



CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS
“MODELACIÓN ESTADÍSTICA DE DATOS PAREADOS
PARA EVALUAR EL EFECTO DEL HURACÁN JOVA EN
ÁRBOLES DE CHAMELA”

T E S I S

PARA OBTENER EL TÍTULO DE:
MAESTRO EN CIENCIAS CON ESPECIALIDAD EN
PROBABILIDAD Y ESTADÍSTICA

PRESENTA:

JOHNATAN ADONAI GARCÍA LÓPEZ

DIRECTORA:

DRA. ELOÍSA DÍAZ-FRANCÉS MURGUÍA

GUANAJUATO, GUANAJUATO, FEBRERO DE 2015.



CIMAT
CENTRO DE INVESTIGACION
EN MATEMÁTICAS A. C.

Centro de Investigación en Matemáticas, A.C.

Acta de Examen de Grado

Acta No.: 087
Libro No.: 002
Foja No.: 087

En la Ciudad de Guanajuato, Gto., siendo las 12:30 horas del día 5 de febrero del año 2015, se reunieron los miembros del jurado integrado por los señores:

DR. ENRIQUE RAÚL VILLA-DIHARCE (CIMAT)
DR. ROGELIO RAMOS QUIROGA (CIMAT)
DRA. ELOÍSA DÍAZ-FRANCÉS MURGUÍA (CIMAT)

bajo la presidencia del primero y con carácter de secretario el segundo, para proceder a efectuar el examen que para obtener el grado de

MAESTRO EN CIENCIAS CON ESPECIALIDAD EN PROBABILIDAD Y ESTADÍSTICA

sustenta

JOHNATAN ADONAI GARCÍA LÓPEZ

en cumplimiento con lo establecido en los reglamentos y lineamientos de estudios de posgrado del Centro de Investigación en Matemáticas, A.C., mediante la presentación de la tesis

"MODELACIÓN ESTADÍSTICA DE DATOS PAREADOS PARA EVALUAR EL EFECTO DEL HURACÁN JOVA EN ÁRBOLES DE CHAMELA "

Los miembros del jurado examinaron alternadamente al (la) sustentante y después de deliberar entre sí resolvieron declararlo (a):

APROBADO

DR. ENRIQUE RAÚL VILLA DIHARCE
Presidente

DR. ROGELIO RAMOS QUIROGA
Secretario

DRA. ELOÍSA DÍAZ-FRANCÉS MURGUÍA
Vocal



CIMAT
DIRECCION
GENERAL

DR. JOSÉ ANTONIO STEPHAN DE LA PEÑA MENA
Director General

**“Modelación estadística de datos pareados
para evaluar el efecto del huracán JOVA en
árboles de Chamela”**

Johnatan Adonai García López

Febrero de 2015

Dedicatoria

*A mis padres Doroteo y Lucía
y a todos mis hermanos
Marcos, Joseluis, Nelson,
Noel Ivan, Alondra, Geovani,
Paul, Monserrat y Ximena.*

Agradecimientos

A Dios, sólo a Él sea la gloria por los siglos de los siglos, por haberme sostenido en todo el trayecto de este proyecto y haber puesto los medios necesarios para culminar con éxito este trabajo.

Al Centro de Investigación en Matemáticas, A.C. (CIMAT) y a todos los investigadores y personal, por permitirme ser parte de ustedes adquiriendo tanto conocimiento como valores.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT), por su apoyo de beca de maestría con la que me permitió desarrollar y concluir una parte importante de mi formación profesional.

A la Dra. Eloísa Díaz-Francés Murguía por aceptar dirigir esta tesis, por todo el apoyo incondicional tanto académico como moral recibido durante la realización de este trabajo. Gracias por cada uno de los consejos, ánimos, y correcciones tras correcciones para poder alcanzar un objetivo juntos.

A proyecto SEP/CONACYT Ref. 179045: “*Respuesta del Socio-ecosistema del Bosque Tropical Seco de la Región de Chamela al Huracán Jova: un evento catastrófico infrecuente*”, gracias por la invitación a participar y contribuir con este trabajo.

Al Dr. Manuel Maass Moreno por facilitarnos amablemente la base de datos de perímetros

IV

de tallos por árbol de un cuadrante de muestreo y explicarnos las características más relevantes de éstos, ya que gracias a su colaboración fue posible proponer modelos estadísticos adecuados.

Al biólogo y M. C. Luis Alfredo Pérez Jiménez(†), al ingeniero Salvador Araiza Méndez, al biólogo Luis Raúl Ahedo Hernández y a Abel Berduzco Robles por su ardua labor en la recolección y depuración de los datos.

A mis sinodales Dr. Enrique Villa Diharce, Dr. Rogelio Ramos Quiroga y Dra. Eloísa Díaz-Francés Murguía por su tiempo, comentarios y correcciones hechas para mejorar este trabajo.

Al lector especial Dr. Manuel Maass Moreno por su importante contribución en la interpretación de resultados de este trabajo.

A la Dra. Angelina Martínez Yrizar y Dr. Diego Pérez Salicrup por sus comentarios bien acertados en cuanto a las variables utilizadas en esta tesis.

A Enriquena Bustamante por el apoyo técnico con el manejo de presupuesto y logística del proyecto.

A mis padres Doroteo y Lucía por que por ellos soy lo que soy y a quienes amo con todas mis fuerzas, quienes siempre estuvieron apoyándome y sobre todo creyeron en mí a pesar de todo. Aunque pasamos momentos malos siempre coincidieron en que debía seguir hasta terminar.

A la gran cantidad de hermanos que Dios me concedió tener, Marcos, Joseluis, Nelson, Noel Ivan, Alondra, Geovani, Paul, Monserrat y Ximena, por su apoyo económico y moral, confianza y amor que siempre me han tenido. Por que en las buenas y en las malas nunca

hemos estado separados, los amo. A mi hija Ximena que siempre me echó porras y me escribía cartas para no sentirme solo.

A mis compañeros, con quienes compartí dos años de mi vida y que juntos llegamos a la meta.

A mis amigos y amigas Adán, Ernesto, Miriam, Merie y Mariana por toda su confianza y amistad.

¡Muchas gracias a todos!

Prefacio

Los desastres naturales como huracanes, sequías, tsunamis, temblores, incendios y otros conllevan muchas consecuencias importantes: ecológicas, económicas, sociales, epidémicas, financieras y políticas. Por ello, el ser humano se ve en la necesidad de evaluar y cuantificar el daño producido por estos fenómenos en diferentes formas para poder prevenir o hacer frente en situaciones similares futuras.

A primera vista puede parecer sencillo la evaluación del daño del impacto de un huracán sobre la vegetación contabilizando los tallos caídos. Sin embargo esto no es así, sobre todo cuando se trata de un ecosistema muy diverso en el que una gran variedad de especies de árboles con historias de vida y proceso evolutivo muy distintos comparten un mismo espacio y tiempo. Evaluar el daño diferencial entre las especies no sólo es interesante sino además indispensable si se quiere lograr un entendimiento del proceso de perturbación a nivel ecosistémico.

En esta tesis se proponen modelos estadísticos para estimar la proporción de árboles dañados en 17 especies de árboles diferentes, en un área de interés en el bosque tropical caducifolio de la costa occidental de México. Asimismo, se desea evaluar el efecto producido por el huracán Jova y eventos ocurridos entre mayo de 2011 y mayo de 2012 en cuanto al crecimiento de área basal en árboles no dañados sobre la Reserva de la Biosfera de Chamela-Cuixmala (RBChC) en octubre de 2011.

Se describirá en detalle el proceso de modelación estadística que se utilizó para estos fines.

Se mostrará también la conexión que los resultados pueden tener con los patrones de lluvias anuales y con el fenómeno de la Oscilación del Sur, El Niño.

La tesis consta de cuatro capítulos, los cuales se han estructurado de la siguiente manera. En el primer capítulo se presentan los conceptos y antecedentes básicos de meteorología y ecología para comprender el contexto de los datos de mediciones de árboles de la RBChC, ubicada en el estado de Jalisco, que se consideraron y modelaron estadísticamente. En el segundo capítulo se presentan y explican los conceptos relevantes de estadística necesarios para el proceso de modelar estadísticamente las diferencias en el área basal de árboles en la región de Chamela y para estimar el número de árboles dañados por especie. En el tercer capítulo se presenta el análisis estadístico propuesto que se aplicó a los datos de árboles de Chamela. Esto incluye el planteo de modelos estadísticos adecuados, la estimación de los parámetros, la validación de los modelos propuestos y la interpretación de los resultados para dar respuesta a las preguntas de interés. Finalmente, como último capítulo se presentan las conclusiones generales de este trabajo. Todo el trabajo computacional se realizó con el lenguaje de programación estadístico gratuito R¹. Algunos de los códigos usados están disponibles a los lectores interesados tras escribir al autor a Johnatan.garcia@cimat.mx.

Los modelos estadísticos propuestos en esta tesis serán de ayuda para entender mejor el crecimiento de área basal en árboles de las 17 especies que se analizaron en la reserva ecológica de Chamela y para evaluar el daño y recuperación de los árboles tras el paso del huracán a nivel destructivo medio categoría 3 y eventos ocurridos durante el periodo en que se tomaron las mediciones. De esta manera se podrán tomar decisiones pertinentes e inteligentes que conduzcan a conservar los ecosistemas y a mejorar la relación con las comunidades humanas que dependen de ellos.

¹<http://www.r-project.org/>

Lista de figuras

2.1. Verosimilitud relativa de θ	19
2.2. Verosimilitud relativa de la reparametrización $\phi = \theta^{-\frac{1}{3}}$	19
2.3. Verosimilitud relativa cuando $0 < t < nk$	23
2.4. Verosimilitud relativa cuando $t = 0$	23
2.5. Gráfica Q-Q de sumas de áreas basales de tallos por árbol de la especie <i>Plumeria rubra</i> antes del paso del huracán JOVA.	30
3.1. Histograma de frecuencias del número de especies con su máximo número de tallos.	44
3.2. Intervalos de verosimilitud para la proporción de daño p_j por especie.	48
3.3. Gráfica Q-Q de sumas de áreas basales de tallos por árbol de la especie <i>Cordia alliodora</i> antes del paso del huracán Jova.	50
3.4. Gráfica Q-Q de sumas de áreas basales de tallos por árbol no dañado de la especie <i>Cordia alliodora</i> después del paso del huracán Jova.	51
3.5. Gráfica Q-Q de diferencias $X_{ij}^* - X_{ij}$ de la especie <i>Cordia allidora</i>	53
3.6. Comparación entre los modelos normal uni y bivariado en cuanto a la diferencia de medias	58
B.1. <i>Colubrina heteroneura</i> 2011	69
B.2. <i>Colubrina heteroneura</i> 2012	69
B.3. <i>Cordia alliodora</i> 2011	70

B.4. <i>Cordia alliodora</i> 2012	70
B.5. <i>Adelia oaxacana</i> 2011	70
B.6. <i>Adelia oaxacana</i> 2012	70
B.7. <i>Exostema caribaeum</i> 2011	70
B.8. <i>Exostema caribaeum</i> 2012	70
B.9. <i>Piptadenia constricta</i> 2011	71
B.10. <i>Piptadenia constricta</i> 2012	71
B.11. <i>Bauhinia unguolata</i> 2011	71
B.12. <i>Bauhinia unguolata</i> 2012	71
B.13. <i>Caesalpinia eriostachys</i> 2011	71
B.14. <i>Caesalpinia eriostachys</i> 2012	71
B.15. <i>Comocladia engleriana</i> 2011	72
B.16. <i>Comocladia engleriana</i> 2012	72
B.17. <i>Plumeria rubra</i> 2011	72
B.18. <i>Plumeria rubra</i> 2012	72
B.19. <i>Guapira macrocarpa</i> 2011	72
B.20. <i>Guapira macrocarpa</i> 2012	72
B.21. <i>Guettarda elliptica</i> 2011	73
B.22. <i>Guettarda elliptica</i> 2012	73
B.23. <i>Bursera instabilis</i> 2011	73
B.24. <i>Bursera instabilis</i> 2012	73
B.25. <i>Helietta lottiae</i> 2011	73
B.26. <i>Helietta lottiae</i> 2012	73
B.27. <i>Tabebuia impetiginosa</i> 2011	74
B.28. <i>Tabebuia impetiginosa</i> 2012	74
B.29. <i>Lonchocarpus eriocarinalis</i> 2011	74
B.30. <i>Lonchocarpus eriocarinalis</i> 2012	74
B.31. <i>Samyda mexicana</i> 2011	74

LISTA DE FIGURAS

XI

B.32. <i>Samyda mexicana</i> 2012	74
B.33. <i>Tabebuia chrysantha</i> 2011	75
B.34. <i>Tabebuia chrysantha</i> 2012	75

Lista de tablas

2.1. Niveles de confianza aproximados de los intervalos de verosimilitud cuando θ es unidimensional.	22
2.2. Diseño pareado	32
2.3. Sueldos mensuales en miles de pesos de cuatro matrimonios.	32
3.1. Medias estimadas y algunos cuantiles empíricos de las sumas de áreas basales de los tallos por árbol (X_{ij}) para antes y después de Jova. La columna μ_{x^*} indica el promedio de áreas basales antes de Jova pero sólo para los árboles que no sufrieron daño.	42
3.2. Número de tallos por árbol por especie (Y_{ij}).	43
3.3. Número de árboles dañados por especie y número de tallos que tenían los dañados antes y después de Jova.	45
3.4. Número de árboles dañados (Z_j) por especie, $j = 1, 2, \dots, 17$	47
3.5. EMVs e intervalos del 14.65 % y del 95 % de confianza para la proporción, p_j , de árboles dañados por especie.	49
3.6. EMVs e intervalo del 95 % de confianza para la diferencia de medias normales de la suma de áreas basales de árboles por especie, antes y después de Jova.	54
3.7. EMVs e intervalos de verosimilitud perfil para los parámetros de la normal bivariada para la suma de áreas basales de tallos para el antes y después de Jova.	56

3.8. Intervalo de verosimilitud-confianza para la diferencia de medias con el mejor modelo para las 17 especies consideradas	57
3.9. EMVs de la normal bivariada para las sumas de áreas basales de tallos para dos parejas de años consecutivos (2000-2001) y (2011-2012).	60
3.10. Lluvias anuales acumuladas en milímetros para dos parejas de años consecutivos.	61
C.1. Resultados obtenidos en la prueba de Mardia aplicada a las 17 especies consideradas.	77

Contenido

Prefacio	VII
Lista de figuras	IX
Lista de tablas	XIII
1. Preliminares	1
1.1. Introducción	1
1.2. Descripción del problema ecológico	1
1.3. Conceptos meteorológicos básicos	3
1.4. Conceptos ecológicos básicos	5
2. Conceptos estadísticos relevantes	9
2.1. Proceso de modelar estadísticamente un fenómeno aleatorio	9
2.2. Modelos estadísticos usados	11
2.2.1. La distribución Binomial	11
2.2.2. La distribución Normal	12
2.2.3. La distribución Normal Bivariada	13
2.3. La función de verosimilitud para estimar los parámetros de un modelo estadístico	14
2.3.1. Propiedad de invarianza frente a reparametrizaciones uno a uno	18
2.3.2. Intervalos de verosimilitud-confianza	20
2.3.3. Función de verosimilitud perfil	25

2.4. Validación de modelos	26
2.4.1. La importancia de los cuantiles y de la función de distribución empírica para validar modelos	27
2.4.2. La prueba de Mardia para normalidad multivariada	29
2.5. Diseño de muestras pareadas	32
2.6. Métodos estadísticos tradicionales de análisis para datos pareados	34
2.6.1. La distribución Normal para las diferencias	34
2.6.2. La distribución Normal Bivariada	35
3. Análisis estadístico	39
3.1. Descripción detallada de los datos	39
3.2. Estadísticas descriptivas relevantes	41
3.3. Modelación de la proporción de daño por especie en Jova con un modelo binomial	46
3.4. Modelo normal para el área basal de tallos antes y después de Jova	50
3.5. Estimación del crecimiento exclusivamente en árboles no dañados	52
3.5.1. Modelo normal para las diferencias $(X_{ij}^* - X_{ij})$	52
3.5.2. Modelo normal bivariado	54
3.5.3. Comparación entre modelos utilizados	57
3.6. Comparación entre dos pares de años hidrológicos similares en cuanto a pa- trones de lluvia	59
4. Conclusiones generales	63
A. La distribución Normal Multivariada	65
B. Gráficas cuantil-cuantil de mediciones en mayo de 2011 y mayo de 2012	69
C. Resultados de la prueba de Mardia	77
Referencias bibliográficas	79

Capítulo 1

Preliminares

1.1. Introducción

En este capítulo se hará la descripción del problema de interés en el contexto ecológico, se proporcionan los conceptos tanto meteorológicos como ecológicos básicos que serán de ayuda para este trabajo. Se incluye con detalle la información de donde provienen los datos.

1.2. Descripción del problema ecológico

La región de Chamela, Jalisco se localiza en la costa occidental de México. El clima que rige la región es tropical, cálido subhúmedo con una marcada estacionalidad. La temperatura promedio anual es de $24.6^{\circ}C$, la máxima promedio es de $30.3^{\circ}C$ y la mínima promedio es de $20.2^{\circ}C$ ¹. La sequía se presenta usualmente de noviembre a junio, algunas veces interrumpida por lluvias ligeras o fuertes en diciembre o enero, fenómeno conocido como canícula. El promedio de días con lluvia apreciable al año es de 52, con 5 episodios de lluvias fuertes aproximadamente por año².

¹García-Oliva et al. (2002).

²<http://www.ibiologia.unam.mx/ebchamela/>

La vegetación dominante de la región es la selva baja caducifolia o bosque tropical caducifolio, la cual alcanza una altura alrededor de 10m donde la mayoría de las plantas se quedan sin hojas en la sequía. Por ello, se nota una marcada diferencia en el paisaje entre la temporada de secas y la de lluvias. Entre los géneros de árboles más abundantes se encuentran: *Lonchocarpus*, *Caesalpinia* y *Cordia*. Los arroyos sostienen una selva más alta, la cual presenta árboles de talla mayor y una composición distinta a la anterior. Dentro de estas comunidades son abundantes las epífitas³ y es común observar a los líquenes⁴ cubriendo los troncos de muchos árboles. Los bejucos son importantes tanto por su abundancia como por su diversidad y explican el porqué los árboles mueren “en pie”.

La Estación de Biología de Chamela (EBC) del Instituto de Biología de la Universidad Nacional Autónoma de México (IBUNAM) está ubicada dentro de la RBChC junto a la costa del estado de Jalisco, México. En octubre de 2011, el ojo del huracán nombrado JOVA pasó justamente sobre la reserva de Chamela. Debido a la gran cantidad de mediciones que se tienen de todo tipo en la EBC desde su fundación en 1971, se decidió aprovechar esta información para evaluar el efecto del huracán en varios aspectos del ecosistema afectado. Para entender bien este efecto de JOVA sobre los árboles de Chamela se obtuvo el apoyo importante de los biólogos y ecólogos participantes en el proyecto CONACyT “*Respuesta del Socio-ecosistema del Bosque Tropical Seco de la Región de Chamela al Huracán Jova: un evento catastrófico infrecuente*”.

El botánico y M.C. Luis Alfredo Pérez Jiménez(†)⁵ fue un pionero en la fundación y el desarrollo de investigación en la EBC, además de ser investigador titular del IBUNAM. Colaboró en numerosos estudios ecológicos, particularmente como parte del “Proyecto Cuencas” -un estudio ecosistémico de largo plazo sobre la estructura y funcionamiento de las selvas

³Epífita: cualquier planta que crece sobre otro vegetal usándolo solamente de soporte, pero que no lo parasita. En ocasiones son llamadas plantas aéreas.

⁴Líquenes: organismos que surgen de la simbiosis entre un hongo y un alga.

⁵12 de enero de 1942 - 27 de marzo de 2009.

bajas. Además de su contribución significativa en lo concerniente a aspectos taxonómicos de la flora de la región, uno de sus legados científicos más importante lo constituye una base de datos única, de más de 25 años, sobre crecimiento de árboles en parcelas permanentes. A su muerte, para continuar con esta labor el Dr. Manuel Maass Moreno del Centro de Investigaciones en Ecosistemas (CIEco) UNAM con apoyo del biólogo Luis Raúl Ahedo Hernández decidieron rescatar la base de datos y seguir registrando información de más años. Ellos facilitaron amablemente los datos de varios años de perímetros de tallos de árboles de diferentes especies de un cuadrante de muestreo de la EBC y explicaron aspectos importantes a considerar a la hora de realizar el análisis estadístico.

