



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

**Modelado estocástico de dependencia  
mediante cópulas para variables  
morfométricas en la región sureste del  
Campo Volcánico Michoacán Guanajuato**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

**ACTUARIO**

P R E S E N T A:

NOMBRE DEL ALUMNO:

Aguilar Tomasini Joel

TUTOR:

Ana Teresa Mendoza Rosas



Ciudad Universitaria, Cd. Mx., 2023



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## ÍNDICE

RESUMEN .....	1
INTRODUCCIÓN .....	2
1. FORMULACIÓN DEL PROBLEMA.....	4
1.1 Objetivos .....	4
1.2 Justificación.....	4
1.3 Antecedentes.....	7
i. Antecedentes geológicos.....	7
ii. Antecedentes de la Teoría de Cópulas .....	8
1.4 Marco tectónico y Geológico.....	9
2. FUNDAMENTOS DE VULCANOLOGÍA .....	12
3. MARCO TEÓRICO .....	18
3.1 Inferencia Estadística.....	18
3.2 Teoría de Cópulas .....	20
4. METODOLOGÍA .....	27
5. APLICACIÓN.....	30
6. DESARROLLO Y RESULTADOS .....	34
6.1 Conos de Escoria .....	34
6.1.1- Edad vs Índice de aspecto .....	34
6.1.2- Edad vs Localización .....	43
6.1.2.1- Edad vs Longitud .....	43
6.1.2.2- Edad vs Latitud .....	47
6.2 Spatter y Domos.....	51
6.2.1- Edad vs Localización .....	51
6.2.1.1- Edad vs Longitud .....	51
6.2.1.2- Edad vs Latitud .....	56
6.2.2- Edad vs Altura.....	61
6.2.3- Edad vs Diámetro basal.....	66
6.3 Conos y Spatter .....	71
6.3.1 Localización vs Elongación .....	71
6.3.1.1 Conos .....	71
6.3.1.2 Spatter .....	78
7. RELACIONES DE DEPENDENCIA .....	84
Despeje .....	83
Regresión .....	87
8. CONCLUSIÓN Y DISCUSIÓN.....	96
9. ANEXO: GLOSARIO PROBABILIDAD .....	101
10. LISTA DE FIGURAS .....	110
11. BIBLIOGRAFÍA.....	115

*“Cuando tratamos de entender la naturaleza, consiste en imaginar que los dioses están jugando a una gran partida de ajedrez, y nosotros no conocemos las reglas del juego, pero se nos permite mirar el tablero (y tratamos de imaginar la reglas a partir de esto)”<sup>1</sup>*

---

<sup>1</sup> Feynman Richard, 1981, “Las reglas del juego”. Parte de la antología “El placer de descubrir”

## **Agradecimientos**

A Ana Teresa Mendoza por la paciencia y guía en la elección y elaboración de este proyecto.

A mis profesores que fueron inspiradores y apasionados por la enseñanza.

A mi hermana, madre, padre y tíos por el apoyo incondicional e inspiración.

A mis hermanos de otra sangre: Talib Zamudio, Andrés Raygoza, Diego Altamirano, Diego Garda y Oswaldo Carranza. Por acompañarme en todos estos años y haber crecido juntos.

A aquellas amistades que la facultad me regaló: Silhi Bernabé, Israel Gómez, Alejandro Bautista, Mariana Escobar, Leonardo Ríos, Sari Salgado, Miguel Gómez, David Vázquez, Alfonso Rojas y Miguel Betanzo.

A Mariana Ibarra A. por introducirme al mundo de la lectura y revolucionar mi forma de ver.

*A la vida, porque tengo grandes amigos y como amigos tengo a grandes personas.*

En memoria de Ana Cristina Galindo Lugo y Víctor Pineda Cabello.

## Resumen

La simulación de propiedades geográficas y morfológicas de los volcanes resulta de un particular interés para el estudio de los eventos magmáticos a través del tiempo. A causa de la incertidumbre presente en estos eventos, es conveniente apoyarse en la teoría de la probabilidad y estadística. Generando, por medio del análisis probabilístico, modelos de dependencia entre variables.

