en cada una de las asignaciones aleatorias.

Otras pruebas para la hipótesis de (3.45) se han propuesto en la literatura. Sin embargo, según Fisher (1993), estos procedimientos son sensibles a los valores atípicos y la suposición de que la población muestreado es de von Mises.

3.8. ESTIMACIÓN DEL PARÁMETRO DE CONCENTRACIÓN COMÚN DE DOS O MÁS DISTRIBUCIONES DE VON MISES

Supongamos que la hipótesis (3.45) no sea rechazada. Es interesante, pues, estimar la concentración común parámetro κ . Si podemos suponer que las q direcciones medias son iguales, podemos considerar la muestra conjunta de tamaño n y aplicar la técnica presentada en la Sub sección 3.3.3. Si no tenemos informaciones sobre las direcciones medias, o sabemos que ellos son diferentes, el procedimiento a ser adoptado depende de la magnitud de κ . El valor mediano $\tilde{\kappa}$ dado en (3.48) puede ser utilizado en la elección del procedimiento a ser adoptado:

1) $\tilde{\kappa} < 2.0$

Sea
$$R_{tot} = R_1 + \dots + R_q \quad y \quad (3.49)$$

$$r_{tot} = \frac{R_{tot}}{n_i}. \quad (3.50)$$

Un estimador para κ puede ser obtenido a partir de la expresión (3.12), utilizándose $\bar{R}_{tot} = \frac{R_{tot}}{n}$ en el lugar de r , es decir:

$$\hat{\kappa} = \begin{cases} 2\bar{R}_{tot} + \bar{R}_{tot}^3 + \frac{5\bar{R}_{tot}^5}{6} & \text{si } \bar{R}_{tot} < 0.53 \\ -0.4 + 1.39\bar{R}_{tot} + \frac{0.43}{1 - \bar{R}_{tot}} & \text{si } 0.53 \leq \bar{R}_{tot} < 0.85 \\ \frac{1}{\bar{R}_{tot}^3 - 4\bar{R}_{tot}^2 + 3\bar{R}_{tot}} & \text{si } \bar{R}_{tot} \geq 0.85 \end{cases} \quad (3.52)$$

Para la construcción de un intervalo de confianza para κ debemos adoptar el método bootstrap (ver 11) del anexo 6).

2) $\tilde{\kappa} \geq 2.0$

La estimación puntual para κ es obtenida de la misma forma que en 1). El intervalo de confianza para κ se construye teniendo en cuenta los resultados asintóticos ya utilizados en la sección 3.3.4 (2). Fijado un coeficiente de confianza $\gamma = 1 - \alpha$, los límites inferior y superior del intervalo son, entonces, dada respectivamente por:

$$\begin{aligned} \kappa_I &= [1 + (1 + 3a)^{1/2}] / 4a, \\ \kappa_S &= [1 + (1 + 3b)^{1/2}] / 4b, \end{aligned}$$

donde:

$$a = (n - R_{tot}) / X_{n-q, 1-\frac{\alpha}{2}}^2$$

y

$$b = (n - R_{tot}) / X_{n-q, \frac{\alpha}{2}}^2$$

donde $X_{n-q, 1-\frac{\alpha}{2}}^2$ y $X_{n-q, \frac{\alpha}{2}}^2$ son, respectivamente los cuantiles de orden $1 - \frac{\alpha}{2}$ y

$\frac{\alpha}{2}$ de la distribución X^2 con $n - q$ grados de libertad.

CAPÍTULO IV

APLICACIÓN DE DATOS CIRCULARES

4.1. CASO: DIRECCIÓN DEL VIENTO EN LA CIUDAD DEL CUSCO

En el presente trabajo se utiliza los datos de direcciones del viento del año 2005 de la ciudad del Cusco (Tabla 4.1). Las características direccionales del viento son de gran importancia para describir el estado del tiempo y pueden ser de gran utilidad para el desarrollo de la agricultura, el servicio aéreo-comercial, etc.

El viento de superficie, se produce en la capa más baja de la atmósfera, que tiene un espesor de unos 8 kilómetros en los polos y unos 16 en el Ecuador.

El movimiento del aire es causado por el desarrollo de gradientes horizontales de presión dentro de la atmósfera, los cuales, en último término, son consecuencia de la radiación solar recibida por la tierra, la atmósfera funciona como una gigantesca máquina térmica en que la diferencia de temperatura existe entre los polos y el ecuador proporciona la energía necesaria para la circulación atmosférica.

Los movimientos verticales en la atmósfera están limitados por el equilibrio existente entre la fuerza gravitatoria y el gradiente vertical de presión (equilibrio hidrostático)

$$\frac{\partial P}{\partial t} = -\rho g$$

Los vientos verticales tienen en general velocidades varias en órdenes de magnitud menores que los horizontales.

TABLA 4.1
DIRECCIÓN DEL VIENTO POR ESTACIONES, CIUDAD DE CUSCO, AÑO
2005

Verano <i>Diciembre 22 - Marzo 21</i>	315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 292.5, 337.5, 337.5, 337.5, 337.5, 337.5, 270, 337.5, 337.5, 337.5, 337.5, 270, 292.5, 270, 337.5, 292.5, 337.5, 337.5, 337.5, 270, 337.5, 337.5, 337.5, 315, 337.5, 292.5, 292.5, 157.5, 157.5, 112.5, 112.5, 45, 90, 90, 135, 67.5, 135, 135, 135, 112.5, 112.5, 112.5, 112.5, 67.5, 112.5, 45, 112.5, 112.5, 67.5, 0, 45, 45, 45, 22.5, 0, 45, 45, 337.5, 270, 292.5, 337.5, 270, 292.5, 315, 315, 270, 292.5, 292.5, 292.5, 292.5, 270, 270, 270, 337.5, 337.5, 337.5, 315, 315, 315, 270
Otoño <i>Marzo 22 - Junio 21</i>	270, 337.5, 337.5, 337.5, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 315, 180, 202.5, 180, 202.5, 202.5, 202.5, 202.5, 180, 202.5, 157.5, 202.5, 157.5, 202.5, 202.5, 202.5, 202.5, 157.5, 247.5, 202.5, 292.5, 337.5, 337.5, 337.5, 337.5, 337.5, 270, 337.5, 337.5, 337.5, 337.5, 270, 292.5, 270, 337.5, 292.5, 337.5, 337.5, 337.5, 270, 337.5, 337.5, 337.5, 315, 337.5, 292.5, 292.5, 157.5
Invierno <i>Junio 22 - Septiembre 22</i>	157.5, 112.5, 112.5, 45, 90, 90, 135, 67.5, 135, 135, 135, 112.5, 112.5, 112.5, 112.5, 67.5, 112.5, 45, 112.5, 112.5, 67.5, 0, 45, 45, 45, 22.5, 0, 45, 45, 337.5, 270, 292.5, 337.5, 270, 292.5, 315, 315, 270, 292.5, 292.5, 292.5, 292.5, 270, 270, 270, 337.5, 337.5, 337.5, 315, 180, 202.5, 180, 202.5, 202.5, 202.5, 202.5, 180, 202.5, 157.5, 202.5, 157.5, 202.5
Primavera <i>Septiembre 23 - Diciembre 21</i>	202.5, 202.5, 202.5, 157.5, 247.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 202.5, 337.5, 337.5, 337.5, 315, 337.5, 292.5, 292.5, 157.5, 157.5, 112.5, 112.5, 45, 90, 90, 135, 67.5, 135, 135, 135, 112.5, 112.5, 112.5, 112.5, 112.5, 67.5, 112.5, 45, 112.5, 112.5, 67.5, 0, 45, 45, 45, 22.5, 0, 45, 45, 337.5, 270, 292.5, 337.5, 270, 292.5, 315, 315, 270, 292.5, 292.5, 292.5, 292.5, 270, 270, 270, 337.5, 337.5, 337.5, 315, 315, 315, 315, 315, 315, 270, 270, 337.5, 337.5, 337.5, 315, 315, 315

Fuente: Laboratorio de Energía Solar DAF – UNSAAC

4.2 SELECCIÓN DE LA MUESTRA

Para mostrar de manera didáctica el análisis descriptivo, la estimación y las pruebas de hipótesis de datos circulares que se distribuyen como una von Mises, se seleccionó en cada estación una muestra aleatoria de tamaño 15 con ayuda del software estadístico R (ANEXO 1).