1.3. Conceptos meteorológicos básicos

Un **ciclón tropical** es un término meteorológico usado para referirse a un sistema tormentoso caracterizado por una circulación cerrada alrededor de un centro de baja presión y que produce fuertes vientos y abundante lluvia. Los ciclones tropicales extraen su energía de la condensación de aire húmedo, produciendo fuertes vientos. Se distinguen de otras tormentas ciclónicas, como las bajas polares, por el mecanismo de calor que las alimenta, que las convierte en sistemas tormentosos de núcleo cálido. Dependiendo de su fuerza y localización, un ciclón tropical puede llamarse depresión tropical, tormenta tropical, huracán, tifón o simplemente ciclón⁶.

Los ciclones tropicales se desarrollan sobre extensas superficies de agua cálida oceánica y pierden su fuerza cuando penetran en tierra. Esta es una de las razones por la que las zonas costeras son dañadas de forma significativa, mientras que las regiones interiores están relativamente a salvo de recibir fuertes vientos. Sin embargo, las fuertes lluvias pueden producir inundaciones o deslaves tierra adentro y las inundaciones costeras pueden producir otras más fuertes hacia el interior en llanuras litorales extensas y de pendiente escasa.

⁶Veáse Melo y Vargas (2003).

Uno de los factores importantes a tener en cuenta cuando se consideran aspectos relacionados con los huracanes es **el fenómeno de la Oscilación del Sur El Niño**⁷ (OSEN). Este es un fenómeno con ciclos de tres a cinco años que se debe al movimiento de rotación terrestre y, por consiguiente, el desplazamiento de las mareas del hemisferio norte al hemisferio sur, siempre dentro de la zona intertropical. El ciclo global del OSEN tiene dos extremos: una fase cálida del mar a nivel superficial cerca de las costas del Pacífico en Perú conocida como El Niño y una fase fría, conocida como La Niña. Cuando el fenómeno no se presenta se dice que se trata de un año Normal.

Otro concepto importante es el **año hidrológico**, el cual se define a partir del inicio de la temporada de lluvias para finalizar con el de secas. El año hidrológico representa una unidad temporal mucho más natural para modelar las lluvias en contraste con el año calendario. Este no es muy común, sin embargo se ha utilizado con éxito en Ortega Ulloa (2013) para analizar estadísticamente las lluvias en Chamela, por lo cual es útil rescatar este concepto para los fines considerados.

Cada fase tiene características particulares en cuanto al clima, cantidad de lluvia en distintos periodos del año, la orientación de las corrientes marítimas, etc. En Chamela, se ha visto que el OSEN afecta de manera importante los patrones de lluvias anuales tanto en la cantidad de lluvia anual como en la distribución de las lluvias en los meses del año. Asimismo existen variaciones en las temporadas usuales en las cuales llueve más o hay sequía, según sea el tipo de año⁸.

Según aspectos tales como la intensidad, presencia o ausencia del fenómeno, los años calendario e hidrológico se clasifican en tres tipos: Niño, Niña y Normal. En esta tesis se

⁷El Niño-Southern Oscillation, ENSO

⁸Veáse Ortega Ulloa (2013).

analizará el efecto de OSEN sobre el crecimiento de los árboles en una parcela delimitada considerada en Chamela que se ha observado por más de treinta años.

En particular, en la región de Chamela durante el fenómeno de la Niña, cae la mayor parte de la lluvia anual de junio a septiembre y en el resto del año casi no llueve. En los años tipo Niño, en el verano suele llover menos que en los otros tipos de años, pero llueve mucho más en el invierno en comparación a los años Niña y normal.

El periodo de ocurrencia del huracán Jova fue del 5 al 12 de octubre de 2011 alcanzando una intensidad máxima de vientos de 205 km/h con rachas de 250 km/h⁹. Empezó como una depresión tropical a partir de una baja presión originada en una zona marítima ubicada al sur del estado de Guerrero, donde se dieron condiciones favorables que permitieron su formación registrada con el número 10-E de la temporada de 2011 en el Pacífico Noroeste. Se desplazó hacia el Sur-Suroeste de Manzanillo, Colima hasta convertirse en huracán el día 8 de octubre por la tarde. Alcanzó la Categoría III en la escala Saffir-Simpson, a 400 km del Suroeste de Manzanillo, Colima. Llegó a tierra en la costa del estado de Jalisco como huracán de Categoría II. Desde su formación, su duración fue de 168 horas (7 días), tiempo en el que recorrió una distancia aproximada de 2,000 km a una velocidad promedio de 12 km/h.

1.4. Conceptos ecológicos básicos

Cuando se evalúan los recursos de un ecosistema se deben tomar más que simples tomas de datos o mediciones; implica todo un análisis a la luz de éstos respecto una situación económica, ecológica o social en que se encuentre el desarrollo del recurso. Estas evaluaciones de recursos naturales suelen ser costosas y requieren una justificación objetiva, que con frecuencia

⁹CONAGUA, <http://www.conagua.gob.mx>, <http://smn.cna.gob.mx/ciclones/tempo2011/pacifico/Jova-p11.pdf>

abarca la función económica y ecológica de los ecosistemas, el uso potencial de la información y los usuarios potenciales de la misma¹⁰.

Dentro de los estudios ecológicos en un ecosistema para la evaluación de sus recursos se encuentra la medición y cálculo del crecimiento o cambio en árboles. Lo cual conlleva a proporcionar información cuantitativa acerca de los mismos, necesaria para el manejo, la planeación y la investigación.

La evaluación de los recursos ecológicos es importante por tres razones:

- a) Los recursos a pesar de ser un recurso natural renovable tienen un ritmo de crecimiento que puede ser superado por la tasa de aprovechamiento de los mismos.
- b) La cuantificación de los recursos permite la toma de decisiones en cuanto a la optimización del uso de suelo, incluida en los planes de manejo forestal.
- c) El conocimiento de los recursos permite definir planes de desarrollo regional integrales que incluyen el crecimiento en el sector industrial y de infraestructura productiva y apoyo a las comunidades rurales.

El **crecimiento** en árboles es el aumento de su masa en el tiempo (en número de árboles, o también el aumento de ciertas características como altura, diámetro, área basal, copa, raíces, valor económico, etc). En este trabajo se considera el crecimiento en el sentido del área basal de los árboles en el cuadrante de muestreo considerada.

En esta tesis se define y se hace uso del concepto de árbol dañado como sigue.

Definición (Árbol dañado). *Se dice que un árbol está dañado si en la medición hecha después del paso del huracán JOVA, perdió uno o más de sus tallos con respecto a la medición anterior. Esto incluye que el árbol completo haya caído al piso.*

¹⁰Veáse Melo y Vargas (2003)

En cualquier otro caso se dirá que el árbol no está dañado. Se define daño en lugar de mortandad, pues se sabe que en los ecosistemas de bosque tropical caducifolio, como es el caso de estudio, los árboles de muchas especies tienen una gran capacidad de recuperarse por medio del rebrote, por lo que la pérdida de algún tallo, no necesariamente representa la muerte del individuo. Además, el daño causado por un huracán es principalmente sobre la copa del árbol, dejando la raíz frecuentemente inalterada, lo que facilita su recuperación por rebrote.

Este concepto servirá al momento de cuantificar el daño causado por Jova en cada una de las especies de árboles consideradas. Sin embargo, existe incertidumbre sobre la causa del daño mismo y cabe la posibilidad que el huracán Jova no fuese causa del daño particular de algunos de los árboles estudiados aquí.

Definición (Área basal). *Se define el área basal (AB) como la superficie de la sección transversal del árbol. Se calcula a partir de los perímetros p_i , $i = 1, 2, \dots, k$, de sus k tallos, donde el perímetro (P) de los árboles se mide a 1.30 m de altura. Sea A_i el área basal del i -ésimo tallo del árbol la cual se define como*

$$A_i = \frac{p_i^2}{4\pi}.$$

Entonces, el área basal del árbol se define como

$$AB = \sum_{i=1}^k \frac{p_i^2}{4\pi} = \frac{1}{4\pi} \sum_{i=1}^k p_i^2.$$

Nótese que esta área es una aproximación a la verdadera área obtenida al cortar transversalmente los tallos, ya que se está suponiendo que los cortes de los tallos son exactamente circulares, lo cual es razonable.

Capítulo 2

Conceptos estadísticos relevantes

En este capítulo se incluyen los conceptos estadísticos relevantes para el desarrollo del presente trabajo. Se puede revisar en mayor detalle las definiciones y resultados aquí utilizados en Sprott (2000), Pawitan (2001), Kalbfleisch (1985), Dobson (2002) y Triola (2000).

2.1. Proceso de modelar estadísticamente un fenómeno aleatorio

En la inferencia estadística se desea encontrar un modelo estadístico que explique bien un fenómeno aleatorio de interés. Puede haber más de un modelo razonable, por lo que la selección del mejor modelo, que puede no ser único, debe ser clara y basada en argumentos sólidos.

Un **modelo estadístico** es un conjunto de modelos probabilísticos que se consideran razonables para describir un fenómeno aleatorio. Por ejemplo, la distribución normal con media θ y varianza unitaria podría ser un modelo estadístico adecuado para describir las alturas de los varones adultos en Guanajuato. El parámetro desconocido θ habrá que estimarlo con base en mediciones que se hagan a hombres guanajuatenses.

Con base en los trabajos de Box (1980) y Sprott (2000 Sección 1.3), los siguientes pasos describen el proceso de modelar estadísticamente un fenómeno natural aleatorio y que se seguirá en esta tesis:

- a) Obtener observaciones del fenómeno aleatorio natural de interés.
- b) Plantear un modelo estadístico que sea razonable, de acuerdo a las características que se observen en los datos del fenómeno. Generalmente se denota con $F(x; \theta)$ una familia de distribuciones específicas con parámetros desconocidos θ .
- c) Verificar si es posible combinar experimentos que se hayan realizado para obtener información sobre el comportamiento de este fenómeno. Los experimentos deben ser homogéneos en el sentido que proporcionen información sobre el mismo parámetro de interés.
- d) Estimar el vector de parámetros θ vía máxima verosimilitud, a través de estimadores puntuales de θ , además de intervalos y regiones de estimación.
- e) Validar el modelo estadístico con la estimación realizada en (d). En caso de que el modelo no describa bien al fenómeno de interés habrá que modificar el modelo considerado o incluso cambiarlo por otro, regresando al punto (b).
- f) Comparar las inferencias sobre los parámetros de interés y las conclusiones obtenidas, bajo otros modelos estadísticos razonables que también se considere describan bien al fenómeno de interés.

Estos pasos forman parte de un proceso iterativo cuya finalidad es encontrar al modelo más sencillo que describa bien a los datos observados. Dicho modelo no tiene porqué ser único, puesto que los modelos son aproximaciones a la realidad. En general, se preferirán modelos simples para describir el fenómeno aleatorio de interés. Entre más simple sea el modelo estadístico, siempre que sea razonable y explique bien a los datos, será más fácil

de entender, interpretar y usar al momento de hacer predicciones acerca del fenómeno bajo estudio.

2.2. Modelos estadísticos usados

2.2.1. La distribución Binomial

Considérese una variable aleatoria X que es el número de éxitos que ocurren en n ensayos independientes Bernoulli¹, donde la probabilidad de éxito en cada ensayo es θ y la probabilidad de fracaso es $1 - \theta$. Se dice que la distribución de dicha variable es Binomial de parámetros n y θ .

Se denotará que la variable aleatoria X sigue una distribución como $X \sim Bin(n, \theta)$ y la función de probabilidad asociada es,

$$f(x; n, \theta) = P(X = x; n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \text{ para } x \in \{0, 1, \dots, n\}, 0 < \theta < 1. \quad (2.1)$$

La función de distribución es,

$$F(x; n, \theta) = P(X \leq x) = \sum_{i=0}^x \binom{n}{i} \theta^i (1 - \theta)^{n-i} \text{ para } x \in \{0, 1, \dots, n\}, 0 < \theta < 1. \quad (2.2)$$

El valor esperado es $E(X) = n\theta$ y la varianza es menor o igual que el valor esperado y es $Var(X) = n\theta(1 - \theta)$. La distribución Binomial pertenece a la familia exponencial de distribuciones, lo cual facilita substancialmente la estimación de sus parámetros.

Un ejemplo de un fenómeno natural que da lugar a una variable aleatoria con distribución Binomial que se considerará en este trabajo es el número de árboles dañados de una especie

¹Una variable aleatoria **Bernoulli** es una variable aleatoria con dos posibles resultados $\{0, 1\}$. Si la variable toma el valor uno se dice que resultó un éxito, en caso contrario, un fracaso. La probabilidad de éxito de una variable aleatoria Bernoulli es θ .

dada que tiene n individuos en la parcela observada.

2.2.2. La distribución Normal

La Normal o Gaussiana es una de las distribuciones de probabilidad continuas que con más frecuencia se utiliza para describir fenómenos reales.

Si X es una variable aleatoria que sigue una distribución Normal con parámetro de localización $\mu \in \mathbb{R}$ y de escala $\sigma > 0$, la función de densidad asociada es,

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right], \text{ para } x \in \mathbb{R}. \quad (2.3)$$

La forma de esta función de densidad es acampanada y simétrica respecto al parámetro de localización que también es su valor esperado, $E(X) = \mu$. La varianza está dada por $Var(X) = \sigma^2$. Para esta distribución tanto la moda como la mediana son iguales a la media. Esta distribución también pertenece a la familia exponencial de distribuciones.

Un ejemplo de una variable aleatoria X que sigue una distribución Normal en esta tesis es la suma de áreas basales de tallos por árbol de una especie dada. Nótese que esta variable aleatoria es suma de variables aleatorias, por lo cual el siguiente teorema será relevante en este trabajo, el cual se conoce como Teorema de Límite Central (TLC).

El TLC asegura que, bajo ciertas condiciones, la suma de variables aleatorias X_1, X_2, \dots, X_n se distribuye aproximadamente como una Normal cuando n es suficientemente grande, sin importar cuál haya sido la distribución original de las variables.

Teorema 2.2.1 (Teorema de Límite Central). *Sean X_1, X_2, \dots, X_n variables aleatorias, independientes e idénticamente distribuidas con media μ y varianza finita positiva $\sigma^2 < \infty$.*

Sea $S_n = X_1 + X_2 + \dots + X_n$, entonces

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mu \right) \leq x \right] = \Phi(x).$$

Donde $\Phi(x)$ es la función de distribución Normal estándar.

La importancia práctica de este teorema radica en que la función de distribución Normal puede usarse para modelar fenómenos aleatorios que sean sumas de variables aleatorias con distribuciones distintas. También permite usar la distribución normal como aproximación de algunas otras funciones de distribución. A continuación se enuncian dos ejemplos.

- a) La distribución binomial de parámetros n y θ , con θ no demasiado cercano a 1 o 0, se distribuye como una normal con media $n\theta$ y varianza $n\theta(1 - \theta)$. Algunos libros recomiendan usar esta aproximación sólo si $n\theta \geq 5$ y $n\theta(1 - \theta) \geq 5$; en cuyo caso se debe aplicar una corrección de continuidad.
- b) La distribución de Poisson para valores grandes de su valor esperado λ es aproximadamente Normal con media y varianza igual a λ .

2.2.3. La distribución Normal Bivariada

Definición (Distribución normal bivariada). Sea (X, Y) un vector aleatorio continuo, bidimensional definido en el plano euclideo \mathbb{R}^2 . Se dice que (X, Y) tiene una distribución Normal Bivariada si su función de densidad de probabilidad conjunta está dada por la expresión siguiente:

$$f(x, y; \theta) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\}, \quad (2.4)$$

donde $\mu_x = E(X)$, $\mu_y = E(Y)$, $\sigma_x^2 = V(X)$, $\sigma_y^2 = V(Y)$ y $\rho = \frac{Cov(X, Y)}{\sigma_x\sigma_y}$.

En general, esta distribución es útil para modelar un vector aleatorio de dos componentes que están correlacionadas. Esta distribución satisface que marginalmente tanto X como Y

siguen una distribución Normal y también se satisface que las variables aleatorias condicionales $X|(Y = y)$ y $Y|(X = x)$ se distribuyen normales. Sin embargo, nótese que puede haber un vector aleatorio con marginales que sean normales y que conjuntamente no sea normal bivariado.

2.3. La función de verosimilitud para estimar los parámetros de un modelo estadístico

Supóngase que se ha propuesto un modelo estadístico con función de probabilidad o de densidad $f(x; \theta)$ para explicar un fenómeno aleatorio de interés. El planteo de $f(x; \theta)$ debe considerar el diseño de experimento con el cual se obtuvieron los datos. El modelo $f(x; \theta)$ depende de un vector de parámetros desconocido θ (que podría ser unidimensional). Se desea usar los datos que resulten del experimento para estimar el valor de θ , y así determinar qué valores de θ hacen más probables a los datos observados. Estos valores de θ constituirán una estimación sensata de este parámetro desconocido. El estadístico y genetista inglés Sir Ronald A. Fisher (1921) fue el primero en definir el concepto de “función de verosimilitud” como una función del vector de parámetros θ que es proporcional a la probabilidad de los datos observados.

Definición (Función de verosimilitud para variables aleatorias discretas). Sea $X = (X_1, X_2, \dots, X_n)$ una muestra de n variables aleatorias discretas independientes e idénticamente distribuidas (v.a. iid) con función de distribución común $F(x; \theta)$ donde θ es un vector de parámetros desconocido de dimensión d , $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$. Dada una realización de la muestra $x = (x_1, x_2, \dots, x_n)$, la función de verosimilitud de θ , denotada por $\mathcal{L}(\theta; x)$, es una función real valuada con dominio en el espacio parametral Θ definida como

$$\mathcal{L}(\theta; x) = C(x) \prod_{i=1}^n P(X_i = x_i),$$

donde $C(x)$ es una función positiva que depende sólo de los datos y no del vector de parámetros θ y que sirve para simplificar la expresión de \mathcal{L} como función de θ .

De esta manera se observa que la función de verosimilitud $\mathcal{L}(\theta; x)$ es proporcional a la probabilidad conjunta de la muestra observada.

La función de verosimilitud la definió Fisher así para variables aleatorias discretas. Sin embargo, se puede hacer una extensión de la definición de verosimilitud para el caso de variables aleatorias continuas, ya que en realidad los datos observados x siempre son discretos puesto que todo instrumento de medición tiene precisión finita. Esto implica que sólo pueden registrarse mediciones de la variable aleatoria con un número finito de decimales. De esta manera, cuando X es una variable aleatoria continua, la observación registrada como $X = x$ debe interpretarse como $X \in [x - \frac{1}{2}\delta, x + \frac{1}{2}\delta]$, donde δ es un número fijo positivo que representa la precisión del instrumento de medición².

Definición (Función de verosimilitud para variables aleatorias continuas). Sea $X = (X_1, X_2, \dots, X_n)$ una muestra de n variables aleatorias continuas independientes e idénticamente distribuidas con función de distribución $F(x; \theta)$ donde θ es un vector de parámetros desconocido de dimensión d , $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$. Dada una realización de la muestra $x = (x_1, x_2, \dots, x_n)$, la función de verosimilitud de θ está dada por

$$\begin{aligned} \mathcal{L}(\theta; x) &= C(x) \prod_{i=1}^n P\left(x_i - \frac{1}{2}\delta < X_i \leq x_i + \frac{1}{2}\delta\right) \\ &= C(x) \prod_{i=1}^n \left[F\left(x_i + \frac{1}{2}\delta; \theta\right) - F\left(x_i - \frac{1}{2}\delta; \theta\right) \right]. \end{aligned}$$

Por el Teorema del Valor Medio para Integrales bajo ciertas condiciones que se cumplen fácilmente, la probabilidad $P\left(x_i - \frac{1}{2}\delta < X_i \leq x_i + \frac{1}{2}\delta\right)$ se puede aproximar bien por $f(x_i; \theta)$ ya que es proporcional a ella en muchas ocasiones. De lo cual se tiene una aproximación continua para la verosimilitud de variables aleatorias continuas, que se enuncia enseguida.

²Veáse Kalbfleisch (1985), p.12.

Definición (Aproximación continua para la verosimilitud de variables aleatorias continuas). Sea $X = (X_1, X_2, \dots, X_n)$ una muestra de n variables aleatorias continuas independientes e idénticamente distribuidas con función de distribución $F(x; \theta)$ donde θ es un vector de parámetros desconocido de dimensión d , $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$. Dada una realización de la muestra $x = (x_1, x_2, \dots, x_n)$, la función de verosimilitud de θ está dada por

$$\mathcal{L}(\theta; x) = C(x) \prod_{i=1}^n f(x_i; \theta).$$

De lo definido anteriormente se observa que usualmente $\mathcal{L}(\theta, x)$ es un producto de varios términos, por lo cual será conveniente trabajar con el logaritmo de esta función.