Estos modelos, a través de las relaciones entre variables primarias y variables secundarias o de apoyo, esbozan variables de interés basándose en leyes físicas, relaciones empíricas o métodos de regresión. Sin embargo, los métodos de regresión no siempre representan adecuadamente la variabilidad de los datos y en su mayoría se limitan a suponer dependencias lineales. Las cópulas son un modelo que no tiene dicha limitante.

El presente estudio busca conocer la dependencia entre variables de morfometría, además de la edad relativa de las estructuras volcánicas, mediante el uso de la teoría de cópulas para medir la correlación no lineal entre las variables disponibles en tres tipos de estructuras volcánicas en particular: Spatter, Domos y Conos. Lo que se validará con diferentes medidas de coeficientes de correlación.

Este método se aplica a datos morfológicos y fechamientos disponibles de la zona sureste del Campo Volcánico Michoacán Guanajuato (CVGM). Dado su número y variedad de estructuras volcánicas, así como su extenso registro histórico y geológico, respaldado con una amplia documentación, se procede a examinar los diversos casos. Los resultados se analizan y justifican la inclusión de las cópulas en los análisis vulcanológicos de campos volcánicos monogenéticos

En la teoría de las cópulas se cuentan con modelos estocásticos que permiten identificar y simular la dependencia entre variables. Determinar las dependencias entre variables aleatorias no se pueden inferir únicamente a partir de las distribuciones de las variables individuales, puesto que requiere conocer la distribución conjunta de las variables. Para resolverlo, se ajustan modelos de distribución bivariados y se proponen modelos de dependencia, basados en cópulas, que son modelos probabilísticos que han tenido aplicaciones en diferentes ámbitos, como finanzas, industria petrolera, economía, entre otros.

En este caso, inicialmente se realizará un análisis univariado obteniendo estadísticas generales y ajustando una función de densidad a cada variable para usarse en el estudio bivariado paramétrico (Cópula de Frank). Posteriormente se estimarán las mejores cópulas no-paramétricas, Kernel Smoothing y Bernstein. Finalmente, se realizarán las simulaciones de las variables y se compararán con los datos originales, por medio de coeficientes de correlación de Kendall, Spearman y Pearson.

Se establecieron relaciones de dependencia no lineales entre diversas variables morfométricas de los volcanes estudiados. Entre los resultados, se encontró una relación negativa entre el índice de aspecto y la edad para los conos volcánicos. Asimismo, se determinó una dependencia negativa entre el diámetro basal y la edad para los domos. Otras variables como la elongación y la longitud presentaron relación positiva para los spatter. Estos análisis no pudieron realizarse mediante regresión debido a que los datos no cumplían con los supuestos del modelo. Esto demuestra la ventaja de aplicar cópulas en este tipo de estudios, al permitir analizar relaciones de dependencia cuando los datos no satisfacen los criterios de regresión clásica.

## Introducción

Los campos volcánicos monogenéticos, son zonas con numerosas estructuras volcánicas donde la mayoría han tenido una única erupción y tienen diversos orígenes; son de interés para los estudios estadísticos dada su gran cantidad de estructuras, lo que permite hacer simulaciones de sus parámetros morfológicos e identificar relaciones entre diversos factores para comprender mejor los procesos volcánicos. Además de ser importantes por la información que brindan sobre la evolución temporal y espacial de la actividad volcánica en la zona, así como comprender patrones.

Por otro lado, los campos monogenéticos presentan grandes desafíos, así como mapearlos de manera completa (debido a sus múltiples centros eruptivos individuales dispersos en un área), ya que los costes monetarios y humanos, que requieren de muestreo, datación y análisis de cada una de los cientos de estructuras presentes. Además, la dificultad física de acceso a ciertas regiones con accidentes geográficos lo que hace la investigación integral de estos sistemas un proceso largo y complejo.

Entre los campos volcánicos monogenéticos, el Campo Volcánico Michoacán-Guanajuato (CVMG) en México es una región compleja y de significativa actividad volcánica. Alberga miles de estructuras volcánicas que presentan gran diversidad petrológica. Sin embargo, solamente una fracción mínima ha sido datada y estudiada en detalle. Esto representa un desafío en fin de comprender la evolución y dinámica del campo.