TABLA 4.2
MUESTRAS ALEATORIAS DE LA DIRECCIÓN DEL VIENTO PARA CADA ESTACIÓN, CIUDAD DE CUSCO, AÑO 2005

Verano	292.5, 337.5, 270, 270, 112.5, 0, 337.5, 45, 270, 90, 112.5, 270,90, 270, 337.5
Otoño	315, 202.5, 202.5, 202.5, 270, 202.5, 315, 180, 202.5, 202.5, 315, 202.5,315, 337.5, 337.5
Invierno	337.5, 315, 202.5, 0, 315, 135, 315, 315, 337.5, 315, 112.5, 315,337.5, 270, 315
Primavera	22.5, 202.5, 337.5, 202.5, 67.5, 112.5, 202.5, 202.5, 315, 337.5, 292.5, 270, 270, 337.5, 135

4.3 REPRESENTACIÓN GRÁFICA Y ANÁLISIS DESCRIPTIVO

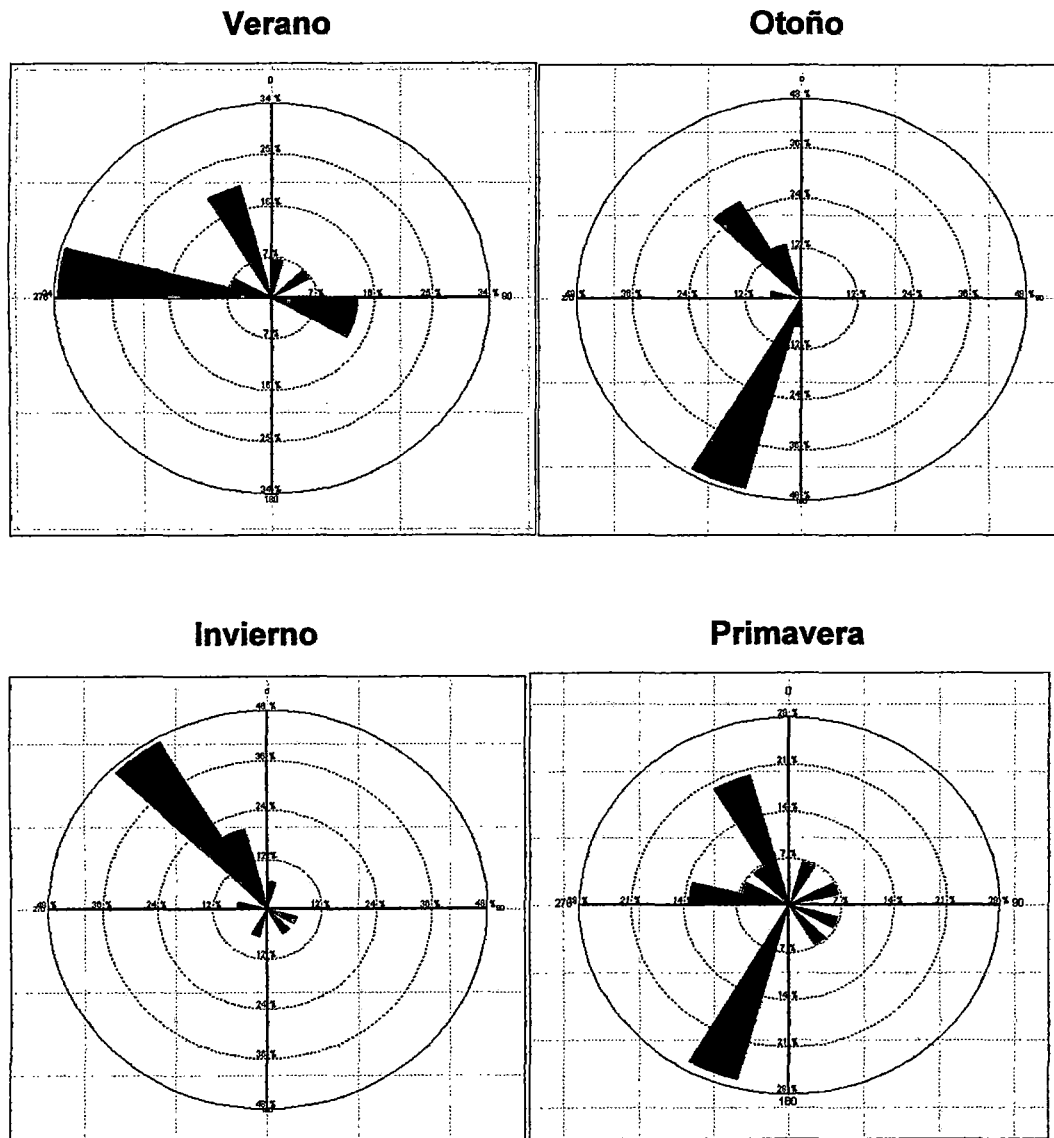
En primer lugar vamos a utilizar un análisis gráfico para verificar visualmente y como primera aproximación la uniformidad de los datos y la existencia de direcciones de preferencia, lo cual será comprobada analíticamente con la prueba de Rayleigh. Si los datos se distribuyen uniformemente no tiene sentido el análisis descriptivo ni la estimación ni la prueba de hipótesis, y si los datos tienen una dirección de preferencia procedemos a verificar si se distribuyen como una von Mises y si los datos se distribuyen como una von Mises realizamos el análisis descriptivo de los datos, la estimación de parámetro y la prueba de hipótesis.

1) REPRESENTACIÓN GRÁFICA

Las direcciones del viento se representan generalmente en sentido horario, empezando desde el norte (0°), al este (90°), al sur (180°), al oeste (270°) y de nuevo hacia el norte (360°). A continuación se muestran las gráficas mediante histogramas de las direcciones del viento de los datos de la Tabla 4.2. Dichas

gráficas fueron obtenidas mediante el software R y la librería “*VecStatGraphs2D*”, cuya sintaxis se muestra en el Anexo 2.

FIGURA 4.1
HISTOGRAMA DE MUESTRAS DE LAS DIRECCIONES DEL VIENTO
SEGUN ESTACIONES, CIUDAD DEL CUSCO, AÑO 2005



En los histogramas se observa que aparentemente las direcciones del viento de las estaciones de verano y de primavera no tienen una dirección de preferencia, mientras que en las direcciones de otoño e invierno existe una dirección de preferencia. Estas primeras conclusiones realizadas visualmente

serán verificadas analíticamente con la prueba de Rayleight en el siguiente apartado.

2) PRUEBA DE RAYLEIGH

En este apartado utilizamos la Prueba de Rayleigh, para verificar analíticamente la uniformidad de los datos circulares, en vista que el método gráfico de la sección anterior nos lleva a una conclusión preliminar sobre la existencia de direcciones de preferencia en algunas estaciones. Para ello verificamos las hipótesis:

H_0 : La distribución de datos es uniforme

H_1 : La distribución de datos es unimodal.

utilizando la función *r.test* de la librería *CircStas* de *R* (ver ANEXO 1), se obtiene los siguientes resultado.

TABLA 4.3
PRUEBA DE RAYLEIGH PARA CADA ESTACIÓN DE LAS DIRECCIONES
DEL VIENTO, CIUDAD DEL CUSCO, AÑO 2005

Estación	r	p-valor
Verano	0.3205089	0.2168784
Otoño	0.5193154	0.0149395*
Invierno	0.5933335	0.0035695*
Primavera	0.2262577	0.4715099

Al observar los p-valores concluimos que las hipótesis nulas para las direcciones del viento de las estaciones de verano ($p=0.21687849$) y primavera ($p=0.4715099$) no se rechazan, esto significa que dichas direcciones se distribuyen uniformemente, esto también se corrobora con los módulos de la dirección media $r=0.3205089$ y $r=0.2262577$, los cuales son muy pequeños, y nos indica que la direcciones del viento de verano y primavera no tienen una dirección de preferencia.

Las estaciones de otoño ($p=0.0149395$) e invierno ($p=0.0035695$) si tienen una dirección de preferencia, es decir no se distribuyen uniformemente, lo cual también se observa en los módulos de las direcciones medias, que son grandes $r=0.5193154$ y $r=0.5933335$.

A continuación calculamos las medidas descriptivas para los datos de dirección del viento de las estaciones de otoño e invierno, luego verificamos si dichos datos se distribuyen como una von Mises, para finalmente calcular las estimaciones paramétricas y verificar las pruebas de hipótesis, para los parámetros de la dirección media (μ) y de concentración (κ).

3) ANÁLISIS DESCRIPTIVO

A continuación se muestra algunas medidas de posición y dispersión muestrales obtenidos con el software R y sus librerías “CircStats” y “VecStatGraphs2D” (Anexos 1 y 2).

TABLA 4.4
ESTADÍSTICOS MUESTRALES DE LAS DIRECCIONES DEL VIENTO,
CIUDAD DEL CUSCO, AÑO 2005

Estadística	Otoño	Invierno
$\bar{\theta}_i$	249.0049	318.9098
$\tilde{\theta}_i$	270	315
r_i	0.5193	0.5933
$\hat{\delta}_i$	1.0654	0.4412
n_i	15	15

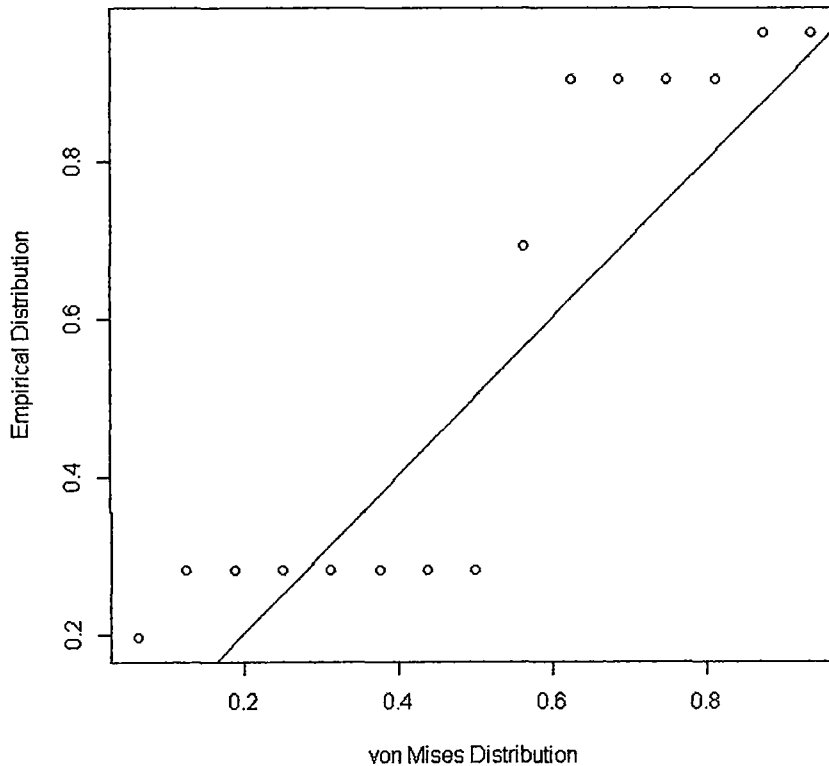
En otoño observamos que los vientos tienen una dirección promedio de 249.0049° que significa una mayor preferencia de los vientos en la dirección Sur sur oeste. Del mismo en invierno la dirección media corresponde a 318.9098° lo cual indica una dirección de preferencia de los vientos en la dirección nor oeste.