Definición (Función de log-verosimilitud). Se define la función log-verosimilitud de θ , denotada por $\ell(\theta; x)$, como el logaritmo natural de $\mathcal{L}(\theta, x)$,

$$\ell(\theta; x) = \ln \mathcal{L}(\theta; x).$$

Definición (Función Score). La función Score o de puntaje, $S_c(\theta; x)$, es el vector de primeras derivadas parciales de la función log-verosimilitud con respecto a θ ,

$$S_c(\theta; x) = \frac{\partial \ell(\theta; x)}{\partial \theta}.$$

Definición (Estimador de máxima verosimilitud). El estimador de máxima verosimilitud (EMV), denotado por $\hat{\theta}$, es el valor de $\theta \in \Theta$ que maximiza la función de verosimilitud $\mathcal{L}(\theta; x)$. Esto es,

$$\hat{\theta} = \arg \max_{\theta \in \Theta} [\mathcal{L}(\theta; x)].$$

Equivalentemente, $\hat{\theta}$ es el valor que maximiza la función log-verosimilitud

$$\hat{\theta} = \arg \max_{\theta \in \Theta} [\ell(\theta; x)], \quad (2.5)$$

esto ya que la función logaritmo es una función monótona.

Usualmente el estimador $\hat{\theta}$ se obtiene al diferenciar la función log-verosimilitud con respecto a cada elemento θ_j de θ y encontrando las raíces de las ecuaciones simultáneas,

$$\frac{\partial \ell(\theta; x)}{\partial \theta_j} = 0, \text{ para } j = 1, 2, \dots, d. \quad (2.6)$$

A (2.6) se les llama ecuaciones de máxima verosimilitud.

Definición (Información observada de Fisher). *La información observada de Fisher, denotada por $\mathbb{I}_{\hat{\theta}}$, es el negativo de la matriz de segundas derivadas parciales de $\ell(\theta; x)$ con respecto a θ , evaluada en el EMV $\hat{\theta}$,*

$$\mathbb{I}_{\hat{\theta}} = - \left[\frac{\partial^2 \ell(\theta; x)}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} \right]. \quad (2.7)$$

Es necesario verificar que las soluciones para $\hat{\theta}$ correspondan al máximo de $\ell(\theta; x)$, lo cual se hace comprobando que (2.7) sea una matriz definida positiva.

Definición (Función de verosimilitud relativa). *La función de verosimilitud relativa de θ está dada por*

$$R(\theta; x) = \frac{\mathcal{L}(\theta; x)}{\mathcal{L}(\hat{\theta}; x)}, \quad (2.8)$$

donde $\hat{\theta}$ es el EMV de θ .

Asimismo se define la función log-verosimilitud relativa como

$$r(\theta; x) = \ln R(\theta; x).$$

La función de log-verosimilitud relativa es una versión estandarizada de la función de log-verosimilitud que es fácil de interpretar. Nótese que $0 \leq r(\theta; x) \leq 1$ para cualquier valor de θ . La función (2.8) proporciona la plausibilidad de cualquier valor especificado de θ relativo al máximo verosímil $\hat{\theta}$, basada en la muestra observada x . Valores de θ cercanos a 1 son más plausibles para el verdadero valor del parámetro que aquellos que están cerca de 0 que hacen la muestra observada poco probable. Kalbfleisch (1985, pág.18) da ciertas recomendaciones al respecto, comentando que un valor de θ que tenga $r(\theta) \leq 0.10$ resulta ser no creíble para la muestra observable.

2.3.1. Propiedad de invarianza frente a reparametrizaciones uno a uno

Un requerimiento básico que se debe pedir a todo procedimiento estadístico es que las conclusiones que se obtengan a partir de una muestra aleatoria x proveniente de algún modelo estadístico $f(x; \theta)$, sean independientes de la parametrización que se haya elegido; asimismo no deben depender de la escala en que se hayan medido las observaciones.

La verosimilitud tiene la propiedad de ser **invariante frente a reparametrizaciones uno a uno**. Sea $\phi = g(\theta)$, donde g es una función continua y biyectiva. La verosimilitud \mathcal{L}^* del nuevo parámetro ϕ se obtiene mediante una simple sustitución algebraica de la verosimilitud $\mathcal{L}(\theta; x)$ del parámetro original θ , esto es:

$$\mathcal{L}^*(\phi; x) = \mathcal{L}[g^{-1}(\phi); x]. \quad (2.9)$$

La verosimilitud L^* se puede obtener como el producto de $f(x_i; \phi^{-1})$ para $i = 1, 2, \dots, n$ o sustituyendo algebraicamente ϕ^{-1} por θ directamente en la verosimilitud de θ .

Debido a la relación (2.9), se tiene que los EMV se relacionan a su vez de la siguiente manera:

$$\begin{aligned} \hat{\phi} &= g(\hat{\theta}) \\ \hat{\theta} &= g^{-1}(\hat{\phi}). \end{aligned}$$

También se cumple la siguiente relación para las informaciones observadas de Fisher,

$$\mathbb{I}_{\hat{\phi}} = \mathbb{I}_{\hat{\theta}} \left(\frac{d\theta}{d\phi} \Big|_{\phi=\hat{\phi}} \right)^2.$$

En general se reparametriza un modelo estadístico en términos de un parámetro particular por dos razones. La primera es cuando el nuevo parámetro tiene una interpretación más clara que sea de interés o que analíticamente simplifique la expresión algebraica del modelo estadístico. La segunda razón es para simetrizar la función de verosimilitud alrededor

del EMV; esto es, se busca que \mathcal{L}^* sea mucho más simétrica que \mathcal{L} . Entre más simétrica sea una función de verosimilitud de θ , será más fácil optimizar numéricamente para estimar los parámetros desconocidos. Además que será razonable aplicar resultados asintóticos a la muestra finita particular que se observó.

Ejemplo. *Considérense cinco observaciones simuladas de una variable aleatoria que sigue una distribución exponencial de parámetro $\theta = 1$.*

En la Figura (2.1) se muestra la gráfica de verosimilitud relativa para el modelo exponencial de parámetro $\theta = 1$ mientras que en la Figura (2.2) se muestra la gráfica de verosimilitud relativa de la reparametrización $\phi = \theta^{-\frac{1}{3}}$.

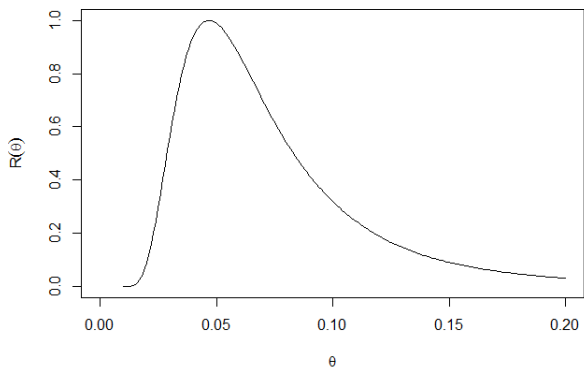


Figura 2.1: Verosimilitud relativa de θ .

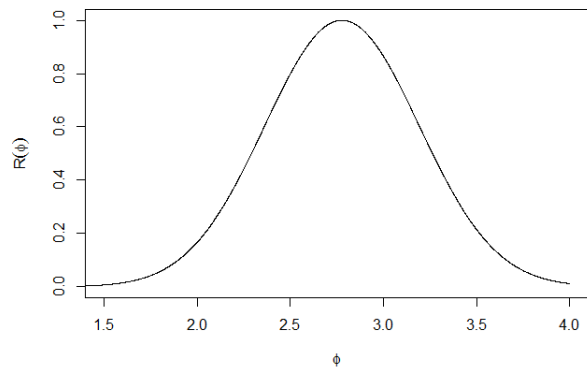


Figura 2.2: Verosimilitud relativa de la reparametrización $\phi = \theta^{-\frac{1}{3}}$.

Puede observarse que con esta reparametrización la función de verosimilitud relativa de ϕ es completamente simétrica.

Otras propiedades de los estimadores de máxima verosimilitud es que son consistentes, suficientes y normales asintóticamente³.

³Veáse Serfling (1980), p. 142 y Pawitan (2001), Teorema 9.7

2.3.2. Intervalos de verosimilitud-confianza

Una forma común de hacer inferencia sobre un parámetro es a través de intervalos o regiones de estimación que indiquen la probabilidad de que dicha región incluya al verdadero valor del parámetro del modelo probabilístico que dio origen a los datos observados. Además, estos intervalos o regiones de verosimilitud indican los valores más plausibles del parámetro dada la información contenida en los datos observados.

Definición (Intervalo de verosimilitud). *Un intervalo o región de verosimilitud de nivel c para θ se define como*

$$IV(c) = \{\theta | R(\theta; x) \geq c\}, \text{ donde } c \in [0, 1]. \quad (2.10)$$

El intervalo de verosimilitud separa los valores más plausibles de θ de los no plausibles a un nivel c de verosimilitud⁴. Nótese que el EMV de θ está contenido en todos los intervalos de verosimilitud, esto puesto que en el EMV, $R(\hat{\theta}; x) = 1$ y para cualquier intervalo de nivel $c \in [0, 1]$ se tiene por tanto que

$$0 \leq c \leq R(\hat{\theta}; x).$$

Los extremos de este intervalo de verosimilitud $\theta_1(x)$ y $\theta_2(x)$ son aleatorios ya que dependen de las observaciones y por tanto cambian para muestras distintas. Por ello este intervalo puede o no incluir al verdadero valor θ_0 , entonces resulta necesario saber con qué probabilidad un intervalo para θ contiene al verdadero valor θ_0 , bajo repeticiones hipotéticas similares a la muestra observada.

Definición (Probabilidad de cobertura). *La probabilidad de cobertura de un intervalo aleatorio, $[\theta_1(x), \theta_2(x)]$, para θ es la probabilidad de que el intervalo incluya al verdadero valor θ_0 , que dio origen a la muestra observada $x = (x_1, x_2, \dots, x_n)$ bajo el modelo $f(x; \theta_0)$,*

$$PC(\theta_0) = P[\theta_1(x) \leq \theta_0 \leq \theta_2(x); \theta = \theta_0]. \quad (2.11)$$

⁴Veáse Sprott (2000), p.14.

La probabilidad de cobertura (2.11) se interpreta como la proporción de veces que en muestras repetidas bajo condiciones similares a la observada, el intervalo $[\theta_1(x), \theta_2(x)]$ incluye al valor verdadero θ_0 cuando el número de repeticiones del experimento con θ fijo en θ_0 , es suficientemente grande.

Definición (Intervalo de confianza). *Un intervalo aleatorio $[\theta_1(x), \theta_2(x)]$ se dice que es intervalo de confianza para θ si su probabilidad de cobertura no depende de θ_0 . Es decir, que para todo valor del parámetro θ_0 , el valor de la probabilidad de cobertura del intervalo es la misma.*⁵

Para un tamaño de muestra fijo, la distribución de probabilidad de θ_1 y θ_2 se puede calcular a partir de la distribución de X . Sin embargo, existen otras formas de asociar un nivel de confianza aproximado a un intervalo de estimación para θ . Particularmente, la probabilidad de cobertura de un $IV(c)$ se puede calcular a través de la distribución asintótica de la estadística de razón de verosimilitudes para θ fijo en θ_0 , la cual se define como

$$D(\theta; x) \equiv -2 \log R(\theta_0; x).$$

Se sabe que $D(\theta; x)$ converge en distribución a una Ji-cuadrada con d grados de libertad, donde $d = \dim(\theta)$ ⁶. Por lo tanto, para n suficientemente grande,

$$\begin{aligned} PC(\theta_0) &= P[IV(c) \supset \theta | \theta = \theta_0] \\ &= P[R(\theta; x) \geq c | \theta = \theta_0] \\ &= P[D(\theta; x) \leq -2 \log(c) | \theta = \theta_0] \\ &\approx P[D(\theta; x) \leq Q_\pi | \theta = \theta_0] = \pi, \end{aligned}$$

donde Q_π es el cuantil de probabilidad de una distribución Ji-cuadrada con d grados de libertad.

⁵Veáse en Kalbfleisch, Cap. 11, un ejemplo de un intervalo de verosimilitud que no es de confianza.

⁶Pawitan (2001), p. 243.

La distribución asintótica de $D(\theta; x)$ y la relación $Q_\pi = -2 \log(c)$, o $c = \exp(-Q_\pi/2)$ permiten dos cosas:

- a) si se desea asignar una probabilidad a un intervalo de verosimilitud de nivel c , se calcula el cuantil Q_π (a partir de la primera expresión) y luego se averigua cuál probabilidad corresponde,
- b) si se desea encontrar el nivel de verosimilitud c para un intervalo que tenga probabilidad π , entonces se calcula c usando la segunda expresión.

En la Tabla (3.8) se muestran los valores más comunes de c y π asociados.

Nivel de verosimilitud c	Nivel de confianza π	Q_π
0.7965	0.50	0.45
0.2585	0.90	2.71
0.1465	0.96	3.86
0.0362	0.99	6.63

Tabla 2.1: Niveles de confianza aproximados de los intervalos de verosimilitud cuando θ es unidimensional.

Ejemplo. Inferencia para el parámetro θ de una distribución Binomial

Considérese una muestra de v.a. iid, $X = (X_1, X_2, \dots, X_n)$ con distribución binomial de parámetros n y θ donde n es conocido. La función de probabilidad para cada variable aleatoria está dada por

$$f(x; \theta) = P(X = x; \theta) = \binom{k}{x} \theta^x (1 - \theta)^{k-x}.$$

La muestra observada es $x = (x_1, x_2, \dots, x_n)$. La función de verosimilitud correspondiente a la muestra x es

$$\mathcal{L}(\theta; x) = C(x) \theta^t (1 - \theta)^{nk-t},$$

donde $t = \sum_{i=1}^k x_i$ y $C(x) = \left[\prod_{i=1}^k \binom{k}{x_i} \right]^{-1}$.

La log-verosimilitud resulta ser

$$\ell(\theta; x) = t \log(\theta) + (nk - t) \log(1 - \theta). \quad (2.12)$$

En el caso en que $0 < t < nk$, la función Score asociada está dada por:

$$S_c(\theta; x) = \frac{t}{\theta} - \frac{nk - t}{1 - \theta}.$$

Para obtener el EMV de θ iguala a cero la función Score obtenida y se despeja θ , de donde

$$\hat{\theta} = \frac{t}{nk}.$$

La información observada de Fisher cuando $0 < t < nk$ es:

$$\mathbb{I}_{\hat{\theta}} = -\frac{t}{\hat{\theta}^2} + \frac{nk - t}{(1 - \hat{\theta})^2}.$$

Nótese que la log-verosimilitud (2.12) no es continua si $t = 0$ o si $t = nk$. En estos casos el EMV es $\hat{\theta} = 0$ y $\hat{\theta} = 1$, respectivamente.

En el caso en que $t = 0$, el intervalo de verosimilitud de nivel c que se considera es de la forma $[0, \theta_1]$, como se muestra un ejemplo en la Figura 2.4.

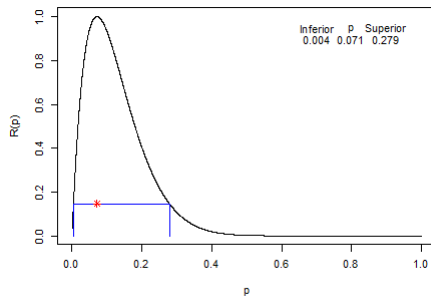


Figura 2.3: Verosimilitud relativa cuando $0 < t < nk$.

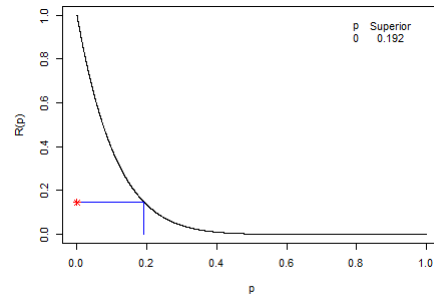


Figura 2.4: Verosimilitud relativa cuando $t = 0$.

Ejemplo. Inferencia para los parámetros μ y σ^2 de una distribución Normal

Considérese una muestra observada de v.a. iid, $x = (x_1, x_2, \dots, x_n)$ proveniente de una distribución Normal de parámetros μ y σ^2 . La función de densidad para cada variable aleatoria está dada por

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2}(x - \mu)^2 \right].$$

La muestra observada es $x = (x_1, x_2, \dots, x_n)$. La función de verosimilitud correspondiente a la muestra x es

$$\mathcal{L}(\mu, \sigma^2; x) = C(x)f(x; \mu, \sigma^2) = C(x)\frac{1}{\sigma^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right],$$

donde $C(x) = (\sqrt{2\pi})^{-n}$.

La log-verosimilitud resulta ser

$$\ell(\mu, \sigma^2; x) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

La función Score asociada está dada por:

$$S_c(\mu, \sigma^2) = \begin{bmatrix} \frac{t_1 - n\mu}{\sigma^2} \\ -\frac{n}{2\sigma^2} + \frac{t_2 - 2\mu t_1 + n\mu^2}{2\sigma^4} \end{bmatrix}.$$

donde $t_1 = \sum_{i=1}^n x_i$ y $t_2 = \sum_{i=1}^n x_i^2$. Luego, los EMV de μ y σ^2 son

$$\begin{aligned} \hat{\mu} &= \bar{x}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned}$$

La matriz de información de Fisher, en este caso, es:

$$\mathbb{I}_{\hat{\mu}, \hat{\sigma}^2} = \begin{bmatrix} -n/\hat{\sigma}^2 & 0 \\ 0 & -n/[2(\hat{\sigma}^2)^2] \end{bmatrix}.$$

2.3.3. Función de verosimilitud perfil

Con frecuencia se tienen modelos estadísticos de los cuales sólo interesa estimar uno de sus parámetros desconocidos. En Estadística es relevante el problema de estimar por separado parámetros de interés en presencia del resto de los parámetros, éstos últimos conocidos como parámetros de estorbo, ya que pueden tener un impacto significativo en las inferencias del parámetro de interés. Un método muy útil para realizar esta estimación se basa en la función de verosimilitud perfil o maximizada⁷.

Definición (Función de verosimilitud perfil o maximizada). Sea $\theta = (\theta_1, \theta_2)$ donde $\dim(\theta) = 1$; dada una realización $x = (x_1, x_2, \dots, x_n)$ de una muestra de tamaño n de v.a. iid, la función de verosimilitud perfil o maximizada del parámetro de interés θ_1 , es una función real valuada no negativa, denotada por $\mathcal{L}_p(\theta_1; x)$, está dada por

$$\mathcal{L}_p(\theta_1; x) = \max_{\theta_2 | \theta_1} \mathcal{L}(\theta_1, \theta_2; x) = \mathcal{L}(\theta_1, \hat{\theta}_2(\theta_1; x); x),$$

donde $\hat{\theta}_2(\theta_1; x)$ es el estimador de máxima verosimilitud restringido (EMVR) de θ_2 para valores fijos de θ_1 .

El dominio de $\mathcal{L}(\theta_1)$ es $\Theta_1 \subset \mathbb{R}$ y su contradominio es $\mathbb{R}^+ \cup \{0\}$.

Ejemplo. Verosimilitud perfil de N en un modelo Binomial de parámetros (N, θ) desconocidos.

Considérese una realización $x = (x_1, x_2, \dots, x_n)$ de una muestra de n v.a. iid $X = (X_1, X_2, \dots, X_n)$ con distribución común Binomial de parámetros N y θ desconocidos. Interesa hacer inferencia sobre el tamaño de población N , por lo cual, el parámetro θ se considera como de estorbo. La probabilidad conjunta de esta muestra x es

$$P(X = x; N, \theta) = \theta^t (1 - \theta)^{Nn-t} \prod_{i=1}^n \binom{N}{x_i},$$

donde $t = \sum_{i=1}^n x_i$.

⁷Ver Pawitan (2001) y Kalbfleisch(1985) Sección 10.3.