Debido a ello, la importancia de este estudio radica en explorar formas de simular los parámetros de las estructuras volcánicas de forma confiable y por medio de variables primarias obtenibles por otros métodos, como el satelital, que nos permitirá a futuro generar estudios y modelos vulcanológicos eficaces y a menor coste.

Para modelar adecuadamente y obtener la dependencia entre variables se requiere conocer la distribución conjunta de probabilidad de las variables. Esta información no puede obtenerse directamente de los datos originales, por lo que se ajustan modelos de distribución a las variables y se conjuntan utilizando cópulas. Esto permite modelar la distribución conjunta (una cópula) que muestra la correlación subyacente entre las variables. El uso de esta técnica resulta útil para modelar dependencias en análisis estadísticos modernos, debido a que otros métodos no consideran la dependencia entre variables o hacen suposiciones restrictivas. Tal como se verá en el marco teórico.

En síntesis, a pesar de la existencia de varios estudios del CVMG, la gran cantidad de estructuras volcánicas sin fechamientos representa un desafío significativo para la comunidad vulcanológica. El uso de modelos estadísticos como las cópulas podría ayudar a llenar parte de este vacío en los datos, proporcionando información adicional sobre las relaciones morfométricas presentes en el campo volcánico.

Este trabajo se dividirá en 8 capítulos dependiendo de lo abordado y el área del conocimiento al que pertenece.

Los contenidos que se abordarán en cada capítulo:

En el capítulo 1 se expone la justificaciones y objetivos del trabajo. Así como un contexto histórico sobre el área de estudio y de la teoría de Cópulas.

En el capítulo 2 se dará un contexto sobre los términos vulcanológicos y variables analizadas en el presente trabajo.

En el capítulo 3 se abordarán las definiciones de carácter más avanzado, abordando con mayor profundidad aquellos conceptos que constituyen el sustento teórico del trabajo como la teoría y construcción de Cópulas, y las medidas de inferencia estadística.

En el capítulo 4 se abordará la metodología.

En el capítulo 5 se describirá la metodología propuesta y forma de limpieza de los datos.

En el capítulo 6 se hará el desarrollo, comparación y discusión de los resultados.

En el capítulo 7 se expondrá el desarrollo matemático para buscar pasar de la cópula obtenida a la expresión de una variable en función de la otra en aquellos resultados que indiquen dependencia.

En el capítulo 8 se dará un resumen de los resultados y discutirán para obtener las conclusiones generales del trabajo, así como las limitaciones identificadas y las posibles líneas de investigación futuras.

En el Anexo se expondrán las definiciones básicas de los conceptos clave relacionados con probabilidad y estadística necesarias para este trabajo. Se realizará una introducción para aquellos lectores menos familiarizados con la temática, además de hablar de la Implementación computacional.

## Capítulo 1- Formulación del Problema

### 1.1 Objetivos

#### a. Generales

Identificar la mejor función bivariada por medio del uso de cópulas que caracterice la distribución entre las variables de cada uno de tres tipos de cuerpos volcánicos de la región sureste del Campo Volcánico Michoacán-Guanajuato. Además de analizar la posible dependencia entre estos factores por medio de medidas de correlación.

#### b. Particulares

Encontrar la cópula que imite la distribución conjunta de la edad con las siguientes variables:

- i. La altura del cono entre el diámetro basal (llamado H/W o Índice de Aspecto)
- ii. La localización de Conos, Domos y Spatter
- iii. El diámetro de Domos y Spatter
- iv. La altura de Domos y Spatter

Además de la cópula para la distribución conjunta de:

- v. La elongación contra la Localización para Conos y Spatter

De la zona suroriental del campo monogenético Michoacán-Guanajuato (CVMG). A fin de buscar una posible relación de dependencia entre las variables por medio de análisis estadísticos de medidas de correlación aplicada a la función (Cópula) encontrada.