4) METODO GRÁFICO PARA PROBAR SI LOS DATOS PROCEDEN DE UNA DISTRIBUCIÓN VON MISES

El gráfico p-p se utiliza para probar si un conjunto de datos circulares se ajusta a una distribución de von Mises, esto ocurrirá si los puntos se concentran alrededor de la recta de pendiente 45°, para ello utilizamos el *software R* y la librería “*CircStats*” (Anexo 1). Al mismo tiempo *R* nos da las estimaciones de los parámetros μ y κ , donde κ , nos indica la concentración de los datos alrededor de la media circular μ .

FIGURA 4.2

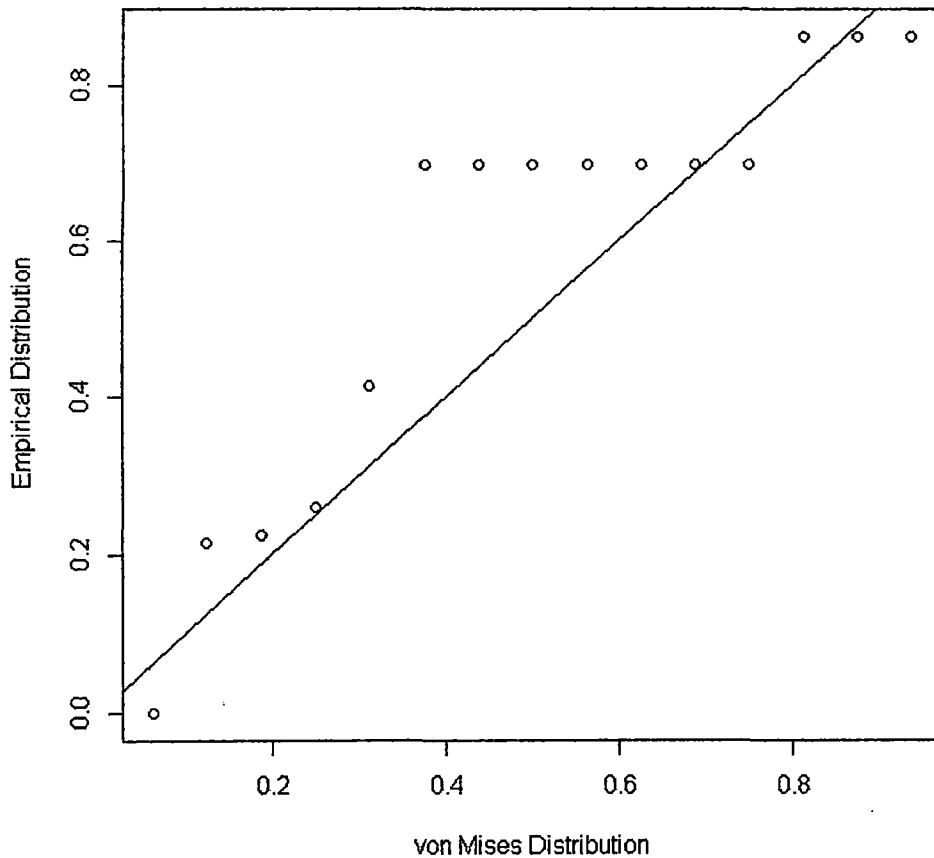
AJUSTE GRÁFICO A UNA DISTRIBUCIÓN VON MISES DE LAS DIRECCIONES DEL VIENTO DE OTOÑO, CIUDAD DEL CUSCO, AÑO 2005.



$$\hat{\mu} = 249^\circ \text{ y } \hat{\kappa} = 1.210$$

Observamos que los puntos se concentran alrededor de la recta por lo que concluimos que las direcciones del viento de invierno se distribuyen razonablemente como una von Mises, donde los parámetros estimados son $\hat{\kappa} = 1.210$ y $\hat{\mu} = 249^\circ$.

FIGURA 4.3
AJUSTE GRÁFICO A UNA DISTRIBUCIÓN VON MISES DE LAS
DIRECCIONES DEL VIENTO DE INVIERNO, CIUDAD DEL CUSCO, AÑO
2005.



$$\hat{\mu} = 318.91^\circ \text{ y } \hat{\kappa} = 1.482$$

Observamos que los puntos se concentran alrededor de la recta por lo que concluimos que las direcciones del viento de invierno se distribuyen razonablemente como una von Mises, donde los parámetros estimados son $\hat{\kappa} = 1.482$ y $\hat{\mu} = 318.91^\circ$.

4.4 INTERVALO DE CONFIANZA (Bootstrap)

Los intervalos de confianza para los parámetros de dirección μ y de concentración κ de von Mises para las muestras de direcciones del viento,

muestra 1 (Otoño) y muestra 2 (Invierno) respectivamente, se calcula con el *software R* y la librería “*CircStats*” (Anexo 1), obteniéndose:

✓ Otoño

$$\mu \in [203.49^\circ, 261.93^\circ]$$

$$\kappa \in [0.37, 5.79]$$

✓ Invierno

$$\mu \in [225.31^\circ, 325^\circ]$$

$$\kappa \in [0.32, 3.18]$$

4.5 PRUEBA DE HIPÓTESIS PARA LA IGUALDAD DE PARÁMETROS DE CONCENTRACIÓN DE DISTRIBUCIONES DE VON MISES

Como las muestras fueron extraídas de poblaciones von Mises, con los valores de los parámetros desconocidos, consideremos probar la hipótesis:

$$H_0 : \kappa_1 = \kappa_2$$

$$H_1 : \kappa_1 \neq \kappa_2$$

Para esto utilizamos la Estadística f_q para probar la hipótesis, dada en la expresión (3.47).

$$f_q = \frac{(n-q) \sum_{i=1}^q (\bar{d}_i - \bar{d})^2}{(q-1) \sum_{i=1}^q \sum_{j=1}^{n_i} (d_{ij} - \bar{d}_i)^2}, \quad q = 2 \text{ número de muestras}$$

Donde: $\bar{\theta}_i = \hat{\mu}_i$ la dirección media en la i -ésima muestra,

$$d_{ij} = |\text{sen}(\theta_{ij} - \bar{\theta}_i)|, \quad j = 1, \dots, n_i$$

$$\bar{d}_i = \sum_{j=1}^{n_i} \frac{d_{ij}}{n_i}, \quad i = 1, \dots, q \quad \text{y}$$

$$\bar{d} = \sum_{i=1}^q \frac{n_i \bar{d}_i}{n}$$

Observemos que d_{ij} puede ser considerado como el “residuo” de la j -ésima observación en el i -ésimo grupo. En la tabla 4.5, se tiene los residuos d_{ij} para $i = 1, 2, j = 1, \dots, 15$.

TABLA 4.5
TENEMOS LOS RESIDUOS DE LA j -ésima OBSERVACIÓN EN EL i -ésimo GRUPO

Para la muestra 1	Para la muestra 2
$d_{1j} = \text{sen}(\theta_{1j} - \bar{\theta}_1) , j = 1, \dots, 15$	$d_{2j} = \text{sen}(\theta_{2j} - \bar{\theta}_2) , j = 1, \dots, 15$
$d_{11} = 0.914$	$d_{21} = 0.319$
$d_{12} = 0.725$	$d_{22} = 0.068$
$d_{13} = 0.725$	$d_{23} = 0.895$
$d_{14} = 0.725$	$d_{24} = 0.657$
$d_{15} = 0.358$	$d_{25} = 0.068$
$d_{16} = 0.725$	$d_{26} = 0.068$
$d_{17} = 0.914$	$d_{27} = 0.068$
$d_{18} = 0.934$	$d_{28} = 0.068$
$d_{19} = 0.725$	$d_{29} = 0.319$
$d_{110} = 0.725$	$d_{210} = 0.068$
$d_{111} = 0.914$	$d_{211} = 0.445$
$d_{112} = 0.725$	$d_{212} = 0.068$
$d_{113} = 0.914$	$d_{213} = 0.319$
$d_{114} = 1.000$	$d_{214} = 0.754$
$d_{115} = 1.000$	$d_{215} = 0.068$

$$\bar{d}_1 = \frac{d_{11} + \dots + d_{115}}{15} = \frac{12.023}{15} = 0.802$$

$$\bar{d}_2 = \frac{d_{21} + \dots + d_{230}}{15} = \frac{4.252}{15} = 0.283$$

$$\bar{d} = \frac{n_1 \bar{d}_1 + n_2 \bar{d}_2}{30} = \frac{15(0.802) + 15(0.283)}{30} = \frac{16.275}{30} = 0.543$$

$$f_2 = \frac{(30-2)[(0.802-0.543)^2 + (0.283-0.543)^2]}{(2-1)[(0.914-0.802)^2 + (0.725-0.802)^2 + \dots + (0.068-0.283)^2]}$$

$$f_2 = \frac{28(0.135)}{1.541} = \frac{3.78}{1.541} = 2.453$$

El procedimiento para obtener la región crítica para la presente prueba de hipótesis depende de los tamaños de las muestras y de las dispersiones muestrales. En primer lugar debemos obtener $\hat{\kappa}$ a partir de la expresión (3.48):