La función de verosimilitud global de N y θ es proporcional a la probabilidad de observar la muestra y está dada por

$$\mathcal{L}(N, \theta; x) = \theta^t (1 - \theta)^{Nn-t} \prod_{i=1}^n \frac{N!}{(N - x_i)!}.$$

La función de verosimilitud perfil de N se obtiene maximizando sobre θ la función de verosimilitud dada anteriormente, para una rejilla de valores fijos de N . Para un valor fijo de N la logverosimilitud asociada es

$$\ell(N, \theta; x) = t \log(\theta) + (Nn - t) \log(1 - \theta) + n \log(N!) - \sum_{i=1}^n \log(N - x_i!).$$

Al derivar esta última expresión con respecto a θ , se obtiene

$$\frac{\partial \ell(N, \theta; x)}{\partial \theta} = \frac{t}{\theta} - \frac{Nn - t}{1 - \theta}.$$

Igualando a cero y despejando θ , se obtiene el EMV de θ restringido a N ,

$$\hat{\theta}(N) = \frac{t}{Nn}.$$

Así, la función de verosimilitud perfil de N , $\mathcal{L}_p(N; x)$, se obtiene al reemplazar el EMVR de θ , es decir,

$$\begin{aligned} \mathcal{L}_p(N; x) &= \mathcal{L}[N, \hat{\theta}(N); x] \\ &= \left(\frac{t}{Nn}\right)^t \left(1 - \frac{t}{Nn}\right)^{Nn-t} \prod_{i=1}^n \binom{N}{x_i}. \end{aligned}$$

En este caso el EMV de N no tiene expresión cerrada y se obtiene maximizando numéricamente sobre N la expresión previa.

2.4. Validación de modelos

Las gráficas cuantil-cuantil y la prueba de mardia para la normal bivariada se usarán aquí como un método de validación de los modelos estadísticos considerados razonables para los datos observados.

2.4.1. La importancia de los cuantiles y de la función de distribución empírica para validar modelos

Se definirá primero la función de cuantiles a través de la cual se obtiene un cuantil de la distribución con probabilidad α .

Definición (Función de cuantiles). *La función cuantil de una distribución de probabilidad se define como la inversa generalizada de la función de distribución correspondiente*

$$Q(\alpha) = F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\},$$

para una probabilidad $\alpha \in [0, 1]$.

Nótese que esta función cuantil devuelve siempre un único valor de x para cada probabilidad llamado el cuantil de probabilidad α , éste valor es el mínimo de x tal que $F(x) \geq \alpha$. En contraste, esta función puede asignar el mismo valor de x a más de una probabilidad α , por lo que no siempre será una función uno a uno. Esto ocurre en el caso de variables aleatorias discretas que tienen una distribución escalonada. En el caso de distribuciones absolutamente continuas, nótese que $Q(\alpha) = F^{-1}(\alpha)$; es decir, coincide con la distribución inversa habitual de la función de distribución correspondiente y para cada probabilidad α hay un cuantil único $Q(\alpha)$.

Una gráfica cuantil-cuantil, denominada aquí también como gráfica Q-Q, es un método gráfico que se usa para comparar dos distribuciones a través de comparar sus cuantiles para un conjunto de probabilidades $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$. Dos distribuciones son iguales si todos sus cuantiles coinciden.

En particular interesa comparar la distribución empírica $F_n(x)$ que se define enseguida, con la distribución estimada propuesta $F(x; \hat{\theta})$ de un modelo considerado razonable para los datos.

Definición (Función de distribución empírica). Sea $x = (x_1, x_2, \dots, x_n)$ una muestra observada de variables aleatorias. Se define la función de distribución empírica como una función escalonada dada por

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(-\infty, x]}(x_i), \quad (2.13)$$

donde \mathbb{I} es la función indicadora.

El Teorema de Glivenko-Cantelli es uno de los resultados fundamentales de la estadística, en dicho teorema se demuestra la convergencia uniforme de la función de distribución empírica de una muestra $F_n(x)$ a la verdadera función de distribución de la variable aleatoria $F(x)$.

Definición (Teorema de Glivenko-Cantelli). Sean X_1, X_2, \dots, X_n variables aleatorias iid, entonces $F_n(x)$ converge uniformemente a $F(x)$, ya que para todo $\epsilon > 0$, se tiene que

$$\lim_{n \rightarrow \infty} P \left(\sup_{-\infty < x < \infty} |F_n(x) - F(x)| > \epsilon \right) = 0.$$

La función de distribución empírica se convierte así en un estimador inmediato y natural para la función de distribución de la variable aleatoria ya que $F_n(x)$ converge uniformemente a la verdadera distribución que generó a las observaciones.

Por la Ley de Grandes Números, dado que $F_n(x)$ se puede ver como el promedio de variables aleatorias Bernoulli por la definición (2.13), se tiene que $F_n(x)$ converge con probabilidad uno a $F(x; \theta_0)$, donde $F(x; \theta_0)$ es el modelo verdadero. Además, por el Teorema de Glivenko-Cantelli

$$\lim_{n \rightarrow \infty} P \left(\sup_{-\infty < x < \infty} |F_n(x) - F(x; \theta_0)| \right) = 0.$$

Es decir, para evaluar la bondad de ajuste de un modelo estadístico propuesto se compararán los cuantiles de la distribución estimada con los datos $F(x; \hat{\theta})$ contra los cuantiles empíricos provenientes de la función de distribución empírica $F_n(x)$, que son precisamente las observaciones ordenadas.

Para una muestra observada de tamaño n , se pueden considerar las n probabilidades siguientes $\alpha_i = \frac{i}{n+1}$, para $i = 1, 2, \dots, n$; éstas son

$$\left\{ \frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1} \right\}.$$

En una gráfica cuantil-cuantil se graficarán los puntos con coordenadas

$$(Q(\alpha_i; \hat{\theta}), x_{(i)}), \text{ para } i = 1, 2, \dots, n.$$

Marron et al. (2004) sugieren agregar una nube de puntos correspondientes a M muestras similares simuladas de tamaño n bajo el mismo modelo estimado, para considerar el ajuste del modelo. En general, usar $M = 100$ funciona bien. Para cada muestra simulada se grafica la gráfica Q-Q asociada. Estos puntos forman una nube con la que se contrastan los puntos de la muestra observada.

La Figura (2.5) muestra un ejemplo de una gráfica Q-Q para la suma de áreas basales de tallos de árboles de la especie *Plumeria rubra* antes del paso del huracán Jova donde se comparan los cuantiles empíricos de $F_n(x)$ que son las observaciones ordenadas contra los cuantiles de un modelo normal estimado. Se encimaron los cuantiles similares de 100 muestras simuladas en color gris bajo el mismo modelo normal estimado.

2.4.2. La prueba de Mardia para normalidad multivariada

Cuando se desea evaluar si un conjunto de datos multivariados siguen una distribución normal multivariada usualmente se verifica primero que marginalmente cada componente del vector aleatorio sea normal univariada a través de gráficas cuantil-cuantil como las descritas en la sección previa para el caso particular de la distribución normal multivariada. Con este fin, Mardia (1974) propuso una prueba para verificar el supuesto de normalidad multivariada con base en extensiones multivariadas de las medidas de sesgo ($\hat{\gamma}_{1,p}$) y curtosis ($\hat{\gamma}_{2,p}$) definidas como sigue:

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n m_{ij}^3 \text{ y } \hat{\gamma}_{2,p} = \frac{1}{n} m_{ii}^2$$

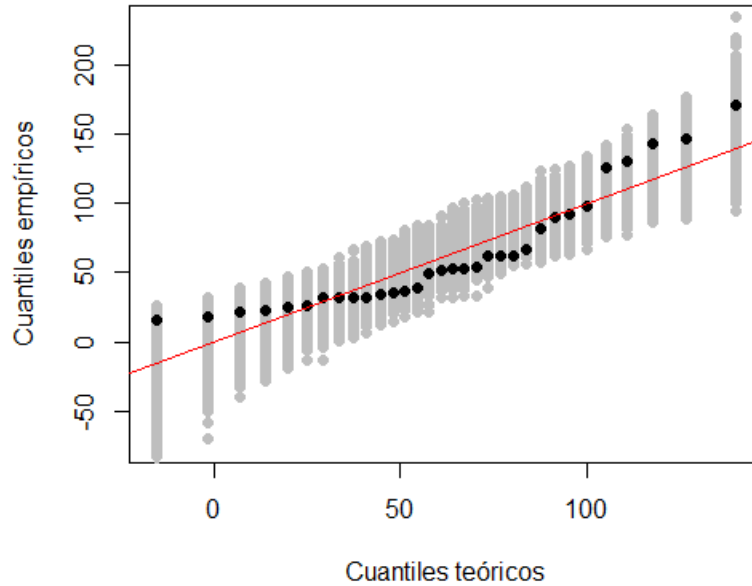


Figura 2.5: Gráfica Q-Q de sumas de áreas basales de tallos por árbol de la especie *Plumeria rubra* antes del paso del huracán JOVA.

donde $m_{ij} = (x_i - \bar{x})S^{-1}(x_j - \bar{x})$ es el cuadrado de la distancia de Mahalanobis y p es el número de variables.

En el caso del estadístico de prueba para el sesgo, éste se distribuye como una Ji cuadrada con q grados de libertad, donde $q = \frac{p(p+1)(p+2)}{6}$. Esto es,

$$\frac{n}{6} \hat{\gamma}_{1,p} \sim \chi^2(q),$$

Similarmente, el estadístico de prueba de la curtoris sigue una distribución normal con media $p(p+2)$ y varianza $\frac{8p(p+2)}{n}$.

Cuando el tamaño de muestra es menor a 20, Mardia (1974) recomendó usar una corrección en el estadístico de prueba del sesgo, para controlar el error tipo I en la prueba de hipótesis asociada. El estadístico de prueba corregido para el sesgo es $\frac{nk}{6} \hat{\gamma}_{1,p}$, donde

$k = (p + 1)(n + 1)(n + 3)/(n(n + 1)(p + 1) - 6)$. Este estadístico también se distribuye como χ^2 con $p(p + 1)(p + 2)/6$ grados de libertad.

2.5. Diseño de muestras pareadas

Cuando se tiene un estudio comparativo de dos poblaciones lo ideal sería que estas poblaciones sean similares en todos los aspectos excepto en la variable de interés que se está estudiando.

Considérese el siguiente ejemplo hipotético. Supóngase que se realiza un estudio de los sueldos de hombres y mujeres y se considera una muestra aleatoria de cuatro matrimonios

Esposo	Esposa
Roman Cruz	Celia Palma de Cruz
Lucio Mireles	Rosa Márquez de Mireles
Reynaldo Perales	Rita Torres de Perales
Marcos García	Lorena Lomelí de García

Tabla 2.2: Diseño pareado

Esta muestra se dice pareada y el tamaño muestral corresponde al número de pares que son cuatro, no al número de individuos que serían ocho. Cuando los datos son pareados interesa comparar las respuestas de cada par. Por lo cual, se analizan las diferencias entre las respuestas que se observaron en cada par.

Salario esposa	Salario del esposo	Diferencia
15	20	-5
45	31	14
50	50	0
16	30	-14
	Promedio de diferencias	-1.25

Tabla 2.3: Sueldos mensuales en miles de pesos de cuatro matrimonios.

La ventaja principal de los diseños pareados es que ayudan a reducir el sesgo por variables

confundentes. Se pueden parear sujetos por edad, sexo, estado de salud, etc. En general, las variables que se eligen para parear son variables que pueden influir en la respuesta. Cuando se comparan los resultados de observaciones pareadas, los efectos de estas variables de pareamiento se cancelan⁸.

Otro tipo de diseño pareado es cuando cada sujeto sirve como su propio control. Por ejemplo, si se tienen dos medicamentos, cada sujeto puede ser tratado con ambos medicamentos en orden aleatorio. Si se tiene sólo un medicamento, se miden las respuestas comparando la situación antes y después del suministro del medicamento. Un ejemplo son mediciones de colesterol de lipoproteínas de baja densidad, tomadas antes y después del tratamiento con un medicamento en los mismos n pacientes.

Si se tienen dos muestras aleatorias dependientes de tamaño n , $(x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n)$ donde cada elemento de la primera muestra es pareja de un elemento de la segunda, entonces estas dos muestras dan lugar a una muestra pareada o a una muestra de diferencias. La muestra de diferencias $\delta_i = y_i - x_i$ se puede pensar como una muestra de la población de diferencias de datos pareados de dos poblaciones. La media de la población de diferencias es igual a la diferencia de las medias poblacionales.

Es crucial incorporar el diseño del experimento en el análisis estadístico que se haga, ya que será fundamental para el éxito en un estudio comparativo. El diseño experimental con muestras independientes requiere que los individuos sean seleccionados aleatoriamente para evaluar el efecto de interés en cada grupo. Con el diseño de muestras pareadas, los individuos en cada par deber ser muy similares y compartir características comunes, mientras que en cada par diferente conviene que sean distintos. De esta forma, la idea de formar parejas de datos es eliminar posibles fuentes de variación desconocidas que estén influyendo en la comparación de interés de las medias μ_X y μ_Y .

⁸Veáse Battacharyya y Johnson (2010)

2.6. Métodos estadísticos tradicionales de análisis para datos pareados

2.6.1. La distribución Normal para las diferencias

En Triola (2000) y Bhattacharyya y Johnson (2010) se describe claramente el método para comparar las medias de dos poblaciones normales usando muestras pareadas.

Considérese una muestra pareada de tamaño n , $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Se tendrán entonces n diferencias $\delta_i = y_i - x_i$, $i = 1, 2, \dots, n$. Se suele buscar que las $\{\delta_i\}_{i=1}^n$ sigan una distribución aproximadamente normal. Sea $\delta = \mu_Y - \mu_X$ la diferencia de medias poblacionales. Nótese que $\hat{\delta} = \bar{y} - \bar{x}$.

En Hogg y Craig (1978) se demuestra para una muestra observada $x = (x_1, x_2, \dots, x_n)$ con media μ y varianza σ^2 , que la cantidad pivotal

$$T = \frac{\sqrt{n-1}(\hat{\mu} - \mu)}{\hat{\sigma}}$$

sigue una distribución t de Student con $(n-1)$ grados de libertad, donde $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$ y $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

Nótese así que, dado que la distribución t-Student es simétrica y está centrada alrededor

del cero, se puede calcular la siguiente probabilidad donde t es el cuantil de probabilidad $\pi/2$,

$$\begin{aligned}
 \pi &= P(-t \leq T \leq t) \\
 &= P(-t \leq \frac{\sqrt{n-1}(\hat{\mu} - \mu)}{\hat{\sigma}} \leq t) \\
 &= P(-t\hat{\sigma} \leq \sqrt{n-1}(\hat{\mu} - \mu) \leq t\hat{\sigma}) \\
 &= P\left(\frac{-t\hat{\sigma}}{\sqrt{n-1}} \leq \hat{\mu} - \mu \leq \frac{t\hat{\sigma}}{\sqrt{n-1}}\right) \\
 &= P\left(\frac{-t\hat{\sigma}}{\sqrt{n-1}} - \hat{\mu} \leq -\mu \leq \frac{t\hat{\sigma}}{\sqrt{n-1}} - \hat{\mu}\right) \\
 &= P\left(\hat{\mu} - \frac{t\hat{\sigma}}{\sqrt{n-1}} \leq \mu \leq \hat{\mu} + \frac{t\hat{\sigma}}{\sqrt{n-1}}\right).
 \end{aligned}$$

Es decir, se ha encontrado así la probabilidad asociada al intervalo de verosimilitud perfil de μ de nivel de verosimilitud $c = \left[\frac{t^2}{n-1} + 1\right]^{-\frac{n}{2}}$ para el cuantil de probabilidad $t_{\alpha/2} = \sqrt{n-1}\sqrt{c^{-2/n} - 1}$.

De esta forma, se encuentra un intervalo de confianza $\alpha \times 100\%$ y de verosimilitud $c = \left[\frac{t^2}{n-1} + 1\right]^{-\frac{n}{2}}$ para la diferencia de medias δ dado por

$$\left(\hat{\delta} - t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n-1}}, \hat{\delta} + t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n-1}}\right),$$

donde $t_{\frac{\alpha}{2}}$ es el cuantil de probabilidad $\alpha/2$ de una distribución t de Student con $(n-1)$ grados de libertad.

En esta tesis se estimará δ y calcularán intervalos de confianza para este parámetro para las 17 especies consideradas.

2.6.2. La distribución Normal Bivariada

Considérese una muestra de v.a. bivariadas iid, $X = [(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$ con distribución Normal Bivariada de parámetros vector de medias,

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix},$$

y matriz de covarianza,

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_x \sigma_y \rho \\ \sigma_x \sigma_y \rho & \sigma_Y^2 \end{pmatrix}.$$

La función de probabilidad para cada variable aleatoria bivariada está dada como en 2.4.

La muestra observada es $(x, y) = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$. La función de verosimilitud correspondiente a la muestra x es:

$$\begin{aligned} \mathcal{L}[\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho; (x, y)] &= -\frac{1}{[2\pi\sigma_x\sigma_y(1-\rho^2)]^n} \exp\left[-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x}\right)^2\right] \\ &\times \exp\left[-\frac{\rho}{1-\rho^2} \sum_{i=1}^n \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x\sigma_y}\right] \\ &\times \exp\left[-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{y_i - \mu_y}{\sigma_y}\right)^2\right]. \end{aligned}$$

La log-verosimilitud asociada a la muestra (x, y) está dada por,

$$\begin{aligned} \mathcal{L}[\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho; (x, y)] &= -\frac{n}{2} \ln[(2\pi)^2] - \frac{n}{2} \ln(\sigma_x^2) - \frac{n}{2} \ln(\sigma_y^2) - \frac{n}{2} \ln(1-\rho^2) \\ &- \frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x}\right)^2 \\ &- \frac{\rho}{1-\rho^2} \sum_{i=1}^n \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x\sigma_y} \\ &- \frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{y_i - \mu_y}{\sigma_y}\right)^2. \end{aligned}$$

En el Apéndice A se muestran los cálculos hechos para encontrar los EMVs de los parámetros para el caso multivariado. Aquí se muestran los EMV en el caso bivariado de los cinco

parámetros desconocidos,

$$\begin{aligned}\hat{\mu}_x &= \bar{x}, \\ \hat{\mu}_y &= \bar{y}, \\ \hat{\sigma}_x^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \\ \hat{\sigma}_y^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2, \\ \hat{\rho} &= \frac{1}{\hat{\sigma}_x \hat{\sigma}_y} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).\end{aligned}$$

Se desea estimar la diferencia de medias de dos poblaciones normales, así como la razón de medias a través de intervalos de verosimilitud perfil los cuales tengan asociados niveles de confianza altos. Por ello, será conveniente reparametrizar la función de verosimilitud asociada al conjunto de variables aleatorias bivariadas.

Considérese la siguiente reparamerización $\delta = \mu_Y - \mu_X$ y $\tau = \frac{\sigma_Y}{\sigma_X}$. Así, la función de verosimilitud resultante será:

$$\begin{aligned}\mathcal{L}[\mu_x, \delta, \sigma_x^2, \tau, \rho; (x, y)] &= \frac{1}{[2\pi\sigma_x^2\tau(1-\rho^2)]^n} \exp \left[-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x} \right)^2 \right] \\ &\times \exp \left[-\frac{\rho}{1-\rho^2} \sum_{i=1}^n \frac{(x_i - \mu_x)(y_i - \delta - \mu_x)}{\sigma_x^2\tau} \right] \\ &\times \exp \left[-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{y_i - \delta - \mu_x}{\sigma_x\tau} \right)^2 \right].\end{aligned}$$

Por la propiedad de invarianza se tiene que:

$$\begin{aligned}\hat{\delta} &= \hat{\mu}_Y - \hat{\mu}_X \\ \hat{\tau} &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\end{aligned}$$

Si se hace $\beta = \frac{\mu_Y}{\mu_X}$ y $\tau = \frac{\sigma_Y}{\sigma_X}$, se tiene que la función de verosimilitud asociada es:

$$\begin{aligned} \mathcal{L}[\mu_x, \beta, \sigma_x^2, \tau, \rho; (x, y)] &= \frac{1}{[2\pi\sigma_x^2\tau(1-\rho^2)]^n} \exp\left[-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x}\right)^2\right] \\ &\times \exp\left[-\frac{\rho}{1-\rho^2} \sum_{i=1}^n \frac{(x_i - \mu_x)(y_i - \beta\mu_x)}{\sigma_x^2\tau}\right] \\ &\times \exp\left[-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left(\frac{y_i - \beta\mu_x}{\sigma_x^2\tau}\right)^2\right]. \end{aligned}$$

Por la propiedad de invarianza se tiene que:

$$\begin{aligned} \hat{\beta} &= \frac{\hat{\mu}_Y}{\hat{\mu}_X}, \\ \hat{\tau} &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}. \end{aligned}$$

En todos los casos, los intervalos de verosimilitud se estimarán numéricamente a través de la función de verosimilitud perfil de cada parámetro.