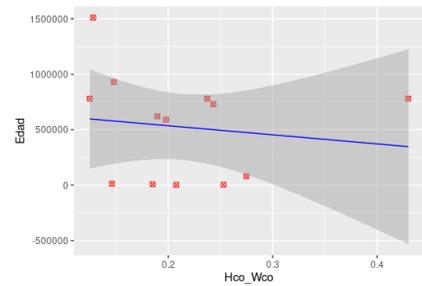
### 1.2 Justificación

A causa de la incertidumbre inherente de las mediciones o a la falta de un conjunto representativo de datos cuantitativos de las estructuras volcánicas, es aconsejable utilizar métodos de estimación. Los cuales son variados, habiendo no-probabilistas como la regresión <sup>1</sup> o probabilistas como los modelos estocásticos de dependencia. Esto permite modelar la variable de interés a través de su relación con otras variables, denominadas variables secundarias o de apoyo.

---

<sup>2</sup> La gráfica de una función de probabilidad se lee como “en el evento estudiado, el valor  $x$  sucede con probabilidad  $y$  ( $y = P(x)$ )”. En cambio, la gráfica de una regresión no asocia probabilidades sino dos valores de variables diferentes, leyéndose como “al valor  $x$  le corresponde al valor  $y$ ” de modo que de querer añadir una definición que incluya la probabilidad cambiaría a “al valor  $x$  le corresponde el valor  $y$  con probabilidad  $1$ ” que termina siendo la definición de un evento no-probabilista.

Los modelos tradicionales de dependencia entre variables primarias y secundarias se basan en leyes físicas, relaciones funcionales empíricas o principalmente en técnicas de regresión. Es decir, modelos deterministas, que no permiten representar adecuadamente la variabilidad de los datos por la incertidumbre inherente que no está siendo considerada y que constituye una parte fundamental de la información.<sup>2</sup> Adicionalmente, el enfoque de regresión a menudo se limita a suponer dependencias lineales entre las variables (Figura 1.1), relaciones que en la práctica pueden ser poco frecuentes o limitadas a ciertos casos particulares.



**Figura 1.1- Ejemplo de Regresión Lineal.** Edad contra Índice de aspecto. Imagen de autoría propia. (Hco\_Wco es la Altura entre el Diámetro basal)

Existe una gran variedad de métodos de simulación estocásticos. Entre los cuales resalta el uso de cópulas por su utilidad en análisis estadísticos modernos. Otros modelos no tienen en cuenta la dependencia entre las variables o parten de hipótesis demasiado restrictivas sobre el tipo de dependencias subyacentes<sup>3</sup>, mientras que las cópulas sí tienen en cuenta esta dependencia y son más flexibles en sus hipótesis, como se verá en el Marco teórico.

Para establecer estas dependencias se requiere del conocimiento de la función de distribución de probabilidad conjunta de las variables y una manera eficiente de obtener estas distribuciones es mediante el uso de Cópulas<sup>4</sup>.

*“En la estadística moderna, la manera sistemática y flexible de modelar dependencias complejas en distribuciones de probabilidad multivariadas es mediante cópulas. Las cópulas son funciones que permiten atrapar la relación de dependencia entre variables aleatorias de manera desacoplada de su comportamiento marginal individual.”<sup>5</sup>*

De esta forma, las cópulas ayudan a la estimación de una función de distribución bivariada, y respectiva dependencia entre variables, sin importar que no se conozca la función de distribución individual de cada variable.

Uno de los parámetros -al que aplicaremos esta metodología es a la división entre la altura y el diámetro basal de un Cono, conocido como el índice de aspecto o juventud o simplemente H/W (por sus siglas en inglés); este índice fue propuesto en el estudio de Wood en 1980 donde lo usa como forma auxiliar de medir edades.

*“La relación altura del cono/diámetro basal se ha utilizado para definir estados de degradación, [...] El clima ha sido señalado como uno de los factores más importantes para poder realizar este tipo de comparaciones, por lo que esta relación sólo puede aplicarse a conos cineríticos que se hayan emplazado en regiones climáticas similares”<sup>6</sup>*

Es decir, se basa en la idea de que una estructura volcánica Monogenética Cónica joven o reciente posee cierto diámetro en la base y una altura dada, similar a un triángulo (Cono en dos dimensiones). Y, conforme pasa el tiempo y, debido a la erosión por factores climáticos como la lluvia o el aire, la altura es lo que se ve más afectado, reduciendo la magnitud del índice. De ahí, propone medir la edad de un Cono por medio de sus variables morfológicas (de forma), a esto se le conoce como Morfocronometría, e incorpora el Índice de

Aspecto dentro de sus parámetros para llevar a cabo la descripción de una estructura volcánica sobre bases cuantitativas.