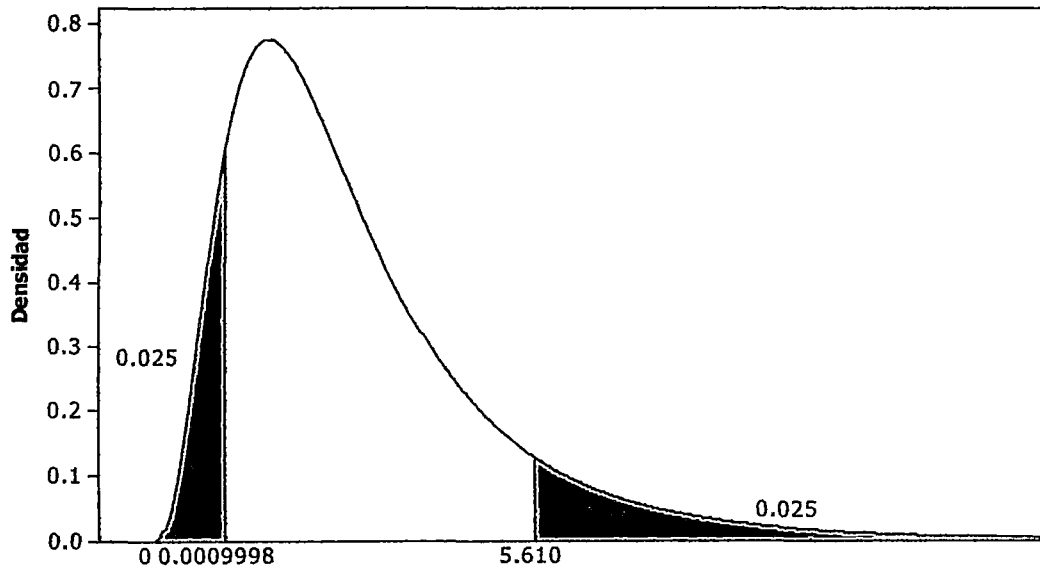
$$\tilde{\kappa} = \text{mediana}(\hat{\kappa}_1, \dots, \hat{\kappa}_q)$$

$$\tilde{\kappa} = \text{mediana}(1.2101 \text{ Y } 1.482)$$

$$\tilde{\kappa} = \frac{1.2101 + 1.482}{2} = \frac{2.692}{2} = 1.346$$

Como $n_i \geq 10$, $\hat{\kappa}_i \geq 1$, $i=1, \dots, q$, los valores críticos de la prueba, a un nivel de significancia $\alpha = 0.05$ están dadas por los cuantiles $F_{0.025;1;28}$ y $F_{0.925;1;28}$ de la distribución F, tal como se muestra en la figura 4.4.

FIGURA 4.4
REGIÓN CRÍTICA PARA LA PRUEBA DE HIPÓTESIS DE LA IGUALDAD DE
LOS PARÁMETROS DE CONCENTRACIÓN



Observamos que $f_2 \in RR(H_0)$ entonces no se debe rechazar la hipótesis nula H_0 por consiguiente se concluye que los parámetros de concentración son iguales.

Luego estimamos el parámetro de concentración común κ , obteniendo \bar{R}_{tot} y $\hat{\kappa}$.

$$R_{tot} = R_1 + \dots + R_q$$

$R_i = n_i(r_i)$ Son las magnitudes de las resultantes en cada muestra

$$R_1 = 15(0.5193) = 7.7895$$

$$R_2 = 15(0.5933) = 8.8995$$

$$R_{tot} = 7.7895 + 8.8995 = 16.689$$

$$\bar{R}_{tot} = \frac{R_1 + R_2}{n} = \frac{16.689}{30}, \text{ Entones } \bar{R}_{tot} = 0.556$$

De (3.52) vamos obtener el $\hat{\kappa}$.

$$\hat{\kappa} = \begin{cases} 2\bar{R}_{tot} + \bar{R}_{tot}^3 + \frac{5\bar{R}_{tot}^5}{6} & \text{si } \bar{R}_{tot} < 0.53 \\ -0.4 + 1.39\bar{R}_{tot} + \frac{0.43}{1 - \bar{R}_{tot}} & \text{si } 0.53 \leq \bar{R}_{tot} < 0.85 \\ \frac{1}{\bar{R}_{tot}^3 - 4\bar{R}_{tot}^2 + 3\bar{R}_{tot}} & \text{si } \bar{R}_{tot} \geq 0.85. \end{cases}$$

Como $0.53 \leq \bar{R}_{tot} < 0.85$, $\hat{\kappa} = -0.4 + 1.39\bar{R}_{tot} + \frac{0.43}{1 - \bar{R}_{tot}} \Rightarrow \hat{\kappa} = 1.341$

4.6 PRUEBA DE HIPÓTESIS PARA LA IGUALDAD DE LAS DIRECCIONES MEDIAS DE DISTRIBUCIONES DE VON MISES

En vista que los parámetros de concentración para las dos muestras son iguales, utilizamos la estadística de Watson-Williams dada en la expresión (3.33), para probar si las direcciones medias son iguales:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$$R^2 = R_1^2 + R_2^2 + 2R_1R_2 \cos(\mu_2 - \mu_1)$$

$$R^2 = (7.7895)^2 + (8.8995)^2 + 2(7.7895)(8.8995) \cos(318.9098 - 249.0049)$$

$$R^2 = 187.516$$

$$R = 13.694$$

TABLA 4.6

TABLA DE ANÁLISIS DE VARIANZA.

Fuente de variación	G.L.	Suma de Cuadrados	Cuadrados medios	F
Entre muestras	q-1 2-1=1	$\sum_{i=1}^q R_i - R$ 16.689- 13.694=2.995	$(\sum_{i=1}^q R_i - R)/(q-1) = a$ 2.995	a/b 6.305
Dentro de muestras	n-q 30-2=28	$n - \sum_{i=1}^q R_i$ 30-16.689=13.311	$(n - \sum_{i=1}^q R_i)/(n-q) = b$ 0.475	
Total	n-1=30-1=29	n-R 30-13.694=16.306		

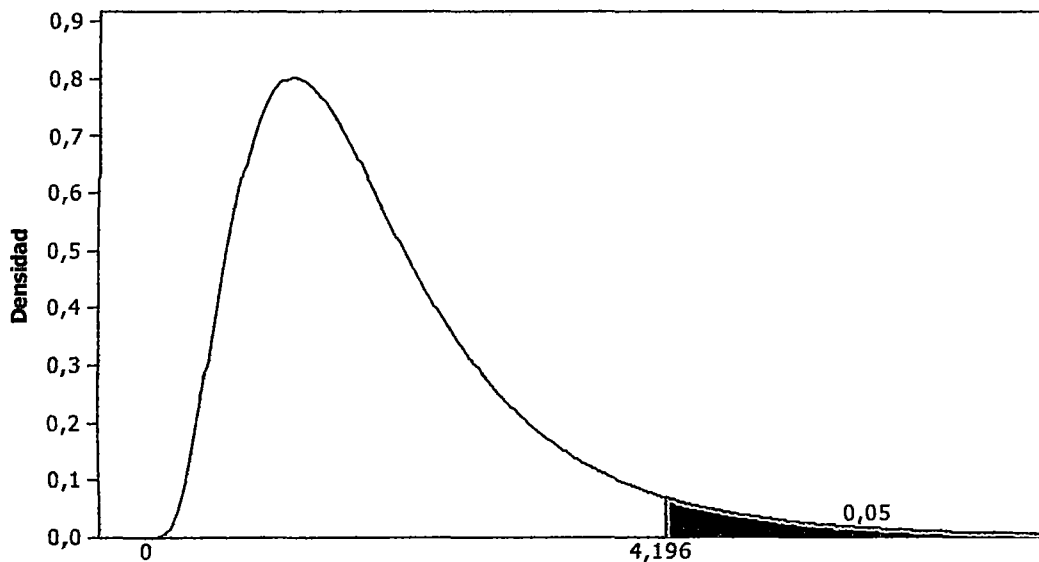
$$F^* = \left(1 - \frac{3}{8\kappa}\right)F \text{ de la expresión (3.33)}$$

$$F^* = \left(1 - \frac{3}{8(1.341)}\right)6.305$$

$$F^* = 4.542$$

F^* se aproxima a una distribución F de Snedecor con $(q-1)$ y $(n-q)$ grados de libertad. La región crítica para la distribución F con un nivel de significación de 0.05 se da en la figura 4.5.

FIGURA 4.5
REGIÓN CRÍTICA PARA LA PRUEBA DE HIPÓTESIS DE LAS DIRECCIONES MEDIAS



Observamos que $F^* \in RR(H_0)$ entonces se debe rechazar la hipótesis nula H_0 y aceptar la hipótesis alterna H_1 por consiguiente se concluye que las direcciones medias del viento en las estaciones de otoño e invierno son diferentes.