Capítulo 3

Análisis estadístico

Se detallará aquí la información relevante sobre los registros de la base de datos y se describirán las variables aleatorias de interés que se modelaran estadísticamente. Se presentan estadísticas descriptivas para las variables de interés así como la parte de estimación puntual y por intervalo tanto de la proporción de daño como del crecimiento de área basal en cada especie. Es decir, se presenta una modelación completa del fenómeno de interés.

3.1. Descripción detallada de los datos

En la estación de Biología de Chamela, desde 1982, el biólogo y M. C. Alfredo Pérez y su equipo de trabajo registraron variables de interés de 7 parcelas distribuidas en 3 altitudes diferentes de 5 cuencas. Aproximadamente cada parcela es de 30 por 80 metros. Se eligió un cuadrante de altitud media para el análisis que se realizará aquí.

En el cuadrante elegido se midieron tallos de árboles de 87 especies en total. Sin embargo, se realizó una depuración de esta base de datos, quedando aquellas especies que tenían 9 o más árboles para modelar la distribución estadística de la población asociada adecuadamente. Las especies que cumplieron éste requisito fueron las siguientes 17: *Colubrina heteroneura*,

Cordia alliodora, *Adelia oaxacana*, *Exostema caribaeum*, *Piptadenia constricta*, *Bauhinia unguolata*, *Caesalpinia eriostachys*, *Comocladia engleriana*, *Plumeria rubra*, *Guapira macrocarpa*, *Guettarda elliptica*, *Bursera instabilis*, *Helietta lottiae*, *Tabebuia impetiginosa*, *Lonchocarpus eriocarinalis*, *Samyda mexicana* y *Tabebuia chrysantha*.

Se identificaron dos grupos de árboles: los dañados y no dañados. Un árbol dañado se define como aquel que perdió al menos uno de sus tallos en la medición hecha en mayo de 2012 con respecto a la hecha en mayo de 2011. Con esto se podrá estimar la proporción de árboles dañados por especie. Además, esto permitirá estimar el crecimiento de área basal en los árboles no dañados que conservaron todos sus tallos.

Las variables aleatorias de interés que se considerará aquí fueron medidas antes y después del paso del huracán Jova, en mayo de 2011 y mayo de 2012. Éstas son:

$$\begin{aligned} X_{ij} &= \text{suma de áreas basales de tallos del árbol } i \text{ de la especie } j \text{ en 2011,} \\ X_{ij}^* &= \text{sumas de áreas basales de tallos del árbol } i \text{ de la especie } j \text{ en 2012,} \\ Y_{ij} &= \text{número de tallos del árbol } i \text{ de la especie } j \text{ en 2011,} \\ Y_{ij}^* &= \text{número de tallos del árbol } i \text{ de la especie } j \text{ en 2012,} \\ Z_j &= \text{número de árboles dañados de la especie } j \text{ en 2012,} \end{aligned}$$

donde $j = 1, 2, \dots, 17$ e $i = 1, 2, \dots, n_j$ y n_j es el número de árboles de la especie j . En este caso $n_j \geq 9$, para las 17 especies consideradas.

Aquí se considera que las variables que están más asociadas al efecto del huracán Jova y eventos ocurridos entre el periodo mayo 2011 y mayo 2012 son Y_{ij} , Y_{ij}^* y Z_j . Así, se podrá evaluar la proporción de árboles dañados por especie. También, las sumas de áreas basales X_{ij} y X_{ij}^* permitirán estimar el crecimiento en área basal de los árboles que posteriormente los ecólogos podrán aprovechar para estimar la biomasa utilizando ecuaciones alométricas.

El orden en el cual se presentan las tablas y figuras siempre será de la especie que sufrió más daño a la que sufrió menos daño; según la magnitud de la proporción del número de árboles dañados con respecto al número total de árboles por especie.

3.2. Estadísticas descriptivas relevantes

Para cada una de las 17 especies consideradas se presentan estadísticas descriptivas en las Tablas (3.1), (3.2) y (3.3), donde se muestran algunos cuantiles empíricos, el promedio y la desviación estándar para las variables de interés.

La Tabla 3.1, en su primera columna etiquetada como “ANTES DE JOVA”, las estadísticas descriptivas correspondientes se obtuvieron considerando todos los árboles pertenecientes a cada especie en mayo de 2011 antes del paso del huracán Jova, tanto dañados como no dañados. Mientras que para la columna etiquetada como “DESPUÉS DE JOVA”, las estadísticas descriptivas sólo se obtuvieron con base en la información de los árboles no dañados registrados en mayo de 2012.

Por ejemplo, en la Tabla 3.1, para antes de Jova la especie *Colubrina heteroneura*, para la suma de áreas basales de tallos por árbol X_{ij} , el promedio en los 16 árboles de esta especie fue 52.71. El árbol con X_{ij} más pequeña observada fue de 12.04, la mediana de los datos para esta especie fue de 28.73 y el valor máximo observado fue de 310.57. En contraste, para después de Jova se registraron las medidas asociadas a los árboles no dañados. Se agregó una columna adicional μ_{x^*} que indica el promedio de sumas de áreas basales antes de Jova pero sólo para los árboles que no sufrieron daño. Por ejemplo, para la *Colubrina heteroneura*, $\mu_{x^*} = 72.33$ y $\mu_y = 75.43$ lo cual indica que los árboles no dañados fueron los que tenían una mayor área basal. En otras especies como la *Comocladia engleriana* ocurrió lo opuesto.

Así, interesará comparar las medias estimadas con muestras observadas antes y después

del huracán Jova para cada especie que se dan en las columnas indicadas por $\hat{\mu}_X$, $\hat{\mu}_Y$ y $\hat{\mu}_{X^*}$.

N.P.	Especie	% Daño	ANTES DE JOVA								DESPUES DE JOVA						
			n_j	μ_x	Min	Q25	Q50	Q75	Max	μ_{x^*}	n_j	μ_y	Min	Q25	Q50	Q75	Max
1	<i>Colubrina heteroneura</i>	44	16	52.71	12.04	17.14	28.73	44.64	310.57	72.33	9	75.43	15.60	19.12	35.77	74.24	312.68
2	<i>Cordia alliodora</i>	38	13	58.37	14.50	24.37	55.46	66.92	200.54	75.10	8	77.93	14.72	48.68	62.62	85.62	206.98
3	<i>Bauhinia unguolata</i>	38	16	29.77	7.33	19.18	25.85	43.15	62.43	25.69	13	26.54	7.64	17.20	21.14	28.73	63.14
4	<i>Comocladia engleriana</i>	28	25	97.52	11.46	22.46	53.08	97.25	447.30	84.80	21	86.51	11.46	23.00	54.79	97.97	371.51
5	<i>Adelia oaxacana</i>	27	15	27.65	11.46	18.10	25.78	31.56	74.03	24.40	11	24.70	12.43	18.99	25.78	29.93	40.29
6	<i>Exostema caribaeum</i>	26	35	39.46	3.90	17.55	25.78	50.34	117.97	34.96	27	35.65	4.48	17.31	26.94	49.74	104.17
7	<i>Piptadenia constricta</i>	25	20	49.54	16.27	25.64	45.84	62.71	163.30	48.86	16	50.94	17.67	25.71	46.99	55.10	173.55
8	<i>Guapira macrocarpa</i>	21	28	61.84	9.63	29.47	50.62	87.32	191.07	66.63	24	68.12	9.63	32.32	58.66	92.48	191.07
9	<i>Guettarda elliptica</i>	21	14	38.12	8.28	17.33	27.66	44.49	93.32	35.21	12	36.58	9.11	16.73	23.32	45.90	95.27
10	<i>Plumeria rubra</i>	19	32	62.33	15.15	32.07	52.34	83.58	170.59	62.47	27	64.92	18.87	33.16	55.46	90.96	158.29
11	<i>Caesalpinia eriostachys</i>	17	12	167.93	5.09	52.54	103.16	240.70	658.98	176.85	10	178.61	5.09	69.85	105.60	232.59	672.08
12	<i>Bursera instabilis</i>	9	23	164.32	18.87	93.39	127.32	241.04	561.50	176.86	21	182.15	42.10	103.13	133.77	267.70	561.50
13	<i>Helietta lottiae</i>	7	14	54.61	17.90	35.95	47.22	55.36	194.98	54.31	13	55.04	17.90	36.78	45.84	55.88	191.07
14	<i>Tabebuia impetiginosa</i>	7	14	106.55	15.60	72.59	95.01	130.05	267.70	107.63	13	109.77	15.60	73.06	99.16	130.53	274.20
15	<i>Lonchocarpus eriocarinalis</i>	6	18	74.68	12.63	35.26	47.79	93.12	223.53	68.55	17	69.37	11.46	36.78	46.99	72.58	225.22
16	<i>Samyda mexicana</i>	0	9	21.52	11.46	14.50	20.37	28.73	36.10	21.52	9	22.18	12.43	14.29	21.93	30.26	35.77
17	<i>Tabebuia chrysantha</i>	0	9	100.11	10.89	33.44	58.76	148.51	346.64	100.11	9	105.27	12.24	36.78	61.22	150.58	377.77

Tabla 3.1: Medias estimadas y algunos cuantiles empíricos de las sumas de áreas basales de los tallos por árbol (X_{ij}) para antes y después de Jova. La columna μ_{x^*} indica el promedio de áreas basales antes de Jova pero sólo para los árboles que no sufrieron daño.

Se puede observar una asimetría en la densidad de las sumas de áreas basales de tallos por especie observando la diferencia que hay entre el promedio y la mediana, tanto de antes como después del huracán Jova. En algunas especies, esta asimetría es más marcada que en otras.

Las sumas de áreas basales de tallos por árbol en promedio para cada especie varían bastante, la razón de ello es que los tallos de árboles pueden ser más gruesos en una especie que en otra y también que el número de tallos de árboles entre una especie y otra pueden ser diferentes. Por tanto es importante analizar cada especie por separado. Se observa que en 11 de las 17 especies después del paso del huracán Jova, el promedio de la suma de las áreas basales fue mayor.

La Tabla 3.2 presenta las estadísticas descriptivas del número de tallos por árbol que hay en cada una de las especies consideradas. En general, el número de tallos por árbol oscilan entre 1 y 7 tallos por árbol.

Solamente para las especies *Exostema caribaeum*, *Guapira macrocarpa* *Plumeria rubra*, el máximo número de tallos por árbol en estas especies disminuyó en uno después del paso del huracán Jova. Esto indica que entre los árboles dañados para estas especies, los árboles que tenían más tallos antes del paso del huracán fueron de los dañados.

N.P.	Especie	% Daño	ANTES DE JOVA							DESPUES DE JOVA						
			n_j	\bar{x}	Min	Q25	Q50	Q75	Max	n_j	\bar{y}	Min	Q25	Q50	Q75	Max
1	<i>Colubrina heteroneura</i>	44	16	2	1	1	1	2	4	9	2	1	1	1	2	4
2	<i>Cordia alliodora</i>	38	13	1	1	1	1	1	1	8	1	1	1	1	1	1
3	<i>Bauhinia unguolata</i>	38	16	2	1	1	1	4	5	10	2	1	1	1	1	5
4	<i>Comocladia engleriana</i>	28	25	3	1	1	2	4	7	18	3	1	1	2	3	7
5	<i>Adelia oaxacana</i>	27	15	1	1	1	1	1	2	11	1	1	1	1	1	2
6	<i>Exostema caribaeum</i>	26	35	1	1	1	1	1	3	26	1	1	1	1	1	2
7	<i>Piptadenia constricta</i>	25	20	1	1	1	1	1	3	15	1	1	1	1	1	3
8	<i>Guapira macrocarpa</i>	21	28	1	1	1	1	1	3	22	1	1	1	1	1	2
9	<i>Guettarda elliptica</i>	21	14	2	1	1	2	2	6	11	2	1	1	1	2	6
10	<i>Plumeria rubra</i>	19	32	1	1	1	1	1	3	26	1	1	1	1	1	2
11	<i>Caesalpinia eriostachys</i>	17	12	2	1	1	1	2	4	10	2	1	1	1	2	4
12	<i>Bursera instabilis</i>	9	23	1	1	1	1	1	1	21	1	1	1	1	1	1
13	<i>Helietta lottiae</i>	7	14	1	1	1	1	1	2	13	1	1	1	1	1	2
14	<i>Tabebuia impetiginosa</i>	7	14	1	1	1	1	1	1	13	1	1	1	1	1	1
15	<i>Lonchocarpus eriocarinalis</i>	6	18	1	1	1	1	1	2	17	1	1	1	1	1	2
16	<i>Samyda mexicana</i>	0	9	1	1	1	1	1	1	9	1	1	1	1	1	1
17	<i>Tabebuia chrysantha</i>	0	9	1	1	1	1	1	2	9	1	1	1	1	1	2

Tabla 3.2: Número de tallos por árbol por especie (Y_{ij}).

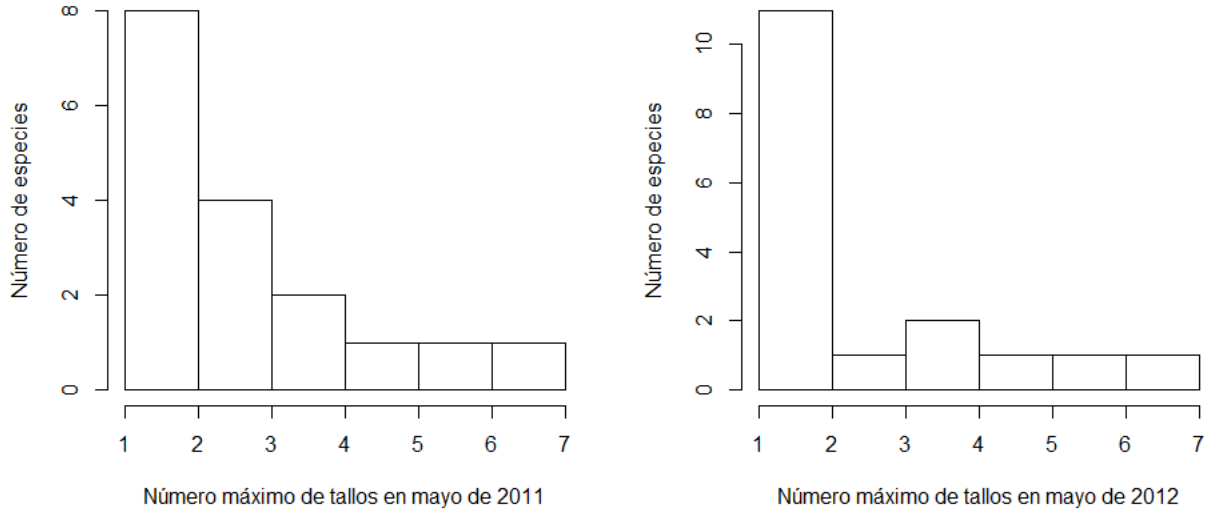


Figura 3.1: Histograma de frecuencias del número de especies con su máximo número de tallos.

La Tabla 3.3 muestra, por especie, el número de tallos correspondientes a cada árbol antes y después del paso del huracán sólo para aquellos árboles que sufrieron daño. La mayoría de los árboles dañados perdieron todos sus tallos como se ve en esta tabla. Por ejemplo, se observa que para la *Colubrina heteroneura*, de 17 árboles que se midieron antes del huracán, 7 de ellos presentaron daño perdiendo todas sus tallos. Estos árboles tenían una o dos ramas antes del huracán. En contraste, para la *Bauhinia unguolata*, se tenían 16 árboles antes de Jova y 6 de ellos resultaron dañados.

N.P.	Especie	n_j	Z_j	% Daño	Número de tallos antes y después de Jova en árboles dañados (y_{ij}, y_{ij}^*)									
1	<i>Colubrina heteroneura</i>	16	7	44	(2,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)			
2	<i>Cordia alliodora</i>	13	5	38	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)					
3	<i>Bauhinia unguolata</i>	16	6	38	(4,1)	(4,1)	(4,0)	(3,2)	(2,0)	(1,0)				
4	<i>Comocladia engleriana</i>	25	7	28	(6,5)	(5,1)	(4,3)	(4,0)	(2,0)	(1,0)	(1,0)			
5	<i>Adelia oaxacana</i>	15	4	27	(1,0)	(1,0)	(1,0)	(1,0)						
6	<i>Exostema caribaeum</i>	35	9	26	(3,0)	(2,1)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
7	<i>Piptadenia constricta</i>	20	5	25	(2,1)	(2,0)	(2,0)	(1,0)	(1,0)					
8	<i>Guapira macrocarpa</i>	28	6	21	(3,2)	(2,1)	(1,0)	(1,0)	(1,0)	(1,0)				
9	<i>Guettarda elliptica</i>	14	3	21	(3,0)	(4,3)	(1,0)							
10	<i>Plumeria rubra</i>	32	6	19	(3,2)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)				
11	<i>Caesalpinia eriostachys</i>	12	2	17	(3,0)	(1,0)								
12	<i>Bursera instabilis</i>	23	2	9	(1,0)	(1,0)								
13	<i>Helietta lottiae</i>	14	1	7	(1,0)									
14	<i>Tabebuia impetiginosa</i>	14	1	7	(1,0)									
15	<i>Lonchocarpus eriocarinalis</i>	18	1	6	(1,0)									
16	<i>Samyda mexicana</i>	9	0	0										
17	<i>Tabebuia chrysantha</i>	9	0	0										

Tabla 3.3: Número de árboles dañados por especie y número de tallos que tenían los dañados antes y después de Jova.

3.3. Modelación de la proporción de daño por especie en Jova con un modelo binomial

Para estimar la proporción de daño por especie a causa de Jova se considerará la variable aleatoria

$$Z_j = \text{número de árboles dañados en la especie } j, \text{ para } j = 1, 2, \dots, 17,$$

para poder estimar la proporción p_j de árboles dañados por cada especie. La variable aleatoria Z_j sigue una distribución binomial de parámetros (n_j, p_j) donde para todas las especies n_j es conocido. De esta forma el interés recae en estimar el valor del parámetro de proporción p_j , que representa también la probabilidad de que un árbol de la especie j -ésima haya tenido uno o más tallos dañados en 2012, en comparación a la medición hecha en 2011.

La Tabla 3.4 presenta el porcentaje observado de pérdida por especie, así como el número de árboles dañados. La especie más dañada fue la *Colubrina heteroneura*, con un daño considerable del 44 %. Le siguen la *Cordia alliodora*, *Adelia oaxacana*, *Exostema caribaeum*, *Piptadenia constricta* con un daño entre el 20 % y el 40 %. Mientras que las doce especies restantes presentaron un daño menor al 20 %. Tanto la *Samyda mexicana* como la *Tabebuia chrysantha* no presentaron daño alguno.

El porqué algunas especies resultaron más dañadas que otras, podría adjudicarse al grosor de las ramas de los árboles por especie o a la ubicación espacial en la que se encontraban los árboles; inclusive también por el tipo de características específicas de árbol en cada especie. Por ejemplo, si los árboles almacenan agua; esto los puede hacer más o menos resistentes, si son muy delgados o altos tienden a tener menor resistencia o mayor elasticidad, etc.

3.3. MODELACIÓN DE LA PROPORCIÓN DE DAÑO POR ESPECIE EN JOVA CON UN MODELO BINOMIAL

N.P.	Especie	n_j	Z_j	% Daño
1	<i>Colubrina heteroneura</i>	16	7	44
2	<i>Cordia alliodora</i>	13	5	38
3	<i>Bauhinia unguolata</i>	16	6	38
4	<i>Comocladia engleriana</i>	25	7	28
5	<i>Adelia oaxacana</i>	15	4	27
6	<i>Exostema caribaeum</i>	35	9	26
7	<i>Piptadenia constricta</i>	20	5	25
8	<i>Guapira macrocarpa</i>	28	6	21
9	<i>Guettarda elliptica</i>	14	3	21
10	<i>Plumeria rubra</i>	32	6	19
11	<i>Caesalpinia eriostachys</i>	12	2	17
12	<i>Bursera instabilis</i>	23	2	9
13	<i>Helietta lottiae</i>	14	1	7
14	<i>Tabebuia impetiginosa</i>	14	1	7
15	<i>Lonchocarpus eriocarinalis</i>	18	1	6
16	<i>Samyda mexicana</i>	9	0	0
17	<i>Tabebuia chrysantha</i>	9	0	0

Tabla 3.4: Número de árboles dañados (Z_j) por especie, $j = 1, 2, \dots, 17$.

La Figura 3.2 y la Tabla 3.5 muestran el intervalo de estimación de verosimilitud que tiene un 95% de confianza para estimar la proporción de árboles dañados por especie, p_j . Cuatro de éstos intervalos no descartan la proporción del 0.5 que corresponden a un daño de la mitad de árboles de esa especie para 4 de las especies más dañadas: *Colubrina heteroneura*, *Cordia alliodora*, *Bauhinia unguolata* y *Adelia oaxacana*. Esto quiere decir que para estas 4 especies es razonable que la verdadera proporción dañada de ellas fuese la mitad.