La presente investigación fundamentará su análisis a variables morfocronométricas (referente a la edad relativa o absoluta de las geoformas) y morfométricas (análisis cuantitativo de la forma, un concepto que abarca el tamaño y la forma) aplicado a un campo volcánico monogenético, por las ventajas en cantidad de información que poseen. Los campos volcánicos son conjuntos de volcanes en un área geográfica específica, y el más grande de México y el mundo (Valentine & Connor, 2015)<sup>7</sup>, que contiene alrededor de  $1,200 \pm 200$  estructuras volcánicas (Hasenaka y Carmichael 1985a, Cebriá et al. 2011a. Pérez-López et al. 2011)<sup>8</sup>, es el CVMG. Esta cifra de estructuras supera en casi seis veces al siguiente campo mexicano en tamaño: el campo volcánico Chichinautzin, que cuenta con 221 estructuras volcánicas<sup>9</sup>.

Una de las razones de importancia del estudio del CVMG es su reciente actividad, siendo en este campo que nació el volcán más joven de todo México: el Parícutín, un volcán monogenético basáltico con erupción hawaiana y estromboliana, así como flujos de lava Aa's.

*“Su nacimiento fue de suma importancia para la ciencia a nivel mundial, ya que es uno de los pocos volcanes en el mundo cuyo nacimiento, desarrollo y fin se ha observado y estudiado a detalle”<sup>10</sup>*

El nacimiento del Parícutín inició con temblores de tierra y posterior salida de magma en un campo de cultivo a través de una falla. En un mes alcanzó una altura de 130 metros, dicha erupción acabó con el pueblo de San Juan Parícutiro y su actividad duró 9 años (única erupción registrada de un volcán monogenético con una duración tan extensa) de 1943 a 1952, teniendo una altura final de 424 metros. Es el único volcán “con dueño” por haber nacido en el campo de cultivo perteneciente al Señor Dionisio Pulido, además del primero en el mundo que se registró su nacimiento y final. Esto ya nos induce la importancia de este campo a nivel nacional e internacional.

Este campo se encuentra dentro del Cinturón Volcánico Transmexicano (CVTM), donde en los últimos años ha habido varias erupciones importantes que han impactado comunidades aledañas, como la erupción del Colima en 2015, el Popocatepetl en 2017 y el Fuego de Colima en 2018-2019. Estas erupciones han forzado evacuaciones, afectado cultivos y cosechas, así como interrumpido vías de comunicación, lo cual genera costes y afecta la economía local y nacional. En nuestro caso particular, el CVMG se encuentra principalmente en un estado agrícola como lo es Michoacán donde se genera el 45% de los aguacates consumidos en México y el 15% de la producción agrícola total mexicana, por lo que el monitoreo y entendimiento de su actividad volcánica conlleva un peso sustancial.

Por casos como éstos resulta importante investigar a fondo las características del campo, especialmente desde la perspectiva de la probabilidad, puesto que resultará en menores costos monetarios que el intentar analizar físicamente cada uno de los conos del campo o datarlos con métodos químicos. En el caso particular de variables morfométricas se pueden obtener con mapas generados de forma satelital, de forma que, de haber dependencia

entre la edad y la forma, podríamos inducir la edad de una estructura sin necesidad de datarlo químicamente. Obteniendo por medio de estos modelos información que nos ayudará a prevenir posible actividad futura y comprender las dinámicas del campo.

A fin de que los estudios vulcanológicos sean eficaces en generar modelos predictivos y de prevención de riesgos para mitigar el impacto antropológico que un evento magmático pueda tener, los Campos Monogenéticos presentan una mina de información para comprender los procesos magmáticos de una zona a través del tiempo y mejorar dichos modelos. Sin embargo, es la misma cantidad de información en estos campos la que presenta retos cómo que para fechar por métodos físico-químicos la gran cantidad de cuerpos o estructuras volcánicas, involucraría gran costo en términos de capital monetario y humano, es por esto, que encontrar una posible relación entre los distintos parámetros morfológicos en términos de dependencia presumiblemente nos ayudará a deducir características de los volcanes del campo monogenético, por medio de variables obtenibles con otros métodos.