CONCLUSIONES

- 1) En el estudio de datos circulares una de las características es que pueden ser expresados en ángulos, y la representación angular o como vector unitario no es única y depende muchas veces de la elección de la dirección y el sentido de rotación por consiguiente debemos tener cuidado tomar en cuenta el punto de partida y el sentido.
- 2) La dirección promedio del viento en la ciudad del cusco en el año 2005 tiene las siguientes características: en la estación de otoño la dirección media fue de 249.0049° que indica en promedio una preferencia de los vientos en la dirección sur sur oeste, mientras que en la estación de invierno la dirección media correspondiente fue de 318.9098° indicándonos una dirección de preferencia en la dirección nor oeste y que las direcciones medias del viento en las estaciones de otoño e invierno son diferentes.
- 3) Dentro de los software estadísticos comerciales y libres, el software estadístico R con su librería *CircStats* es el más adecuado para realizar los diferentes análisis estadísticos de datos direccionales.

RECOMENDACIONES

- 1) Existen muchas aplicaciones en diferentes campos de la ciencia tales como la agricultura, biología, medicina, meteorología, etc., que utilizan datos circulares, por lo que se sugiere que la Carrera Profesional de Matemática implemente dentro de su currícula asignaturas para el estudio y tratamiento de ese tipo de datos.

- 2) Se recomienda seguir investigando en este campo por ser muy amplio, como por ejemplo en el análisis de datos bivariados o multivariados.

BIBLIOGRAFIA

- [1] Abramowitz, M. e Stegun, I.A. (1970). *Handbook of Mathematical functions*. New York: National Bureau of Standards.
- [2] Artes, R. (1997). Extensões da teoria das equações de estimação generalizadas a dados circulares e modelos de dispersão. Tese de Doutorado, IME USP.
- [3] Batschelet, E. (1981) *Circular Statistics in Biology Data*. London: Academic Press.
- [4] Cabrera, J., Schmidt-Koenig, K. e Watson, G.S. (1991). The statistical analysis of circular data. Chapter 10 (pp. 285-305) in *Human Understanding and animal Awareness*, Editors P.P.G. Bateson e P.H. Klopfer. Perspectives in Ethology Series, Volume 9. New York: Plenum Press.
- [5] Cacsire, G. D. (1997) Inferência sobre medidas de posição e dispersão em dados circulares. Tese para grau mestre em Estatística, Universidade de São Paulo-Brazil.
- [6] Conover, W. J. (1971). *Practical nonparametric Statistics*. Wiley, New York.
- [7] Fernández, M. (2012) Estimadores de mínima distância para dados circulares. Tese de Licenciatura em Matemáticas, Universidad de Buenos Aires-Argentina.
- [8] Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.
- [9] Fisher, N. I. (1986). Robust comparison of dispersion for samples of direction data.
Aust. J. Statist. **28**, 213-219.

- [10] Fisher, N. I. e Hall, P. G. (1989). Bootstrap confidence regions for directional data. *J. Am. Statist. Assoc.* **84**, 996-1002.
- [11] Fisher, N. I. e Lewis, T. (1983). Estimating the common mean direction of several circular or spherical distribution whit, differing dispersions. *Biometrika***70**, 333-41
- [12] Fisher, N. I. e Powell, C. McA. (1989). Statistical analysis of two-dimensional palaeocurrent data: Methods and Examples. *Aust. J. Earth Sci.* **36**, 91-107.
- [13] Hall, P. G. (1992). *The bootstrap and Edgeworth Expansion*. New York: Springer.
- [14] Harrison, D., Kanji, R. J. (1986). *Analysis of variance for circular data*. *J. Appl. Statist.* **13**, 123-138.
- [15] Harrison, D., Kanji, R. J. (1988). The development of analysis of variance for circular data. *J. Appl. Statist.* **15**, 197-224.
- [16] Jammalamadaka S. R. y SenGupta A. (2001) *Topics in Circular Statistics*. World Scientific, Singapore.
- [17] Lockhart, R. A. e Stephens, M. A. (1985). Tests of fit for the von Mises distribution. *Biometrika***72**, 647-52.
- [18] Mardia, K.V. (1967). A nonparametric test for the bivariate two-sample location problem. *J. Roy. Stat. Soc. B* **29**, 320-342.
- [19] Mardia, K.V. (1969). On the null distribution of a non-parametric test for the bivariate two-sample problem. *J. Roy. Stat. Soc. B* **31** 98-102.
- [20] Mardia, K.V. (1972). *Statistics of Directional Data*. Academic Press.

- [21] Neter, J. ; Kuter, M. H. ; Nachtsheim, C. J. Wasserman, W. (1996). Applied linear statistical models. London; Irwin. 1
- [22] Núñez, G. (2010) Análisis Bayesiano de Modelos Lineales para Datos Direccionales considerando la Distribución Normal bajo Proyección. Tesis de Doctorado en Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa-México.
- [23] Rao, J. S. (1966). Large sample tests for homogeneity of angular data. Appendix to a paper by S. Sengupta and J. S. Rao. *Sankhya, Ser. B*, **28**, 172-174.
- [24] Stephens, M. A. (1969). Tests for the von Mises distribution. *Biometrika***56**, 149-60.
- [25] Stephens, M. A. (1982). Use of the von Mises distribution to analyse continuous proportions. *Biometrika* **69**, 730-7.
- [26] Upton, G. J. G. (1973). Single-sample tests for the von Mises distribution. *Biometrika***60**, 87-99.
- [27] Watson, G.S. (1983). Statistics on spheres. New York: Wiley.
- [28] Wheeler, S. e Watson, G. S. (1964). A distribution-free two-samples test on a circle. *Biometrika***51**, 256-257.
- [29] Zar, J.H. (1974). Bioestatistical analysis. Prentice-Hall, Englewood Cliffs, N.J.

Páginas de internet

[30] Jaimes P. y Pineda S. J. (2012), "*Una introducción al análisis de datos circulares con algunas aplicaciones*" Tesis de licenciatura en Matemáticas, Universidad Industrial de Santander, Bucaramanga.

repositorio.uis.edu.co/jspui/bitstream/123456789/7207/2/143012.pdf

[31] Izbicki R. (2008), "*Análisis de dados circulares*" Projeto de Iniciação Científica, Instituto de Matemática e Estatística da Universidade de São Paulo - IME-USP, Brasil.

www.ime.usp.br/~rzbicki/projetoCirc.pdf

ANEXO 1

SOFTWARE ESTADÍSTICO R con la librería “*CircStats*”.

- Se pueden construir vectores. Construyamos un vector de nombre *z* y escribimos su valor en la ventana de R Console. La expresión *c()* significa el conjunto de valores dados entre paréntesis.

```
>z <- c(aquí se coloca el vector de datos) # después de este numeral podemos comentar de que trata los datos colocados.
```

- Para graficar la figura de puntos o ramas de datos circulares.

```
> circ.plot(rad(z), stack=TRUE, bins=52, shrink=2, dotsep=10, cex=0.7)
```

- Para graficar el diagrama de rosa de datos circulares.

```
> rose.diag(rad(z), bins=52, prop=2.5, pts=FALSE, shrink=2, cex=0.7)
```

```
> circ.mean(rad(z))
```

```
> deg(circ.mean(rad(z)))
```

```
> pp.plot(z)
```

```
> pp.plot(rad(z), ref.line=TRUE)
```

```
> circ.disp(rad(z))
```

Generate uniform data and create several rose diagrams.

Some optional parameters may be needed to optimize plots.

```
> data <- runif(50, 0, 2*pi)
```

```
> rose.diag(z, bins = 18, main = 'Uniform Data')
```

```
> rose.diag(z, bins = 18, main = 'Stacked Points', pts=TRUE)
```

> Generate von Mises data and create several rose diagrams.

```
> z <- rvm(25, 0, 5)
```

```
> rose.diag(z, bins=18, pts=TRUE) # Points fall out of bounds.
```

```
> rose.diag(z, bins=36, prop=1.5, pts=TRUE, shrink=1.5)
```

```
> aleatVonMises<-rvm(10,45,1) ;aleatVonMises
```

```
> aleatVonmises<-rvm(10,45,1);aleatVonmises
```

```
> circ.summary(rad(dat))
```

```
> vm.bootstrap.ci(z, bias=FALSE, alpha=0.05, reps=200, print=TRUE)
```

```
> r.test(rad(x), degree=FALSE)
```


ANEXO 2

SOFTWARE ESTADÍSTICO R con la librería “*VecStatGraphs2D*”.

```
> x <- c(aquí se coloca el vector de datos)

> DrawHistogram(x, ClassSize = 15, Direction = 1) # grafica el histograma de datos
circulares en sentido antihorario por ejemplo la dirección del viento.

> DrawDistribution(x, Direction = 1)

> DrawHistogram(x, ClassSize = 15, Direction = 2) # grafica el histograma de datos
circulares en sentido horario.

> DrawDistribution(x, Direction = 2)

> AllAzimuthStatistics(z)
```

ANEXO 3

SOFTWARE ESTADÍSTICO R con la librería "bootstrap".

```
> x <- c(aquí se coloca el vector de datos)
```

```
> vm.bootstrap.ci(x, bias=FALSE, alpha=0.05, reps=200, print=TRUE)
```

Para Encontrar intervalos de confianza no paramétricos Bootstrap.

```
> city.boot <- boot(city, ratio, R = 999, stype = "w", sim = "ordinary")
```

```
boot.ci(city.boot, conf = c(0.90, 0.95), type = c("norm", "basic", "perc", "bca"))
```

```
> boot.ci(boot.out, conf = 0.95, type = "all", index = 1:min(2,length(boot.out$t0)),  
var.t0 = NULL, var.t = NULL, t0 = NULL, t = NULL, L = NULL, h = function(t) t, hdot =  
function(t) rep(1,length(t)), hinv = function(t) t, ...)
```