Para las otras 13 especies restantes, la pérdida estimada fue de menos de la mitad de sus árboles, dentro de los límites indicados por cada intervalo dado en la Tabla (3.5), con una probabilidad del 95%. En contraste, hubo cinco especies cuyos intervalos incluyen al cero.

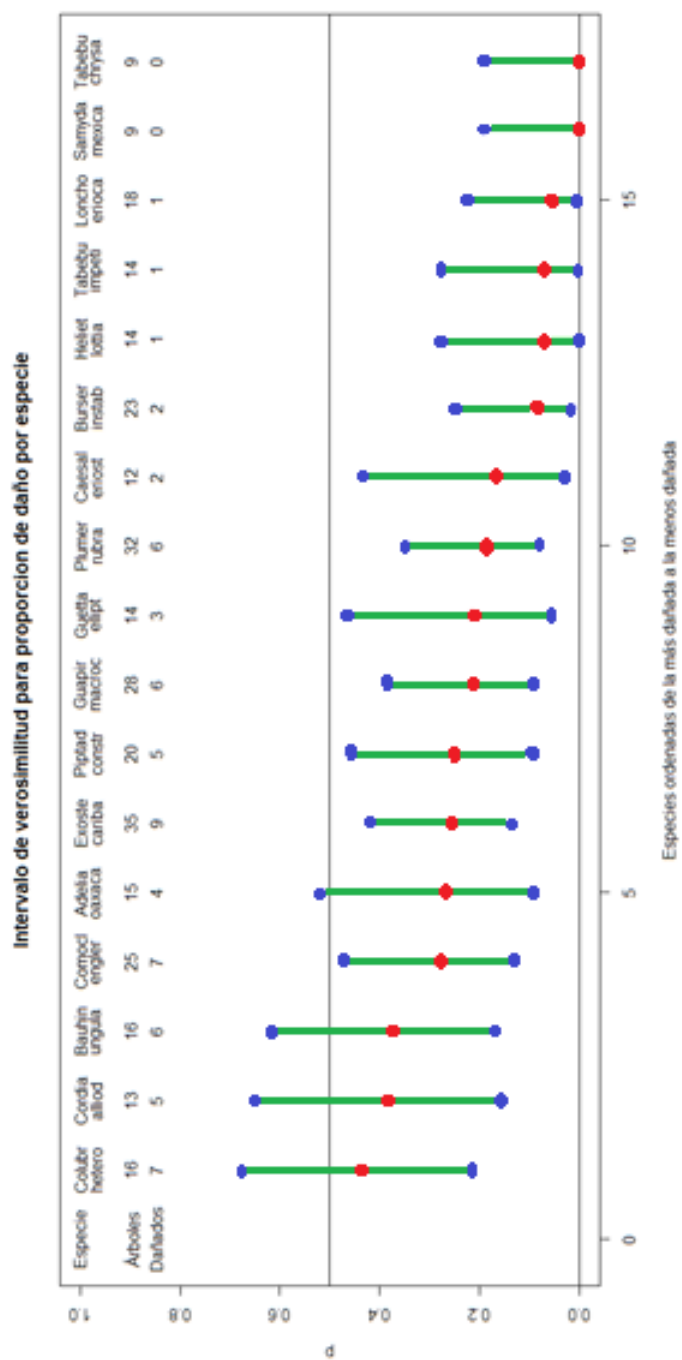


Figura 3.2: Intervalos de verosimilitud para la proporción de daño p_j por especie.

3.3. MODELACIÓN DE LA PROPORCIÓN DE DAÑO POR ESPECIE EN JOVA CON UN MODELO BINOMIAL

N.P.	Especie	\hat{p}_j	$IV_{p_j}(0.1465)$
1	<i>Colubrina heteroneura</i>	0.44	(0.22,0.68)
2	<i>Cordia alliodora</i>	0.38	(0.16,0.65)
3	<i>Bauhinia unguolata</i>	0.38	(0.17,0.62)
4	<i>Comocladia engleriana</i>	0.28	(0.13,0.47)
5	<i>Adelia oaxacana</i>	0.27	(0.09,0.52)
6	<i>Exostema caribaeum</i>	0.26	(0.13,0.42)
7	<i>Piptadenia constricta</i>	0.25	(0.10,0.46)
8	<i>Guapira macrocarpa</i>	0.21	(0.09,0.39)
9	<i>Guettarda elliptica</i>	0.21	(0.06,0.47)
10	<i>Plumeria rubra</i>	0.19	(0.08,0.34)
11	<i>Caesalpinia eriostachys</i>	0.17	(0.03,0.43)
12	<i>Bursera instabilis</i>	0.09	(0.02,0.25)
13	<i>Helietta lottiae</i>	0.07	(0,0.28)
14	<i>Tabebuia impetiginosa</i>	0.07	(0,0.28)
15	<i>Lonchocarpus eriocarinalis</i>	0.06	(0,0.22)
16	<i>Samyda mexicana</i>	0	(0,0.19)
17	<i>Tabebuia chrysantha</i>	0	(0,0.19)

Tabla 3.5: EMVs e intervalos del 14.65 % y del 95 % de confianza para la proporción, p_j , de árboles dañados por especie.

Para estas especies (*Helietta lottiae*, *Tabebuia impetiginosa*, *Tabebuia chrysantha*, *Lonchocarpus eriocarinalis* y *Samyda mexicana*) la proporción de daño poblacional fue pequeña o nula.

El hecho de que el EMV no es el punto medio de estos intervalos indica que la función de verosimilitud de p_j es en general muy asimétrica para la mayoría de las especies.

3.4. Modelo normal para el área basal de tallos antes y después de Jova

En la Figura 3.3, se muestra la gráfica Cuantil-Cuantil para la variable aleatoria X_{i2} , de la especie *Cordia alliodora* antes del huracán Jova contra los cuantiles teóricos de una distribución normal. Se identifica que no existe algún patrón en los árboles dañados, es decir, están distribuidos entre los árboles más delgados a los más gruesos de esa especie y no tienen una característica común que los distinga. Comportamientos similares ocurrieron para todas las demás especies. En este caso se observó que un modelo normal ajustaba bien marginalmente a todas las especies en cuanto a la medición X_{ij} .

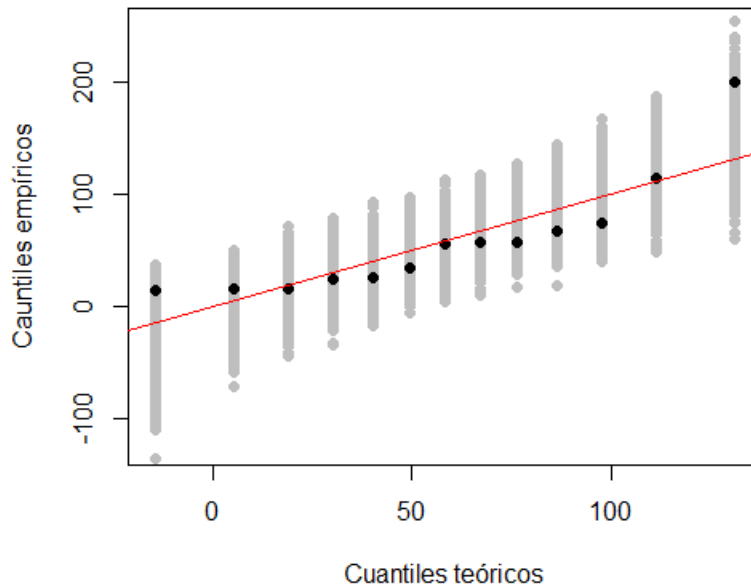


Figura 3.3: Gráfica Q-Q de sumas de áreas basales de tallos por árbol de la especie *Cordia alliodora* antes del paso del huracán Jova.

3.4. MODELO NORMAL PARA EL ÁREA BASAL DE TALLOS ANTES Y DESPUÉS DE JOVA 51

Al realizar una gráfica Cuantil-Cuantil para cualquiera de las especies de los árboles después del huracán Jova, claramente se identifican dos grupos, el de los árboles dañados y el de los no dañados. Esto debido a que en todas las 17 especies consideradas, los árboles que se dañaron fue prácticamente en la totalidad de sus tallos. Por esta razón para los árboles no dañados, es de interés estimar y cuantificar el crecimiento o pérdida de grosor por especie. Por lo cual en la siguiente sección se estimará el crecimiento de los árboles considerando un modelo normal para las diferencias entre el antes y después, así como también un modelo normal bivariado.

A manera de ejemplo se muestra la gráfica Cuantil-Cuantil normal para la especie *Cordia alliodora*, donde cada observación corresponde a aquellos árboles que no sufrieron daño después del paso del huracán y que ajustan muy bien marginalmente a una distribución normal. Algo similar ocurre para las demás especies.

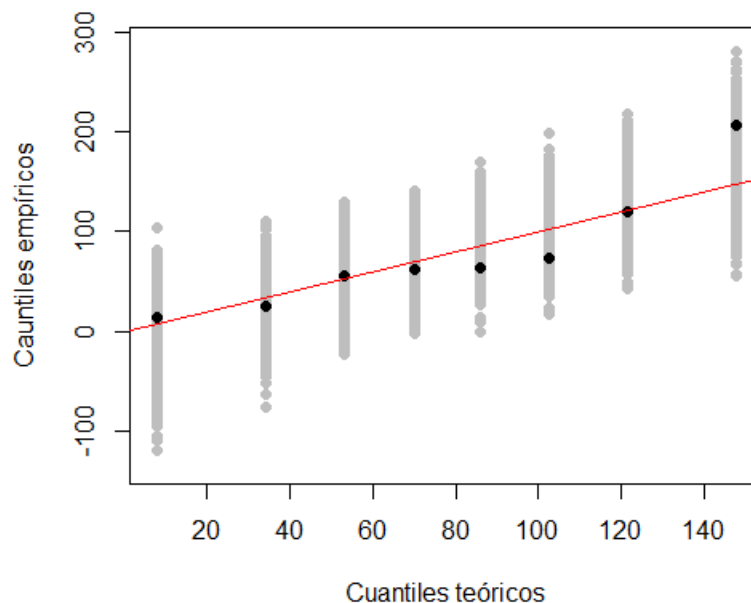


Figura 3.4: Gráfica Q-Q de sumas de áreas basales de tallos por árbol no dañado de la especie *Cordia alliodora* después del paso del huracán Jova.

3.5. Estimación del crecimiento exclusivamente en árboles no dañados

Para cuantificar el crecimiento en los árboles no dañados, por especie se consideró un modelo normal para las diferencias de los datos pareados de los después y antes de Jova que es lo que comúnmente se hace en la práctica. También se contempló un modelo normal bivariado para aquellas especies que eran descritas bien por esta distribución, para así poder tomar en cuenta a la correlación, la cual el primer modelo ignora por completo.

3.5.1. Modelo normal para las diferencias $(X_{ij}^* - X_{ij})$

Como se describió en la Sección 2.6.1, se consideraron las diferencias $X_{ij}^* - X_{ij}$ entre el después de Jova y el antes de Jova por cada especie. Para ello primero se procedió a verificar que en efecto un modelo normal ajustara bien a las diferencias $\delta_{ij} = X_{ij}^* - X_{ij}$, para $i = 1, 2, \dots, n_j$. Como ya se había verificado a través de gráficas cuantil-cuantil que $X_{.j}^*$ y $X_{.j}$ fueron normales para todas las especies, basta con mostrar ahora una gráfica cuantil-cuantil para las $\delta_{.j}$. Para la especie *Cordia Alliodora* se presenta esto en la Figura (3.5).

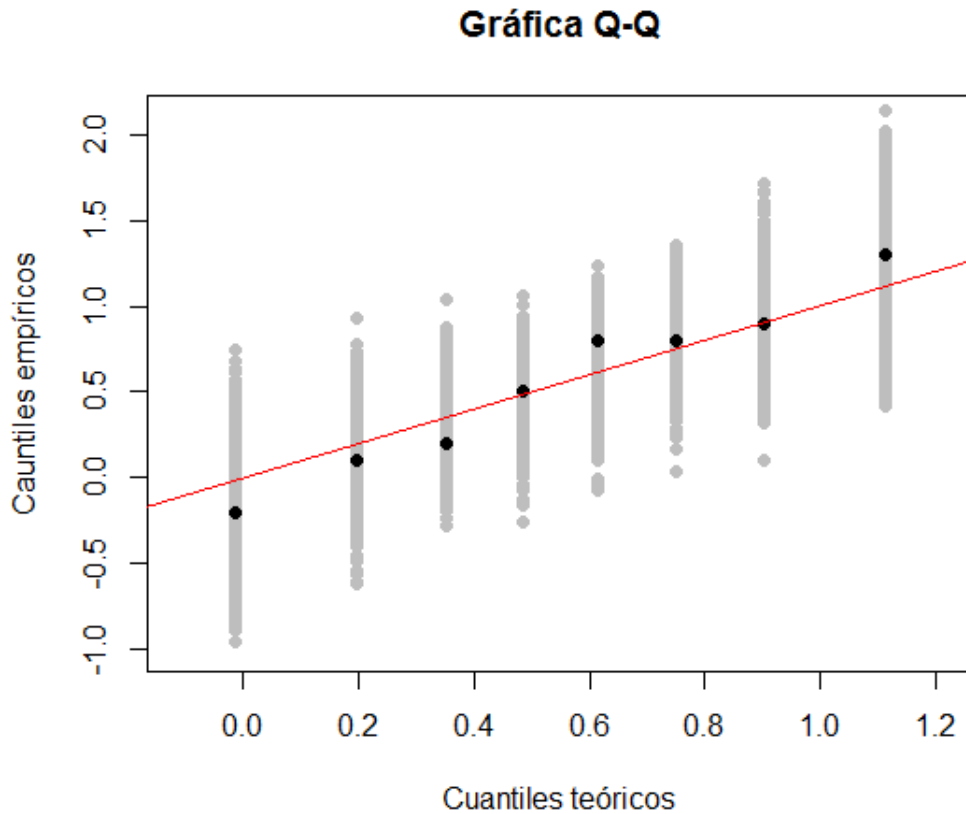


Figura 3.5: Gráfica Q-Q de diferencias $X_{ij}^* - X_{ij}$ de la especie *Cordia allidora*.

Una vez realizado esto, se procedió a estimar el crecimiento en áreas basales de los árboles en cada especie, para lo cual se obtuvo un intervalo del 95% de confianza para la diferencia de medias normales $\mu_{X_{ij}^*} - \mu_{X_{ij}}$, como se describió en la Sección 2.3.2.

La Tabla 3.6 muestran dichos intervalos, en los cuales se observa que no todos excluyen al cero. Los intervalos que yacen por arriba del cero sustentan que sí hubo un crecimiento en las medias. Con base en ellos, la especie que presenta el mayor crecimiento es la *Bursera instabilis*, seguida de la *Plumeria rubra*, la tercera fue *Colubrina heteroneura* que sorprendentemente fue la especie con proporción más alta de daño.

Los intervalos que incluyen al cero corresponden a especies para las cuales no hay evidencia en los datos que haya habido crecimiento. Cabe resaltar que de las 17 especies hubo 13 especies cuyos intervalos cayeron a la derecha del cero y por tanto sustentan que sí hubo crecimiento en las medias de sus áreas basales.

N.P.	Especie	$\hat{\delta}_j$	$IV_{\hat{\delta}_j}(0.1465)$
1	<i>Colubrina heteroneura</i>	3.10	(1.14,5.06)
2	<i>Cordia alliodora</i>	2.83	(0.93,4.73)
3	<i>Bauhinia unguolata</i>	0.30	(-0.02,0.61)
4	<i>Comocladia engleriana</i>	0.69	(0.39,0.99)
5	<i>Adelia oaxacana</i>	2.07	(0.85,3.29)
6	<i>Exostema caribaeum</i>	0.85	(0.14,1.56)
7	<i>Piptadenia constricta</i>	1.88	(-0.57,4.33)
8	<i>Guapira macrocarpa</i>	1.71	(0.65,2.77)
9	<i>Guettarda elliptica</i>	2.45	(1.41,3.48)
10	<i>Plumeria rubra</i>	1.49	(0.72,2.26)
11	<i>Caesalpinia eriostachys</i>	1.36	(0.33,2.40)
12	<i>Bursera instabilis</i>	5.30	(3.00,7.59)
13	<i>Helietta lottiae</i>	0.73	(-0.12,1.57)
14	<i>Tabebuia impetiginosa</i>	2.15	(0.99,3.30)
15	<i>Lonchocarpus eriocarinalis</i>	0.82	(0.21,1.43)
16	<i>Samyda mexicana</i>	0.67	(0.06,1.28)
17	<i>Tabebuia chrysantha</i>	5.17	(-0.92,11.26)

Tabla 3.6: EMVs e intervalo del 95 % de confianza para la diferencia de medias normales de la suma de áreas basales de árboles por especie, antes y después de Jova.

3.5.2. Modelo normal bivariado

Como se describió en la Sección 2.2.3, el modelo de la normal bivariada considera la correlación, aspecto importante a tener en cuenta cuando se tienen datos pareados normales ya que

se podrían obtener resultados incompletos de no tomarla en cuenta en la interpretación de los resultados como se verá más adelante.

Al tomar en cuenta la correlación, la estimación de la diferencia de medias se vuelve más precisa y se mejora; esto sujeto a que el modelo bivariado sí describa bien a los datos. Para verificar si este modelo describía bien a los datos de las 17 especies, se les aplicó la prueba de normalidad multivariada de Mardia (1974). Con base en esta prueba se vio que el modelo bivariado era razonable para 5 de las 17 especies consideradas. Las estimaciones de los parámetros de interés se muestran en la Tabla 3.7 sólo para las especies a las que sí les quedó bien el modelo bivariado.

En la Tabla C.1 del Apéndice C se muestran los resultados de la prueba de Mardia aplicada a cada una de las 17 especies consideradas, de las cuales se verifica que cinco de ellas siguen una distribución normal bivariada. Estas especies son: *Cordia alliodora*, *Adelia oaxacana*, *Exostema caribaeum*, *Tabebuia impetiginosa* y *Samyda mexicana*.

Los estimadores de máxima verosimilitud y los intervalos de estimación para la razón de medias, $\beta = \frac{\mu_y}{\mu_x}$, resultaron ser mayores que uno. Esto significa que la media de áreas basales de árboles para después del paso del huracán es ligeramente mayor a la media de antes, lo que se traduce en un crecimiento de área basal en los árboles.

Al analizar la estimación de la diferencia de medias de áreas de basales $\delta = \mu_y - \mu_x$ se concluye de manera similar a lo que indicó la estimación de β que sí hay evidencia de crecimiento en las medias de las áreas basales estas especies.

Por último, si se considera la correlación estimada $\hat{\rho}$, se puede observar que es alta para estas especies, lo cual quiere decir que los datos (X_{ij}, X_{ij}^*) para $j = 1, 2, \dots, 17$, de mayo de 2011 y mayo de 2012 están altamente correlacionados por especie y que al tomar en cuenta

en el modelo propuesto esta correlación se mejorará notablemente las inferencias sobre la comparación de medias, siendo el intervalo de estimación más angosto y preciso.

N.P.	Especie	$\hat{\beta}_j$	$IV_{\beta_j}(0.1465)$	$\hat{\delta}_j$	$IV_{\delta_j}(0.1465)$	$\hat{\tau}_j$	$IV_{\tau_j}(0.1465)$	$\hat{\rho}_j$	$IV_{\rho_j}(0.1465)$
1	<i>Cordia alliodora</i>	1.04	(1.02,1.06)	2.83	(1.44,4.22)	1.03	(1.01,1.06)	0.9994	(0.9981,0.9997)
2	<i>Adelia oaxacana</i>	1.04	(1.03,1.06)	2.07	(1.26,2.89)	1.06	(1.04,1.09)	0.9989	(0.9977,0.9994)
3	<i>Exostema caribaeum</i>	1.03	(1.01,1.06)	0.85	(0.17,1.53)	1.04	(1.00,1.09)	0.9969	(0.9924,0.9984)
4	<i>Tabebuia impetiginosa</i>	1.02	(1.01,1.03)	2.15	(1.19,3.10)	1.02	(1.01,1.04)	0.9996	(0.9991,0.9998)
5	<i>Samyda mexicana</i>	1.03	(1.00,1.06)	0.67	(0.07,1.27)	1.01	(0.94,1.09)	0.9934	(0.9804,0.9970)

Tabla 3.7: EMVs e intervalos de verosimilitud perfil para los parámetros de la normal biva-riada para la suma de áreas basales de tallos para el antes y después de Jova.