### 1.3 Antecedentes

#### i. Antecedentes geológicos

En México, un país con una gran actividad sísmica e histórica actividad volcánica, la vulcanología es una materia de interés científico e importancia pública. Ya que es el sexto país con más volcanes activos en el mundo<sup>11</sup>, con un conteo en torno a 46 zonas volcánicas. La mayoría ubicados en el Cinturón Volcánico Transversal (o Transmexicano, CVTM), un sistema de volcanes que cruza el país de oeste a este, incluyendo los volcanes más conocidos como el Popocatepetl, Citlaltépetl e Iztaccíhuatl.

El CVMG se localiza en la porción central del CVTM, distribuido entre la porción centro-norte del estado de Michoacán, y el sur del estado de Guanajuato, ubicado entre las longitudes 100°40' – 102°50' O y latitudes 18°50' – 20°50' N<sup>12</sup>.

El primer informe geológico que incluyó un conteo de estructuras sobre el Campo Volcánico Michoacán-Guanajuato fue realizado por *Hasenaka* en 1985, junto con Carmichael, en el que se realizaron estudios morfométricos y de petrología. Bajo su criterio la entidad recibió su actual denominación (CVMG), con un primer conteo de 1040 estructuras<sup>13</sup>.

Recientemente trabajos más detallados (Johnson et al. 2008; Ownby et al. 2011; Guillbaud et al. 2012, entre otros) han generado una gran cantidad de datos petrológicos, geoquímicos y edades isotópicas y radiométricas, importantes para comprender las dinámicas del campo. Sin embargo, en cada estudio realizado se encuentran nuevos cuerpos volcánicos y la cantidad sigue en aumento. Demant<sup>14</sup> (1981) y Connor<sup>15</sup> (1987) mencionan que el CVMG alberga más de 1280 edificios volcánicos, con uno de los últimos conteos en 1,400 por Marcial<sup>16</sup> (2018), no obstante, cómo se verá más adelante en este trabajo,

dejando aún áreas considerables sin registros. Asimismo, los volcanes con edades isotópicas y radiométricas son menores a 400, quedando abajo del 30% del conteo total de estructuras volcánicas en el campo.

Para proponer un método de análisis de los datos actuales a fin de modelar los faltantes, nos enfocaremos en la zona sureste del campo, donde las edades fueron obtenidas principalmente de Gómez- Calderón, 2021<sup>17</sup>; además de: Carmichael et al. 2006; Guilbaud et al 2011; Guilbaud et al., 2012; Cardona, 2015; Ruiz, 2017; Maghoub et al.2017. Por métodos químicos como Carbono 14, Thorio-Uranio y Argón-Argón, además de otros por Paleomagnetismo. Dichas edades fueron calibradas con 2s (2- desviación estándar) por Gómez-Calderón. Entre las comparaciones a las que aplicaremos nuestra metodología de análisis de dependencias se encuentra un parámetro llamado el índice de Aspecto, propuesto por Wood, que propone inferir la edad por medio de la morfología de los Conos, sin embargo, se ha visto debatida su utilidad mientras que métodos similares de datación por medio de la morfología han sido usados en algunos estudios modernos como el de Corona y Gómez-Calderón (2018)<sup>18</sup>.

La obtención de edades relativas por medio de la morfología se basa en una suposición de correlación previa existente entre dichas variables, este estudio busca analizar la correlación subyacente entre distintos pares de variables, incluidos los útiles para morfocronometría, mediante el uso de cópulas.

## ii. Antecedentes de la teoría de cópulas

La palabra "cópula" proviene del latín *copula* y significa "unión" o "vínculo". La idea fundamental detrás de las cópulas es proporcionar una forma de describir la estructura de dependencia entre variables aleatorias

Las cópulas inician a mediados del siglo XX, cuando los matemáticos empiezan a investigar cómo modelar dependencia en pares de variables aleatorias. Propuestas por primera vez en 1959 por el matemático polaco Jan Mikusinski en su libro "Operational Calculus"<sup>19</sup> y formalizándose el mismo año con Harry Joe, definiéndola como una función que une las distribuciones marginales de las variables aleatorias a su distribución conjunta. Esta definición proporcionó una forma elegante de estudiar y describir la dependencia sin perder información sobre las distribuciones marginales. Aún en 1959 el estadista de origen polaco Alfred J. Sklar hace su aporte con su artículo "Fonctions de répartition à n dimensions et leurs marges", en el que establece el Teorema de Sklar<sup>20</sup>.