ANEXO 04

Salida de R de estadística descriptiva e intervalos de confianza para las direcciones de Otoño

```
> y <- c(315, 202.5, 202.5, 202.5, 270, 202.5, 315, 180, 202.5, 202.5, 315, 202.5, 315, 337.5, 337.5)
```

```
[1] "*****"
[1] "****"
[1] "****" CIRCULAR STATISTICS - AZIMUTH "****"
[1] "****"
[1] "*****"
[1] "NUMBER OF OBSERVATIONS= 15"
[1] "MEAN AZIMUTH = 249.0049"
[1] "MEAN MODULE = 0.5193"
[1] "CIRCULAR STANDARD DEVIATION = 1.1448"
[1] "CIRCULAR VARIANCE = 0.4807"
[1] "VON MISES PARAMETER = 1.2101"
[1] "CIRCULAR DISPERSAL = 1.0654"
[1] "SKEWNESS COEFFICIENT OF AZIMUTH = -0.7826"
[1] "KURTOSIS COEFFICIENT OF AZIMUTH = -1.769"
```

```
> vm.bootstrap.ci(y, bias=FALSE, alpha=0.05, reps=200, print=TRUE)
Bootstrap Confidence Intervals for Mean Direction and Concentration
Confidence Level: 95 %
Mean Direction: Low = 0.41 High = 1.43
Concentration Parameter: Low = 0.37 High = 5.79
$mu.ci
 2.5% 97.5%
0.4143672 1.4315515
```

```
$kappa.ci
 2.5% 97.5%
0.3716413 5.7931878
```

```
$mu.reps
[1] 0.3445837 0.4161565 0.4852267 0.5784930 0.5950466 0.6024249 0.6456615 0.6781043 0.6944424
[10] 0.7253868 0.7338107 0.7338107 0.7485320 0.7619778 0.7907881 0.7964006 0.7964006 0.7964006
[19] 0.8233980 0.8332639 0.8607526 0.8691541 0.8803497 0.8822615 0.8827989 0.8848523 0.8856962
[28] 0.8882814 0.8882814 0.8942978 0.8980878 0.9016602 0.9069561 0.9087047 0.9102153 0.9159822
[37] 0.9244737 0.9363928 0.9399703 0.9442628 0.9475167 0.9537544 0.9553959 0.9601878 0.9601878
[46] 0.9618979 0.9712268 0.9712268 0.9767342 0.9784558 0.9785432 0.9809919 0.9813474 0.9835940
[55] 0.9988275 0.9997775 1.0030645 1.0051311 1.0131144 1.0143386 1.0143386 1.0210527 1.0272167
[64] 1.0272167 1.0272167 1.0272167 1.0406097 1.0422863 1.0462742 1.0488565 1.0488565 1.0580018
[73] 1.0623822 1.0623822 1.0623822 1.0641314 1.0661960 1.0661960 1.0685255 1.0715384 1.0752172
[82] 1.0801007 1.0881935 1.1019132 1.1034970 1.1041283 1.1087840 1.1102868 1.1143007 1.1149615
[91] 1.1149615 1.1149615 1.1180938 1.1207663 1.1207663 1.1276504 1.1276504 1.1303325 1.1350175
[100] 1.1448055 1.1458404 1.1467016 1.1499655 1.1499655 1.1551816 1.1551816 1.1570269 1.1599247
[109] 1.1608218 1.1608218 1.1638434 1.1675029 1.1675029 1.1675029 1.1675029 1.1700208 1.1705717
[118] 1.1748599 1.1770297 1.1770297 1.1770297 1.1770297 1.1838363 1.1838363 1.1838363 1.1838798
[127] 1.1838798 1.1839487 1.1978511 1.1978511 1.1988419 1.1988419 1.2034718 1.2038805 1.2038805
[136] 1.2083608 1.2097993 1.2097993 1.2147654 1.2147654 1.2173230 1.2173230 1.2173230 1.2173230
[145] 1.2209131 1.2235270 1.2235270 1.2235270 1.2235270 1.2300854 1.2300854 1.2300854 1.2325306
[154] 1.2397432 1.2416701 1.2416701 1.2421064 1.2467777 1.2501242 1.2501242 1.2517246 1.2517246
[163] 1.2564451 1.2569280 1.2569280 1.2629540 1.2642787 1.2642787 1.2721375 1.2727134 1.2738540
[172] 1.2790273 1.2790273 1.2790273 1.2821018 1.2977310 1.2977310 1.3003491 1.3003491 1.3039844
[181] 1.3198305 1.3229168 1.3246800 1.3391745 1.3586972 1.3788749 1.3856648 1.3856648 1.4131913
[190] 1.4279927 1.4310940 1.4493928 1.4613587 1.4983875 1.7426640 1.9230770 5.8933450 5.9820615
[199] 6.1672311 6.2600433
```

```
$kappa.reps
[1] 6.0640203 2.2501191 1.7693511 3.7164285 0.5624163 0.9791511 0.9480825 1.2198442
[9] 1.3885543 0.3729170 2.0339510 1.3906224 1.0395419 0.5875141 1.4238000 2.7668437
[17] 2.6604858 0.4461757 0.4641809 0.6551644 1.7779824 0.8432248 0.8569502 0.5573792
[25] 1.0089154 0.3729170 2.2289642 1.0162188 0.6613647 1.3929309 1.2198442 2.2321614
[33] 1.5484225 1.4238000 2.6438333 0.9966881 0.9898734 3.5349189 1.2105934 1.2290447
[41] 1.4558610 1.1178276 1.0776689 2.2289642 0.8640286 0.8683823 2.2299206 4.0032222
[49] 0.9362243 0.8605135 1.7375309 0.9793283 3.2722664 1.4099445 1.7525992 0.8554886
[57] 0.9821124 1.5908628 1.2215779 0.3719345 1.2290447 0.8640286 1.7475195 1.0862071
[65] 2.6379451 1.7375309 0.9821124 0.9847176 0.8569502 1.7779824 1.1778695 0.5252585
[73] 1.5424461 0.8640286 1.2290701 1.8897219 0.5482438 0.3983214 1.3114166 0.6703553
```

Estimación de Parámetros y Pruebas de Hipótesis en Datos Circulares

[81]	1.7692999	0.3460880	0.7179161	0.5624163	1.7368276	2.2974975	1.7474903	1.0559936
[89]	0.5126435	0.6293050	3.4386977	1.2215645	2.2357499	2.2501191	1.7375384	1.4021303
[97]	0.5480020	0.8824733	0.6653863	3.5811555	0.3602085	1.2334536	5.7755362	1.0776689
[105]	0.5624163	1.2555686	1.5288142	0.8554820	1.4238000	1.4238000	1.2215779	3.2382696
[113]	6.4012848	0.7792814	1.0089154	2.2843439	0.2737328	2.6400235	1.1261603	2.2974975
[121]	1.5288142	0.9793283	0.3983214	0.4093518	1.3942076	1.1178276	1.3942076	0.9944401
[129]	1.4099445	1.4099445	1.7392351	0.6653863	1.3929309	1.4454423	0.6653863	1.5586207
[137]	2.3706141	2.6604858	1.4099445	1.2188872	0.8683823	0.7639989	13.0037672	1.4164363
[145]	1.7375384	1.5373802	1.4850916	0.5624163	1.9937433	2.0033145	2.0086581	2.2289642
[153]	1.5288142	0.9014275	2.1546331	0.4483758	0.3809792	0.6762591	0.8735697	1.9951322
[161]	0.8825162	0.8776670	1.7392351	1.7693511	1.1657651	2.6400422	1.5586207	2.2321614
[169]	1.4073789	0.6558078	5.9091641	1.7474903	0.7076806	0.9821124	1.7375384	0.1988568
[177]	3.2592336	6.4012848	1.5760164	5.7902140	4.1157604	5.7902140	1.3891640	2.0151728
[185]	1.7475195	1.0089154	0.2540822	1.7419003	1.5288142	0.3983214	1.1035017	1.2188872
[193]	0.5480001	1.2254357	1.4021303	0.8569502	1.3891640	1.3891640	2.6400235	1.7392213