3.5.3. Comparación entre modelos utilizados

En la Tabla 3.8 se muestran los intervalos de verosimilitud-confianza para la diferencia δ obtenidos del mejor modelo para cada una de las 17 especies consideradas.

N.P.	Especie	$\hat{\delta}_j$	$IV_{\delta_j}(0.1465)$	Mejor modelo
1	<i>Colubrina heteroneura</i>	3.10	(1.14,5.06)	Normal para diferencias
2	<i>Cordia alliodora</i>	2.83	(1.44,4.22)	Normal bivariado
3	<i>Bauhinia unguolata</i>	0.30	(-0.02,0.61)	Normal para diferencias
4	<i>Comocladia engleriana</i>	0.69	(0.39,0.99)	Normal para diferencias
5	<i>Adelia oaxacana</i>	2.07	(1.26,2.89)	Normal bivariado
6	<i>Exostema caribaeum</i>	0.85	(0.17,1.53)	Normal bivariado
7	<i>Piptadenia constricta</i>	1.88	(-0.57,4.33)	Normal para diferencias
8	<i>Guapira macrocarpa</i>	1.71	(0.65,2.77)	Normal para diferencias
9	<i>Guettarda elliptica</i>	2.45	(1.41,3.48)	Normal para diferencias
10	<i>Plumeria rubra</i>	1.49	(0.72,2.26)	Normal para diferencias
11	<i>Caesalpinia eriostachys</i>	1.36	(0.33,2.40)	Normal para diferencias
12	<i>Bursera instabilis</i>	5.30	(3.00,7.59)	Normal para diferencias
13	<i>Helietta lottiae</i>	0.73	(-0.12,1.57)	Normal para diferencias
14	<i>Tabebuia impetiginosa</i>	2.15	(1.19,3.10)	Normal bivariado
15	<i>Lonchocarpus eriocarinalis</i>	0.82	(0.21,1.43)	Normal para diferencias
16	<i>Samyda mexicana</i>	0.67	(0.07,1.27)	Normal bivariado
17	<i>Tabebuia chrysantha</i>	5.17	(-0.92,11.26)	Normal para diferencias

Tabla 3.8: Intervalo de verosimilitud-confianza para la diferencia de medias con el mejor modelo para las 17 especies consideradas

En la Figura 3.6 se muestran los intervalos de confianza obtenidos usando un modelo normal para la comparación de la diferencia de medias para las 17 especies consideradas y también los intervalos de verosimilitud obtenidos a través del modelo normal bivariado para la diferencia de medias en aquellas especies a las cuales les quedó bien el modelo bivariado.

Se observa que los intervalos bajo el modelo bivariado son más angostos que bajo el

primer modelo, mostrando que al modelar mejor la correlación entre los pares, se puede dar una estimación más precisa para δ .

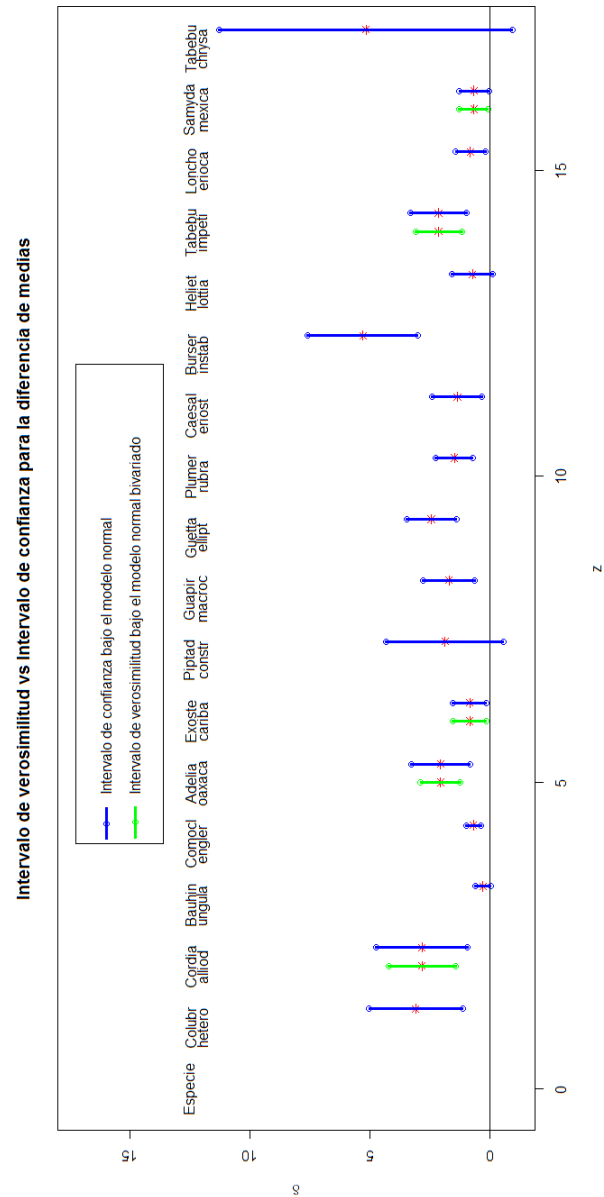


Figura 3.6: Comparación entre los modelos normal uni y bivariado en cuanto a la diferencia de medias

3.6. Comparación entre dos pares de años hidrológicos similares en cuanto a patrones de lluvia

En general, se encontró que la mayoría de las especies consideradas presentaron un crecimiento significativo de diferente magnitud, según la especie. Sin embargo, resulta interesante comparar ese crecimiento con la de algún otro par de años para ver si hay diferencias y poder identificar el efecto aislado del huracán Jova.

En la base de datos original existe información de otros años anteriores. Sin embargo, se cuenta con información de árboles distintos a los del par de años consecutivos 2011-2012. A pesar de esto, la comparación del crecimiento se puede hacer por especie aunque los árboles no sean los mismos. Los años elegidos para realizar la comparación fueron los años 2000 y 2001, porque estos años siguieron patrones de lluvias similares al par (2011-2012) siendo el primer año La Niña y el segundo normal.

También ajustó bien un modelo normal para las diferencias entre los pares de datos (X_{ij}, X_{ij}^*) para el par de años 2000-2001. La comparación entre los pares de años considerados se presenta en la Tabla 3.9.

Se observa que en la mayoría de las especies la razón de diferencias de medias entre el par de años (2011-2012) y el par (2000-2001) es mayor a 1. Esto se da con excepción de dos especies, la *Exostema caribaeum* y la *Guettarda elliptica*, lo cual indica que el crecimiento fue mayor en el año hidrológico 2011-2012 con excepción de las dos especies mencionadas.

N.P.	Especie	Años 2000-2001			Años 2011-2012			$\left \frac{\hat{\delta}_2}{\hat{\delta}_1} \right $
		$\hat{\mu}_x$	$\hat{\mu}_y$	$\hat{\delta}_1$	$\hat{\mu}_x$	$\hat{\mu}_y$	$\hat{\delta}_2$	
1	<i>Colubrina heteroneura</i>	56.70	57.27	0.57	70.64	73.61	2.98	5.25
2	<i>Cordia alliodora</i>	66.94	68.97	2.03	86.83	90.23	3.40	1.68
3	<i>Bauhinia unguolata</i>	17.41	17.81	0.40	26.87	27.84	0.97	2.43
4	<i>Comocladia engleriana</i>	69.74	70.51	0.78	83.20	84.81	1.61	2.07
5	<i>Adelia oaxacana</i>	18.47	18.36	-0.11	22.30	22.66	0.36	3.41
6	<i>Exostema caribaeum</i>	29.89	30.57	0.68	39.43	40.04	0.62	0.91
7	<i>Piptadenia constricta</i>	26.03	26.62	0.59	41.19	42.62	1.43	2.41
8	<i>Guapira macrocarpa</i>	54.96	54.82	-0.14	66.93	68.43	1.50	10.49
9	<i>Guettarda elliptica</i>	26.09	27.54	1.46	36.60	37.92	1.32	0.91
10	<i>Plumeria rubra</i>	43.87	45.18	1.31	62.47	64.92	2.45	1.87
11	<i>Caesalpinia eriostachys</i>	156.50	157.45	0.96	191.38	193.43	2.05	2.14
12	<i>Bursera instabilis</i>	149.50	153.26	3.76	183.59	189.15	5.56	1.48
13	<i>Helietta lottiae</i>	40.16	40.50	0.34	53.06	53.73	0.67	1.96
14	<i>Tabebuia impetiginosa</i>	83.54	84.36	0.83	104.05	106.37	2.32	2.81
15	<i>Lonchocarpus eriocarinalis</i>	57.53	57.52	-0.01	68.02	68.77	0.76	71.01
16	<i>Samyda mexicana</i>	18.65	18.80	0.15	22.77	23.40	0.63	4.19
17	<i>Tabebuia chrysantha</i>	100.10	102.05	1.95	125.29	131.68	6.39	3.27

Tabla 3.9: EMVs de la normal bivariada para las sumas de áreas basales de tallos para dos parejas de años consecutivos (2000-2001) y (2011-2012).

Sin embargo, de la información que proporciona la Tabla 3.9, no es posible atribuirle exclusivamente al huracán Jova el efecto del crecimiento en los árboles. Esto debido a que la cantidad de lluvia anual en 2011 fue substancialmente mayor que en la del año 2000 como se muestra en la Tabla 3.10. Aún si se omitiesen las lluvias acumuladas en el mes de octubre, mes en el cual pasó el huracán Jova, sigue existiendo una diferencia considerable de lluvia anual registrada con el paso del huracán, por lo cual no se le puede atribuir solamente al agua adicional que trajo el huracán Jova el crecimiento de los árboles en el 2011, sino más bien que en dicho año llovió mucho más en comparación al año 2000. Es decir, los años 2000 y 2001 fueron años secos en contraste con 2011 y 2012. Por tanto, la última columna de la Tabla 3.9 permite evaluar para cada especie el efecto en su crecimiento al contar con una mayor cantidad de lluvia acumulada anual. Para todas las especies el crecimiento es notoriamente

3.6. COMPARACIÓN ENTRE DOS PARES DE AÑOS HIDROLÓGICOS SIMILARES EN CUANTO A PATRONES DE LLUVIA

mayor si hay más agua, como era de esperarse.

Año hidrológico	Tipo de año	LLuvia anual promedio	LLuvia anual acumulada	LLuvia anual acumulada excepto octubre
2000-2001	Niña débil	848.97	568	505
2011-2012	Niña débil		1178.5	857.5

Tabla 3.10: Lluvias anuales acumuladas en milímetros para dos parejas de años consecutivos.

A través del análisis hecho, se observó que cada especie de las 17 consideradas en esta tesis sufrió cambios de diferente magnitud. Este cambio no necesariamente es atribuible en su totalidad al huracán Jova, ya que las mediciones se tomaron en mayo de 2011 y mayo de 2012, por lo cual cualquier otro evento que haya sucedido en el transcurso de este tiempo pudo haber influido en el daño observado en cada especie.

Por otro lado, en la estimación del crecimiento en árboles que no sufrieron daño, se encontró que con el modelo de diferencias normales era razonable para todas las especies mientras que el modelo normal bivariado era razonable para cinco de las 17 especies. Debido a que el modelo normal bivariado sí considera a la correlación alta existente en los pares de datos las estimaciones de los intervalos de verosimilitud-confianza son más precisos como muestra la Figura 3.6.

Al analizar el efecto del OSEN y las lluvias anuales acumuladas, se encontró que si bien el huracán Jova trajo consigo mucha agua para la reserva de Chamela, no necesariamente el crecimiento de los árboles en cada especie es atribuible solamente al efecto del huracán. Esto pues, en el año 2011 la cantidad de lluvia anual acumulada fue ligeramente el doble de magnitud en comparación con el año 2000.

Capítulo 4

Conclusiones generales

Con respecto a una evaluación global del daño ocasionado tanto por el huracán Jova como algún otro evento ocurrido entre mayo de 2011 y mayo de 2012 se determinó que sí hubo proporciones importantes de daño de magnitudes no mayores al 60 % de los árboles en cada especie de las 17 consideradas, y los efectos del daño fueron distintos para cada especie. Los factores antrópico quedaron fuera de sospecha como factor de daño pues el cuadrante de muestreo se localiza en una zona dentro de una reserva con la vegetación muy bien conservada. Aunque existen otros posibles factores de corte natural como lo son las sequías y plagas, la ocurrencia de árboles caídos en la vegetación natural es muy poco frecuente, ya que la estructura de la vegetación es dominada por tallos delgados y de poca altura, acompañados de una abundancia de bejucos o lianas lo cual promueve que los árboles mueran “en pie”.

Se observó también un crecimiento significativo en las áreas basales de los árboles no dañados de cada especie. Este hecho no puede atribuirse directamente al huracán Jova sino posiblemente al efecto de una mayor cantidad de agua recibida en el año 2011-2012 al ser años consecutivos La Niña seguido de un normal, en contraste con el par de años más secos del 2000-2001. Un argumento a favor de esto es que el agua es el factor limitante más importante de este tipo de vegetación. Asimismo, se ha documentado que con el paso del huracán Jova, se dio un aumento significativo en el ingreso de materia orgánica producto de un aumento en

la caída de hojarasca y material vegetal por efecto de los fuertes vientos. Esto probablemente generó un efecto fertilizador en el suelo que, aunado a una mayor disponibilidad de agua, detonaron una mayor productividad.

Por otro lado, en el aprendizaje de modelación para datos pareados se recomienda tomar en cuenta la estructura de los datos, realizar una verificación correcta de los supuestos de cada método, identificar grupos o poblaciones diferentes como lo fue en este caso el de los árboles dañados y el segundo grupo de los árboles no dañados que permitan proponer un mejor modelo para los datos.

Se recomienda usar el modelo normal bivariado cuando sea razonable para estimar la diferencia de medias y obtener intervalos de verosimilitud para este parámetro de interés ya que sí incorpora la marcada correlación existente en los pares de datos obteniéndose así una estimación más precisa. Cuando esto no sea posible, para las especies restantes en Chamela se pudo estimar la diferencia de medias tras suponer que las diferencias de áreas basales para un mismo árbol en los años consecutivos considerados seguía una distribución normal univariada.

En cuanto a la gravedad de Jova, fue un huracán que trajo daños no muy severos en el cuadrante de muestreo que se estudió. La mayor cantidad de lluvia en el par de años 2011-2012 en contraste con 2000-2001 sí trajo consigo un crecimiento mayor en las áreas basales de los árboles. Por tanto, a pesar que la proporción de daño por especie no rebasó la mitad de árboles para la mayoría de especies, habrá que evaluar el ritmo y velocidad de recuperación del ecosistema en los años siguientes.

Apéndice A

La distribución Normal Multivariada

En este apéndice se muestra la distribución normal multivariada y se obtienen los estimadores de máxima verosimilitud de sus parámetros. Los teoremas auxiliares utilizados sobre diferenciación matricial pueden ser consultados con mayor detalle en Schott (1997).

Definición (Distribución normal multivariada). *Se dice que el vector $X = (X_1, X_2, \dots, X_n)$ posee una distribución normal multivariada si su función de distribución de probabilidades es*

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right], \quad (\text{A.1})$$

y se denota por

$$X \sim N_n(\mu, \Sigma).$$

La densidad normal multivariada es una generalización de la densidad normal univariada para $p > 2$ dimensiones.

Teorema A.0.1. *Si A es un escalar, entonces $\text{tr}(A) = A$. En particular, si $x_{n \times 1}$ es un vector y $A_{n \times n}$ una matriz, entonces $x^t A x$ es un escalar y*

$$x^t A x = \text{tr}(x^t A x) = \text{tr}(x x^t A) = A x x^t.$$

Teorema A.0.2. *Sean $A_{p \times p}$ matriz simétrica y no singular, $B_{p \times p}$ matriz y $x_{p \times 1}$ un vector. Entonces:*

a)

$$\frac{\partial}{\partial A} \text{tr}(AB) = B + B^t - \text{diag}(B).$$

b)

$$\frac{\partial}{\partial x} (x^t Ax) = 2Ax.$$

c)

$$\frac{\partial}{\partial A} \ln |A| = 2A^{-1} - \text{diag}(A^{-1}).$$

Teorema A.0.3. Sean X_1, X_2, \dots, X_n variables aleatorias normales multivariadas $N_p(\mu, \Sigma)$, y $x = (x_1, x_2, \dots, x_n)$ una muestra observada de las variables aleatorias, entonces los estimadores de máxima verosimilitud para μ y Σ son

$$\hat{\mu} = \bar{x},$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t = \frac{1}{n} W = \frac{n-1}{n} S,$$

donde $W = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t$.

Demostración. La función de verosimilitud asociada está dada por

$$\mathcal{L}(\mu, \Sigma; x) = \frac{1}{(\sqrt{2\pi})^{np} |\Sigma|^{\frac{n}{2}}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^t \Sigma^{-1} (x_i - \mu) \right].$$

Nótese que

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)^t \Sigma^{-1} (x_i - \mu) &= (x_1 - \mu)^t \Sigma^{-1} (x_1 - \mu) + \dots + (x_n - \mu)^t \Sigma^{-1} (x_n - \mu) \\ &= \text{tr}[\Sigma^{-1} (x_1 - \mu)(x_1 - \mu)^t] + \dots + \text{tr}[\Sigma^{-1} (x_n - \mu)(x_n - \mu)^t] \\ &= \text{tr}[\Sigma^{-1} (x_1 - \mu)(x_1 - \mu)^t + \dots + \Sigma^{-1} (x_n - \mu)(x_n - \mu)^t] \\ &= \text{tr} \left[\Sigma^{-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^t \right] \\ &= \text{tr}(\Sigma^{-1} A) \end{aligned}$$

Donde,

$$\begin{aligned}
A &= \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^t \\
&= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)(x_i - \bar{x} + \bar{x} - \mu)^t \\
&= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - \mu)][(x_i - \bar{x}) + (\bar{x} - \mu)]^t \\
&= \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t + \sum_{i=1}^n [(x_i - \bar{x})(x_i - \mu)^t + (\bar{x} - \mu)(x_i - \bar{x})^t] + \sum_{i=1}^n (\bar{x} - \mu)(\bar{x} - \mu)^t \\
&= \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t + (\bar{x} - \mu)(x_i - \bar{x})^t + \sum_{i=1}^n (\bar{x} - \mu)(\bar{x} - \mu)^t
\end{aligned}$$

La última igualdad se da ya que

$$\sum_{i=1}^n [(x_i - \bar{x})(x_i - \mu)^t + (\bar{x} - \mu)(x_i - \bar{x})^t] = 0.$$

Así,

$$\begin{aligned}
A &= \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^t \\
&= \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t + n(\bar{x} - \mu)(\bar{x} - \mu)^t \\
&= W(x) + n(\bar{x} - \mu)(\bar{x} - \mu)^t
\end{aligned}$$

Por lo tanto, se puede escribir la verosimilitud asociada como

$$\mathcal{L}(\mu, \Sigma; x) = |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} [W + n(\bar{x} - \mu)(\bar{x} - \mu)^t] \right\}.$$

De esta manera la log-verosimilitud asociada es

$$\ell(\mu, \Sigma; x) = -\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}(\Sigma^{-1} W) - \frac{n}{2} (\bar{x} - \mu)^t \Sigma^{-1} (\bar{x} - \mu).$$

Para obtener el EMV de μ , se deriva la log-verosimilitud previa respecto a μ , lo cual queda

$$\frac{\partial}{\partial \mu} \ell(\mu, \Sigma; x) = -n \Sigma^{-1} (\mu - \bar{x}).$$

Igualando a cero se obtiene que:

$$\hat{\mu} = \bar{x}. \quad (\text{A.2})$$

Se verifica que

$$\frac{\partial^2}{\partial \mu \partial \mu^t} \ell(\mu, \Sigma; x) = -\frac{n}{2} \frac{\partial}{\partial \mu} [2\Sigma^{-1}(\mu - \bar{x})] = -n\Sigma^{-1} < 0,$$

pues $n > 0$ y Σ definida positiva. Por tanto (A.2) es máximo.