ANEXO 05

Salida de R de estadística descriptiva e intervalos de confianza para las direcciones de Invierno

```
> vm.bootstrap.ci(z, bias=FALSE, alpha=0.05, reps=200, print=TRUE)
Bootstrap Confidence Intervals for Mean Direction and Concentration
Confidence Level: 95 %
Mean Direction:          Low = -0.78   High = 0.96
Concentration Parameter: Low = 0.32   High = 3.18
$mu.ci
  2.5%      97.5%
-0.7801076 0.9580485
```

```
$kappa.ci
  2.5%      97.5%
0.3185635 3.1764485
```

```
$mu.reps
[1] 0.001679086 0.009789622 0.018559850 0.038077122 0.050149255 0.050776859
[7] 0.054794144 0.075483263 0.085602700 0.103827101 0.106366318 0.118480402
[13] 0.122649643 0.124540435 0.155363906 0.160432945 0.161076172 0.168149776
[19] 0.171877452 0.173230994 0.173837651 0.187186636 0.190417499 0.202563189
[25] 0.202708159 0.209000367 0.209000367 0.220057656 0.235661655 0.242786103
[31] 0.245140997 0.246436839 0.252588381 0.257640711 0.258382626 0.279511373
[37] 0.283039400 0.289638061 0.290175669 0.297975230 0.300208276 0.314970111
[43] 0.319211488 0.320458619 0.322664427 0.325418871 0.346496680 0.352836661
[49] 0.362837790 0.365650823 0.390320589 0.392237803 0.393631571 0.402462613
[55] 0.406770867 0.408330090 0.410868421 0.416230168 0.419916716 0.423871589
[61] 0.424004854 0.424004854 0.434027234 0.434993660 0.435425413 0.439187162
[67] 0.440239102 0.440819994 0.441772783 0.446224903 0.446636376 0.446636376
[73] 0.449291938 0.450768983 0.451772281 0.458259205 0.458259205 0.458259205
[79] 0.462526764 0.470070953 0.484062133 0.488388862 0.491108904 0.492244571
[85] 0.497354981 0.498122923 0.501108110 0.507544955 0.510220408 0.512570067
[91] 0.512570067 0.518018303 0.530816429 0.532033912 0.535194621 0.537238374
[97] 0.537443572 0.549090482 0.551970246 0.560198248 0.561931259 0.561931259
[103] 0.565874502 0.568004937 0.569332943 0.574247097 0.583332762 0.586956552
[109] 0.587987971 0.594853558 0.595741549 0.597922699 0.609972447 0.610853559
[115] 0.617828172 0.640255213 0.643360869 0.645914085 0.645914085 0.665689123
[121] 0.671578578 0.671983549 0.680449060 0.680545580 0.682089902 0.685598943
[127] 0.696806266 0.700262301 0.702494631 0.703338782 0.705465759 0.708992105
[133] 0.717037245 0.717037245 0.725370288 0.731136364 0.737793295 0.742537779
[139] 0.752791862 0.753877172 0.769262975 0.779337905 0.784791006 0.796516909
[145] 0.817351357 0.821182714 0.825585072 0.833352224 0.843269579 0.854408962
[151] 0.886526750 0.912821914 0.914617304 0.929562380 0.931220449 0.949103366
[157] 0.951944727 0.957213773 0.990604693 0.990604693 1.034678814 1.064277021
[163] 1.144099621 4.924716212 5.236263681 5.286639508 5.347538907 5.387524459
[169] 5.506040613 5.530185623 5.564593716 5.609373736 5.620673516 5.624590399
[175] 5.772057559 5.811706525 5.860160540 5.861599452 5.861599452 5.877844394
[181] 5.947072588 5.951405563 6.014826686 6.018438347 6.027780559 6.062762757
[187] 6.062762757 6.081275701 6.083617253 6.091675675 6.114153014 6.133067923
[193] 6.135556596 6.157488787 6.165475641 6.191559524 6.212098232 6.222117213
[199] 6.269404435 6.269618862
```

```
$kappa.reps
[1] 1.1155271 2.8848872 4.1712944 0.5705052 0.8532711 0.6871824 1.2084828
[8] 0.7700133 1.3625433 1.0122310 1.0324670 1.0467757 1.0237424 1.7769478
[15] 0.8978741 0.8062099 1.0864719 1.0925645 0.5449282 0.2204151 1.7200948
[22] 1.7531828 0.4394857 2.3416591 1.3959478 0.8919095 1.5425408 0.8581673
[29] 0.6850353 1.0550229 0.5560623 1.4651708 3.7782368 1.4135591 0.8308988
[36] 1.6086442 0.8782490 0.6722534 0.6697485 1.2376466 1.5752274 1.8160973
[43] 3.1360758 2.2458630 0.8782490 1.0700498 1.2307483 1.2299845 3.1762263
[50] 1.1172087 0.8594139 1.3519923 1.2248878 0.8827844 0.7986639 1.3426277
[57] 1.6854004 1.8885060 0.7564950 1.7054881 0.3865490 0.7487746 1.4551331
[64] 0.7803944 0.6874954 1.3400691 0.4766998 1.2106536 0.7314680 1.0463728
[71] 1.2614286 0.7514297 0.4571625 2.2593349 1.0467757 0.2939949 0.8539082
[78] 0.9543897 0.8679019 1.1072366 1.4161186 0.5469930 2.2215237 1.4275076
[85] 1.9340281 2.0662776 1.3326682 0.7678384 0.6663711 1.3023947 0.5570936
[92] 1.2411502 1.3426277 2.5871360 0.8499729 0.7518411 0.9824703 0.7609359
[99] 4.5387076 0.2277086 1.6447275 0.8880697 0.7688158 0.6850353 1.0169203
[106] 0.6640459 1.1368159 1.2735162 0.3110655 1.3381557 1.3400691 1.2311334
[113] 0.9046630 0.4394699 0.7324369 0.3547471 0.4282517 1.4339291 1.1015830
[120] 0.6983854 0.6977739 1.4540375 4.2508068 2.8893045 1.8080139 0.8621365
[127] 2.8065407 1.2039126 0.8382057 1.2088936 0.7841291 2.1702561 1.8923200
```

Estimación de Parámetros y Pruebas de Hipótesis en Datos Circulares

[134]	2.2037257	0.9479595	0.4193189	1.7455613	0.2331089	0.9786786	0.6978562
[141]	0.7915747	2.0355052	0.4424803	0.7455950	0.8010765	0.9361713	1.9561444
[148]	0.6207778	1.0181554	0.9106686	1.0521320	0.6777639	1.0194636	1.4151146
[155]	1.3122740	1.0974834	0.5386239	1.0008434	1.0989750	1.1372391	0.8341312
[162]	0.4585412	1.1145646	0.9204813	1.7076180	0.7579816	0.4157973	0.9996507
[169]	0.9981537	1.7769478	1.2691437	0.6134636	1.6677752	0.3975253	1.5790049
[176]	0.8626814	2.2088141	1.2888819	1.1799945	1.1199553	0.6392612	0.5255867
[183]	1.9281396	0.3554907	0.7794566	1.1148461	1.4135591	1.1730717	0.6308641
[190]	0.7782226	1.0541151	1.4338740	3.1851127	0.6702350	1.3316478	0.3187557
[197]	1.5196952	1.5610212	0.5438289	0.6404562			

ANEXO 6

MÉTODOS BOOTSTRAP

Los métodos bootstrap se refieren a una clase de procedimientos estadísticos computacionales que son frecuentemente útiles en la realización de pruebas de hipótesis, o en la obtención de medidas de variabilidad de estimadores puntuales, en situaciones en las cuales procedimientos estadísticos analíticos no son válidos o conocidos.

Efron y Tibshirani (1986), describieron la técnica bootstrap de la siguiente forma: supongamos que los datos observados.

$$\theta = (\theta_1, \dots, \theta_n)$$

Consisten en observaciones de una variable aleatoria Θ con función de distribución desconocida F en Ω , el espacio muestral común de las observaciones. Tenemos una estadística de interés, digamos $X(\theta)$, cuya distribución es desconocida. A partir de un proceso de remuestreo de las observaciones, y utilizando un método de Monte Carlo, podemos generar la distribución de $X(\theta)$ de la siguiente forma:

- **Paso 1: Remuestreo.** Usando un generador de números aleatorios, retiramos, independientemente, una muestra bootstrap, digamos:

$$\theta_1^*$$

- **Paso 2: Estimador bootstrap.** En la muestra bootstrap calculamos la estadística de interés, obteniendo:

$$X(\theta_1^*)$$

- **Paso 3: Repetición.** Repetimos los pasos 1 y 2 un gran número de veces, digamos B veces, obteniendo las muestras bootstrap:

$$\theta_1^*, \dots, \theta_B^* \text{ y}$$

los B valores observados de la estadística:

$$(\theta_1^*), \dots, (\theta_B^*)$$

- **Paso 4: Intervalo de confianza.** Ordenamos las estadísticas bootstrap en orden creciente y obtenemos los percentiles de acuerdo con el coeficiente de confianza deseado. A partir de los intervalos así construidos podemos también realizar pruebas de hipótesis.

Informaciones adicionales sobre los datos pueden ser incorporadas al procedimiento, de forma que se pueda mejorar.

Describiremos en las próximas secciones técnicas bootstrap que pueden ser aplicadas a diferentes problemas relacionados con los datos circulares.

1) Intervalo de confianza para la dirección media μ

Dada una muestra aleatoria $\theta_1, \dots, \theta_n$, generar una muestra bootstrap siguiendo los siguientes pasos:

- **Paso 1: Remuestreo.** Generar n números aleatorios u_1, \dots, u_n a partir de la distribución uniforme $U[0,1]$, y calcular $i_1 = \text{parte entera de } nu_1 + 1, \dots, i_n = \text{parte entera de } nu_n + 1$.

La muestra bootstrap es:

$$\theta_I^* = (\theta_{i_1}^*, \dots, \theta_{i_n}^*)$$

- **Paso 2:** calcular el estimado bootstrap $\hat{\mu}_1$ utilizando la expresión (2.2).
- **Paso 3:** repetirlos pasos 1 y 2 hasta obtener un total de B estimaciones de bootstrap de la dirección media; $\hat{\mu}_1, \dots, \hat{\mu}_B$.
- **Paso 4:** construcción del intervalo de confianza.

Obtener

$$v_b = \hat{\mu}_b - \bar{\theta}, b = 1, \dots, B.$$

donde $\bar{\theta}$ es la media en la muestra original, y reordenamos los v_b en orden creciente,

obteniendo: $v_{(1)} \leq \dots \leq v_{(B)}$.

Para α dado, encontrar $m = \text{parte entera de } \frac{B\alpha}{2}$ y $l = B - m$. El intervalo de confianza es:

$$(\mu_l, \mu_s), \text{ donde } \mu_l = v_{(m+1)} + \bar{\theta}, \mu_s = v_l + \bar{\theta}.$$

2) Prueba de hipótesis sobre la dirección media μ en problemas con una muestra

Nuestro objetivo es probar la hipótesis $H_b: \mu = \mu_0$, donde μ_0 es una dirección especificada. En la muestra original, calculamos:

$$\psi_1 = \theta_1 - (\bar{\theta} - \mu_0), \dots, \psi_n = \theta_n - (\bar{\theta} - \mu_0),$$

donde $\bar{\theta}$ es la dirección media muestral.

- **Paso 1: Remuestreo.** Encontrar una muestra bootstrap considerando como muestra original ψ_1, \dots, ψ_n .
- **Paso 2:** calcular la estadística $T = T(\theta_1^*, \dots, \theta_n^*, \bar{\theta})$ de (4.8).
- **Paso 3:** repetirlos pasos 1 y 2 hasta obtener un total de B valores de la estadística, es decir T_1, \dots, T_B .
- **Paso 4: Región crítica bilateral.** Reorganizarlos valores $|T_1|, \dots, |T_B|$ en orden creciente obteniendo $T_{(l)} \leq \dots \leq T_{(B)}$. Sea α el nivel de significancia de la prueba $m_\alpha = \text{parte entera de } B(1-\alpha) + 1$. Rechaza la H_0 , si $|T| > |T_{(m_\alpha)}|$, donde T es calculado.

3) Comparación de las direcciones medias de dos o más poblaciones

Nuestro objetivo es probar la hipótesis

$$H_0: \mu_1 = \dots = \mu_q = \mu$$

a partir de las q muestras observadas:

$$\theta_{i_1}, \dots, \theta_{i_q}, \quad i = 1, \dots, q.$$

Obtener:

$$\psi_{i_1} = \theta_{i_1} - \bar{\theta}_i, \dots, \psi_{i_q} = \theta_{i_q} - \bar{\theta}_i, \quad i = 1, \dots, q,$$

donde $\bar{\theta}_i$ es la dirección media observada en la i-ésima muestra.

- **Paso 1:** a partir de cada muestra

$$\psi_{i1}, \dots, \psi_{in_i}, \quad i = 1, \dots, q.$$

generar una muestra bootstrap considerando el Paso 1 en 4.15.1, obteniendo

$$\theta_i^* = (\theta_{i1}^*, \dots, \theta_{in_i}^*), \quad i = 1, \dots, q.$$

- **Paso 2:** calcular la estadística Y_q
 - **Paso 3:** repetirlos pasos 1 y 2 hasta obtener un total de B valores de la estadística, Y_1^*, \dots, Y_B^* .
 - **Paso 4:** reordenarlos valores $|Y_1^*|, \dots, |Y_B^*|$ en orden creciente obteniendo $Y_{(1)}^*, \dots, Y_{(B)}^*$.
- Sea α el nivel de significancia de la prueba es $m_\alpha = \text{parte entera de } B(1-\alpha) + 1$.

Rechaza H_0 si $|Y_q| > Y_{(m_\alpha)}^*$.

4) Estimación de la dirección media común de dos o más distribuciones

Tenemos $q \geq 2$ muestras de datos circulares:

$$\theta_{i1}, \dots, \theta_{in_i}, \quad i = 1, \dots, q.$$

de distribuciones unimodales con la misma dirección media μ . La forma de combinar las muestras depende de los tamaños de las muestras y de las dispersiones muestrales será o no de la misma magnitud. Adoptando el método adecuado, calculamos inicialmente $\hat{\mu}$ de la forma descrita en esta sección. Para construir el intervalo de confianza para μ , con coeficiente de confianza $\gamma = 1 - \alpha$, adoptamos el siguiente procedimiento:

- **Paso 1:** volver a muestrear cada muestra obteniendo:

$$\theta_i^*, \quad i = 1, \dots, q.$$

- **Paso 2:** calcular el estimado de la dirección media común μ , digamos $\hat{\mu}_1^*$ adoptando el mismo método empleado para el cálculo de $\hat{\mu}$.
- **Paso 3:** repetirlos pasos 1 y 2 hasta obtener un total de B estimaciones de la dirección media,

$$\hat{\mu}_1^*, \dots, \hat{\mu}_B^*.$$

- **Paso 4:** obtener

$$v_b = \hat{\mu}_b^* - \hat{\mu}, \quad b = 1, \dots, B. \text{ y,}$$

reordenar los v_b en orden creciente, obteniendo; $v_{(1)} \leq \dots \leq v_{(B)}$.

Para α dado, encontrar $m = \text{parte entera de } \frac{B\alpha}{2}$. El intervalo de confianza es:

(μ_l, μ_s) , donde

$$\mu_l = v_{(m+1)} + \hat{\mu}, \quad \mu_s = v_{(l)} + \hat{\mu}, \quad \text{y } l = B - m.$$

5) Intervalo de confianza para la dirección media μ de una $VM(\mu, \kappa)$

Dadas las estimaciones $\hat{\mu}$ (2.2), $\hat{\kappa}$ (3.12) obtenidas de la muestra original, proceder de la siguiente forma:

- **Paso 1: Remuestreo.** Simular n valores $\theta_1^*, \dots, \theta_n^*$ a partir de una $VM(\hat{\mu}, \hat{\kappa})$

Los demás pasos son análogos a los pasos 2, 3 y 4 escritos en 1).

6) Prueba de hipótesis sobre la dirección media de una población $VM(\mu, \kappa)$

Considerando el proceso remuestreo descrito en 5) procedemos como en 2) considerando ahora la estadística T_n dada en (3.47).

7) Intervalo de confianza para el parámetro de concentración en una $VM(\mu, \kappa)$

Para determinar un intervalo de confianza para κ , considerar el proceso de remuestreo descrito en 4) para la construcción de B muestras bootstrap. En cada muestra obtener $\hat{\kappa}_i^*, i=1, \dots, B$, que ordenados en forma creciente producen:

$$\hat{\kappa}_{(1)}^* \leq \dots \leq \hat{\kappa}_{(B)}^*.$$

Un intervalo de confianza para κ con un coeficiente de confianza $\gamma = (1 - \alpha)$ es:

$(\hat{\kappa}_{(l)}^*, \hat{\kappa}_{(m)}^*)$ donde $l =$ parte entera de $(B\alpha/2 + 1/2)$ y $m = B - l$.

8) Prueba de hipótesis sobre el parámetro de concentración en una $VM(\mu, \kappa)$

Nuestro objetivo es probar $H_0: \kappa = \kappa_0$. Esta prueba es realizada considerando el intervalo de confianza bootstrap en 7). Rechace H_0 si κ_0 no pertenece al intervalo $(\hat{\kappa}_{(l)}^*, \hat{\kappa}_{(m)}^*)$.

9) Prueba de igualdad de direcciones medias de dos o más distribuciones von Mises

Tenemos q muestras:

$$\theta_{i1}, \dots, \theta_{in_i}, \quad i = 1, \dots, q$$

de la distribución von Mises. En cada muestra calculamos $\hat{\mu}_i, \hat{\kappa}_i, i = 1, \dots, q$.

- **Paso 1: Remuestreo.** Simular una muestra $\psi_{i1}, \dots, \psi_{in_i}$ de tamaño n_i a partir de una

$VM(0, \hat{\kappa}_i), i = 1, \dots, q$ es obtener:

$$\theta_{i1}^* = \psi_{i1} + \hat{\mu}_i, \dots, \theta_{in_i}^* = \psi_{in_i} + \hat{\mu}_i, \quad i = 1, \dots, q$$

- **Paso 2: Estadística bootstrap.** Calcular el valor de la estadística Y_q dada en la expresión (3.28).

Los siguientes pasos son análogos a los pasos 3 y 4 en 3).

10) Estimación de la dirección media común de 2 o más distribuciones von Mises

- **Paso 1: Remuestreo.** A partir de cada muestra $\theta_{i1}, \dots, \theta_{in_i}, i = 1, \dots, q$, construir una muestra $(\theta_{i1}^*, \dots, \theta_{in_i}^*), i = 1, \dots, q$ de la misma forma que en el paso 1 de 9).

Estimación de Parámetros y Pruebas de Hipótesis en Datos Circulares

- **Paso 2: Estimación bootstrap.** Calcular la estimación bootstrap $\hat{\mu}_w^*$, para este conjunto de q muestras, utilizando la expresión (3.36).

Los siguientes pasos son análogos a los pasos 3 y 4 en 4).

11) Estimación del parámetro de concentración común de dos o más distribuciones de von Mises

- **Paso 1: Remuestreo.** A partir de las muestras originales, construimos muestras bootstrap de la misma forma que en el paso 1 de 9).

- **Paso 2: Estimador bootstrap de K .** Calcular el estimador \hat{K}_1^* para este conjunto de q muestras en la misma manera que \hat{K} fue calculado en la expresión (3.50).

- **Paso 3: repetir los pasos 1 y 2 B veces,** obteniendo K_1^*, \dots, K_B^* , que ordenados producen $K_{(1)}^* \leq \dots \leq K_{(B)}^*$. Un intervalo de confianza para K con un coeficiente de confianza $\gamma = (1 - \alpha)$ es:

$$\left(\hat{K}_{(l+1)}^*, \hat{K}_{(m)}^* \right),$$

donde $l = \text{parte entera de } \left(B \frac{\alpha}{2} + \frac{1}{2} \right)$ y $m = B - l$.