Una vez hecho esto, se procede a calcular la logverosimilitud perfil de Σ ,

$$\begin{aligned} \ell_p(\Sigma) &= \ell(\hat{\mu}(\Sigma), \Sigma; x) \\ &= \ell(\hat{\mu}, \Sigma; x) \\ &= -\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}(\Sigma^{-1}W). \end{aligned}$$

Derivando con respecto a Σ^{-1} obtiene:

$$\frac{\partial}{\partial \Sigma^{-1}} \ell_p(\Sigma) = n\Sigma - \frac{n}{2} \text{diag}(\Sigma) - W + \frac{1}{2} \text{diag}(W)$$

Igualando a cero y despejando Σ :

$$\hat{\Sigma} = \frac{1}{n}W.$$

Por lo tanto, se obtiene que los EMV son:

$$\begin{aligned} \hat{\mu} &= \bar{x}, \\ \hat{\Sigma} &= \frac{1}{n}W. \end{aligned}$$

Apéndice B

Gráficas cuantil-cuantil de mediciones en mayo de 2011 y mayo de 2012

En este apéndice se incluyen las gráficas Q-Q de las 17 especies consideradas en esta tesis, donde se verifica que las variables X_{*j} y X_{*j}^* para $i = 1, 2, \dots, j$, siguen una distribución aproximadamente normal.

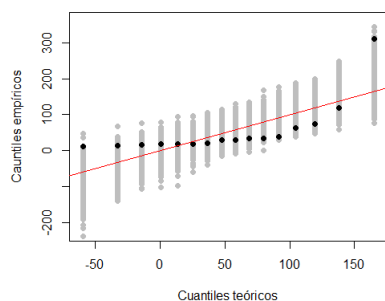


Figura B.1: *Colubrina heteroneura* 2011

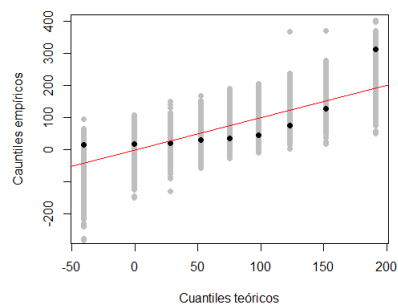


Figura B.2: *Colubrina heteroneura* 2012

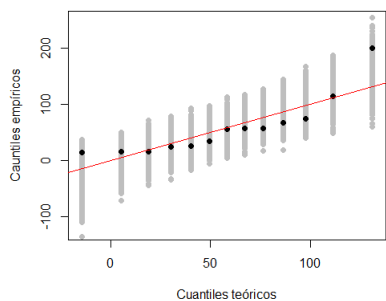


Figura B.3: *Cordia alliodora* 2011

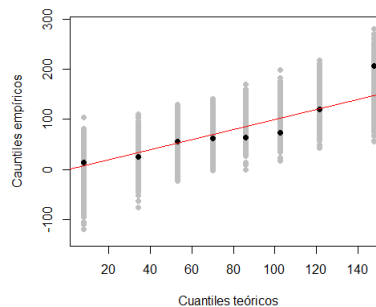


Figura B.4: *Cordia alliodora* 2012

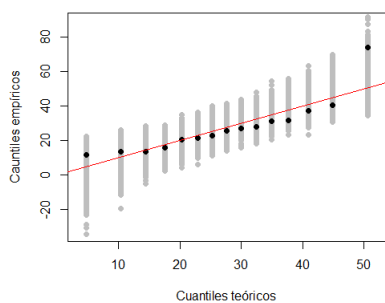


Figura B.5: *Adelia oaxacana* 2011

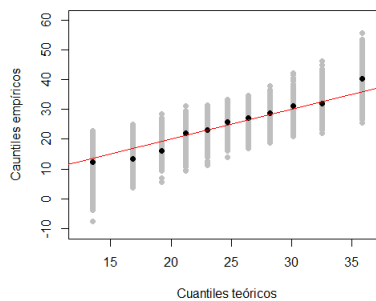


Figura B.6: *Adelia oaxacana* 2012

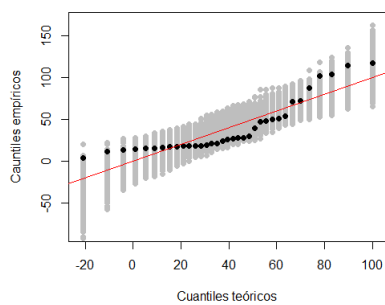


Figura B.7: *Exostema caribaeum* 2011

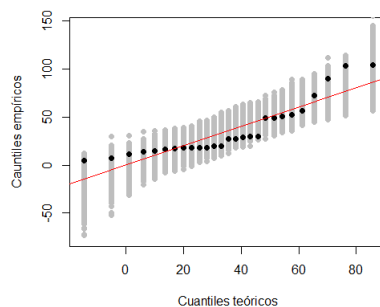
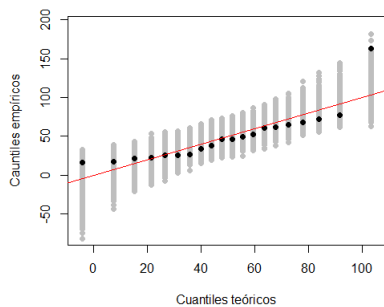
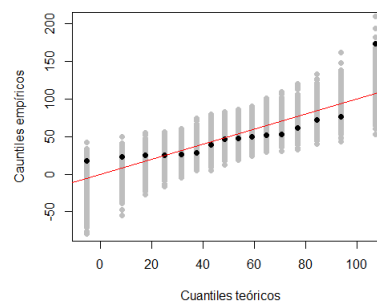
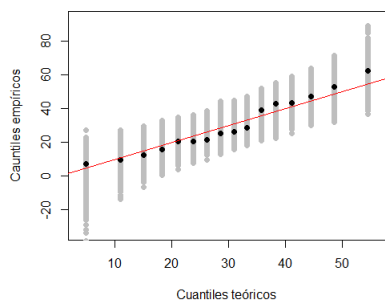
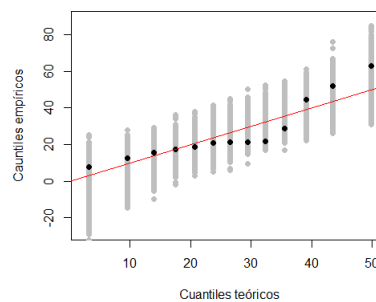
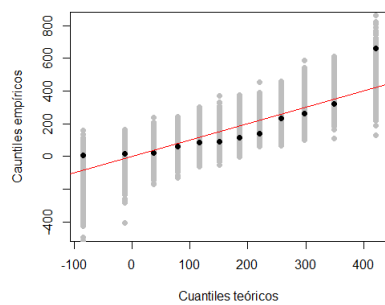
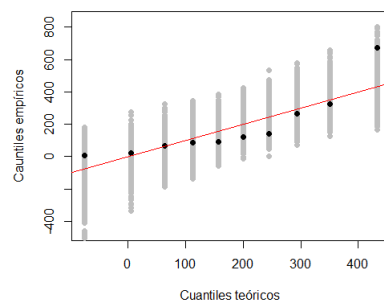


Figura B.8: *Exostema caribaeum* 2012

Figura B.9: *Piptadenia constricta* 2011Figura B.10: *Piptadenia constricta* 2012Figura B.11: *Bauhinia unguolata* 2011Figura B.12: *Bauhinia unguolata* 2012Figura B.13: *Caesalpinia eriostachys* 2011Figura B.14: *Caesalpinia eriostachys* 2012

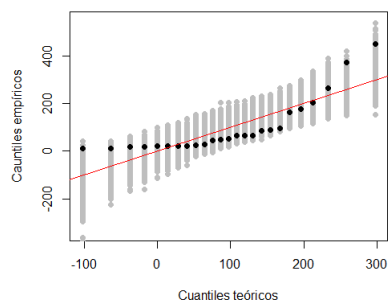


Figura B.15: *Comocladia engleriana* 2011

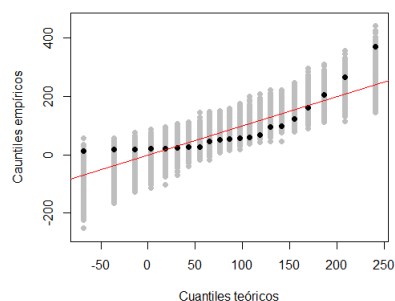


Figura B.16: *Comocladia engleriana* 2012

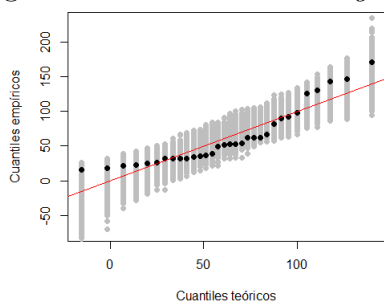


Figura B.17: *Plumeria rubra* 2011

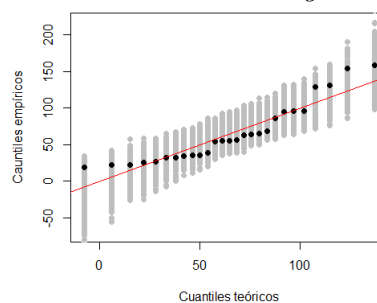


Figura B.18: *Plumeria rubra* 2012

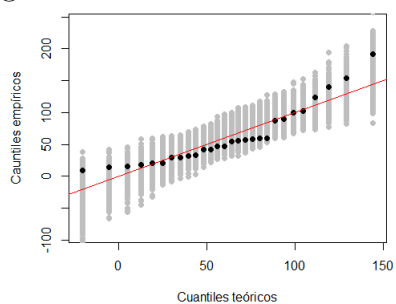


Figura B.19: *Guapira macrocarpa* 2011

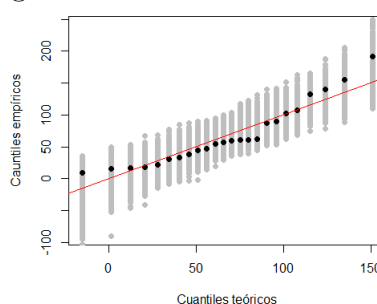
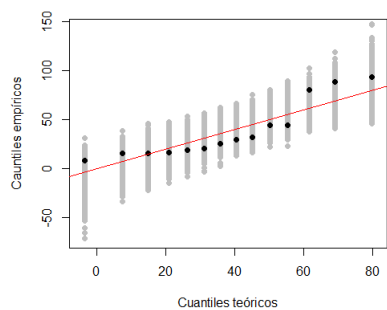
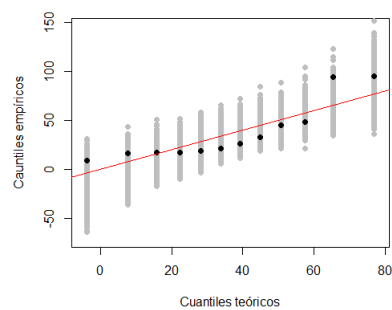
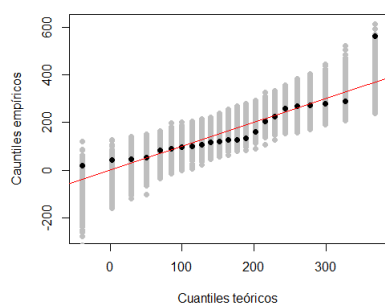
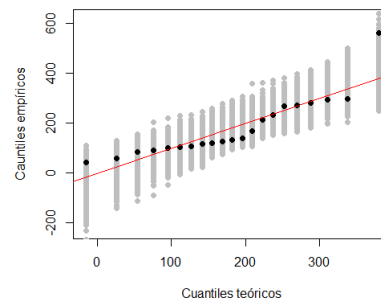
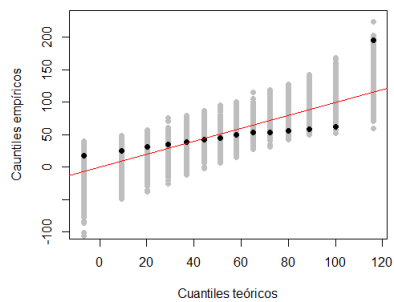
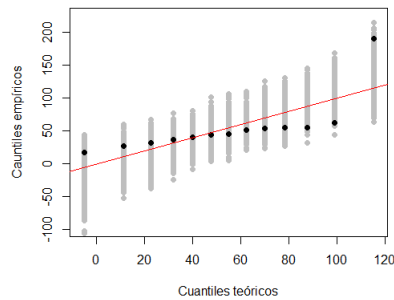


Figura B.20: *Guapira macrocarpa* 2012

Figura B.21: *Guettarda elliptica* 2011Figura B.22: *Guettarda elliptica* 2012Figura B.23: *Bursera instabilis* 2011Figura B.24: *Bursera instabilis* 2012Figura B.25: *Helietta lottiae* 2011Figura B.26: *Helietta lottiae* 2012

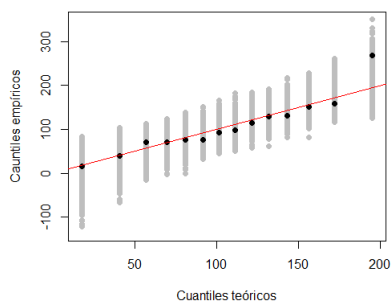


Figura B.27: *Tabebuia impetiginosa* 2011

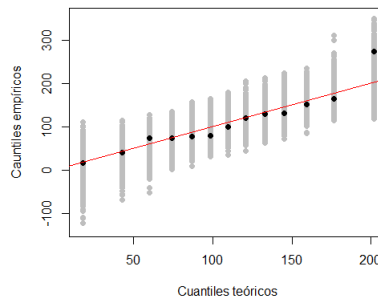


Figura B.28: *Tabebuia impetiginosa* 2012

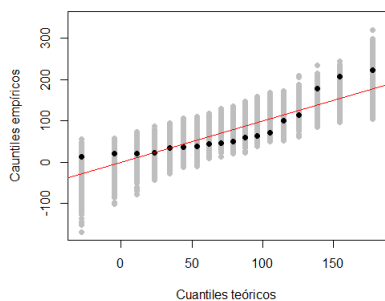


Figura B.29: *Lonchocarpus eriocarinalis* 2011

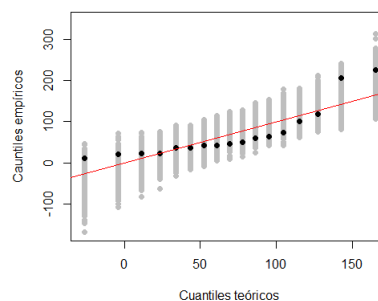


Figura B.30: *Lonchocarpus eriocarinalis* 2012

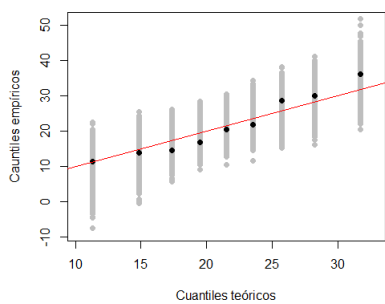


Figura B.31: *Samyda mexicana* 2011

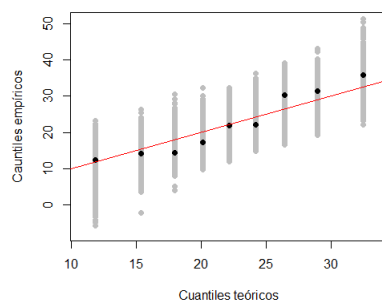


Figura B.32: *Samyda mexicana* 2012

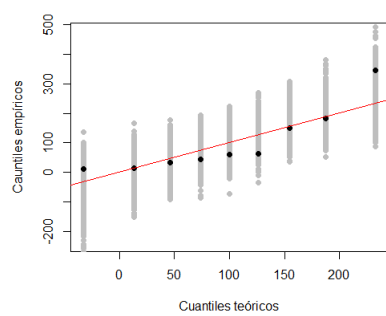


Figura B.33: *Tabebuia chrysantha* 2011

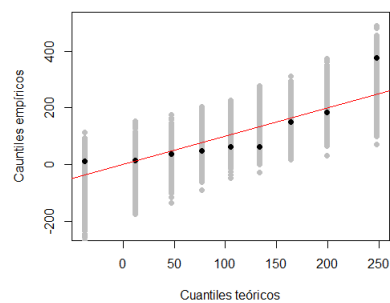


Figura B.34: *Tabebuia chrysantha* 2012

Apéndice C

Resultados de la prueba de Mardia

Los resultados obtenidos al aplicar la prueba de Mardia introducida en la Sección 2.4.1, para verificar el supuesto de normalidad bivariada en cada una de las 17 especies consideradas son los siguientes.

N.P.	Especie	n_j	$\hat{\gamma}_{1,p}$	p-valor sesgo	$\hat{\gamma}_{2,p}$	p-valor curtosis	Normal bivariada razonable
1	<i>Colubrina heteroneura</i>	16	16.81	0.0021	9.92	0.4713	NO
2	<i>Cordia alliodora</i>	13	3.78	0.4359	6.58	0.6159	SI
3	<i>Bauhinia unguolata</i>	16	12.84	0.0121	10.14	0.3342	NO
4	<i>Comocladia engleriana</i>	25	10.71	0.0000	16.59	0.0000	NO
5	<i>Adelia oaxacana</i>	15	4.38	0.3573	6.62	0.5681	SI
6	<i>Exostema caribaeum</i>	35	1.97	0.0643	7.36	0.6774	SI
7	<i>Piptadenia constricta</i>	20	6.21	0.0024	12.07	0.0418	NO
8	<i>Guapira macrocarpa</i>	28	4.96	0.0005	10.32	0.1554	NO
9	<i>Guettarda elliptica</i>	14	18.86	0.0008	11.29	0.1544	NO
10	<i>Plumeria rubra</i>	32	5.01	0.0002	10.83	0.0062	NO
11	<i>Caesalpinia eriostachys</i>	12	14.75	0.0052	9.16	0.6471	NO
12	<i>Bursera instabilis</i>	23	9.75	0.0000	16.54	0.0000	NO
13	<i>Helietta lottiae</i>	14	25.75	0.0000	12.26	0.0546	NO
14	<i>Tabebuia impetiginosa</i>	14	5.66	0.2261	7.45	0.8042	SI
15	<i>Lonchocarpus eriocarinalis</i>	18	10.99	0.0267	9.13	0.5618	NO
16	<i>Samyda mexicana</i>	9	7.79	0.0996	7.44	0.8328	SI
17	<i>Tabebuia chrysantha</i>	9	17.49	0.0016	9.71	0.5210	NO

Tabla C.1: Resultados obtenidos en la prueba de Mardia aplicada a las 17 especies consideradas.

Referencias bibliográficas

- [1] Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*. Boston: Chapman and Hall.
- [2] Fisher, R. A. (1991). *The Design of Experiments*. Oxford: Oxford University Press. 8a Edición.
- [3] García-Oliva, F., Camou, A. y Maass, J. M. (2002). El Clima de la Región Central de la Costa del Pacífico Mexicano. En: Noguera, F.A., Vega, J.H. R., García, A.N. A y Quesada, M. A (Eds.). *Historia Natural de Chamela*. Instituto de Biología, UNAM. México. pp:525-542.
- [4] Hogg, R. V. y Craig, A. T. (1978). *Introduction to Mathematical Statistics*. Nueva York: Macmillan Publishing Co. Inc.
- [5] Johnson, R. A. y Bhattacharyya, G. K. (2010). *Statistics, Principles and Methods*. Nueva York: John Wiley and Sons, Inc. 6a Edición.
- [6] Kalbfleisch, J. G. (1985). *Probability and Statistical Inference*. Nueva York: Springer-Verlag.
- [7] Mardia, K. V. (1974). Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies. *Sankhya*, Serie B, V.36, No. 2, pp. 115-128.

- [8] Marron, J. S., Campos, F. H., Smith, F. D. y Samorodnitsky, G. (2004). Variable Heavy Tails in Internet Traffic. V. 18, pp. 27. Performance Evaluation, V.58, pp. 261-284.
- [9] Melo, O. A. C. y Vargas, R. R. (2003). *Evaluación ecológica y silvicultural de ecosistemas boscosos*. Universidad de Tolima.
- [10] Ortega, G. U. (2013). *Modelos estadísticos para eventos de lluvia y para sus máximos anuales en Chamela y Maiquetía*. Tesis de Licenciatura en Matemáticas. Universidad de Guanajuato.
- [11] Pawitan, Y. (2001). *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. Oxford: Oxford University Press.
- [12] Schott, J. R. (1997). *Matrix Analysis for Statistics*. Nueva York: John Wiley and Sons.
- [13] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Nueva York: John Wiley and Sons.
- [14] Sprott, D. A. (2000). *Statistical Inference in Science*. Nueva York: Springer-Verlag.
- [15] Sprott, D. A. (2001). The Estimation of Ratios From Paired Data. Nueva York: Springer-Verlag. En: Ahmed, S. E. y Reid, N (Eds.). *Empirical Bayes and Likelihood Inference*. Lecture Notes in Statistics. V. 148, pp. 141-159.
- [16] Triola, M. F. (2009). *Estadística*. Addison-Wiley.