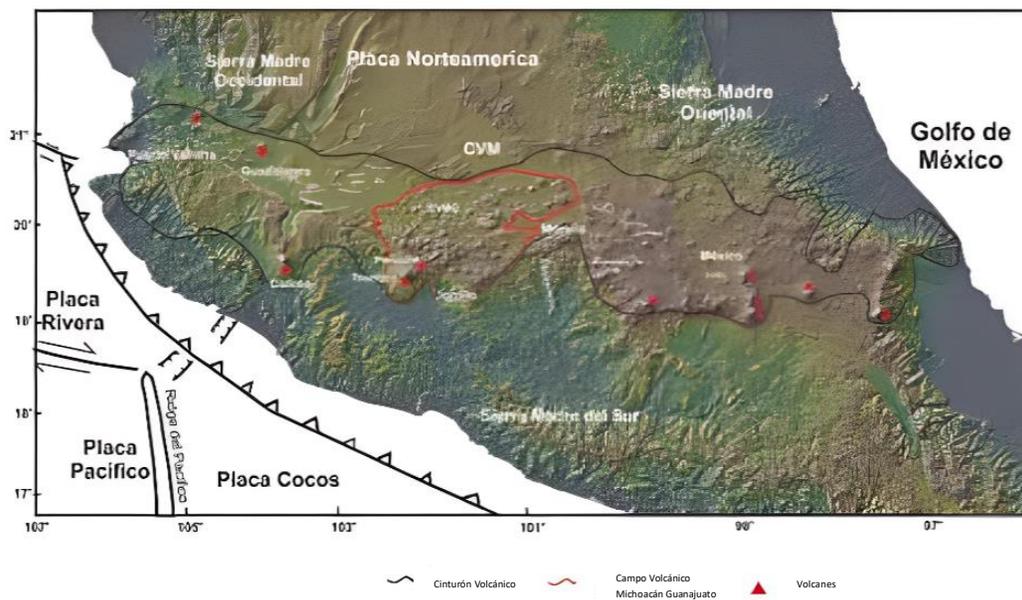
Sin embargo, es hasta 1980 que se usa por primera vez en campos como la economía, finanzas y ciencias actuariales; es entonces cuando cobran su importancia, desarrollando con el tiempo numerosas familias de cópulas, cada una con propiedades y características específicas las cuales se mencionan en el Marco Teórico. (Capítulo 3)

En la actualidad para el estudio de dependencia en variables pertenecientes a las Ciencias Naturales, contamos con estudios previos en términos de la Ingeniería de Petróleos y en estudios de Acuíferos. Mientras que, para la aplicación de estos métodos al estudio vulcanológico, este trabajo resultará pionero.

#### 1.4 Marco tectónico y geológico

El CVTM es un arco volcánico continental originado por uno de los márgenes convergentes más complejos del mundo<sup>21</sup> y que se caracteriza por presentar una amplia variedad composicional desde basanitas, shoshonitas, basaltos y andesitas, hasta riolitas metaluminosas, peralcalinas o peraluminosas, inclusive detectándose algunas riolitas trondhjemíticas.<sup>22</sup> El CVTM y los campos volcánicos asociados como el CVMG son producto de la interacción de las placas de Rivera y Norteamérica.

Debido a que gran cantidad de rocas del CVTM han sido difíciles de categorizar usando los sistemas de clasificación geoquímicos clásicos, Gómez-Tuena y colaboradores (2018) proponen un grupo geoquímico denominado tipo intraplaca. Las rocas tipo intraplaca se pueden encontrar en el frente volcánico de los campos volcánicos de Chichinautzin y Michoacán – Guanajuato (Figura 1.2), en donde sus diferentes edificios volcánicos de tipo monogénico constituyen alineamientos volcánicos en dirección N-S y NNO-SSO, respectivamente.<sup>23</sup>

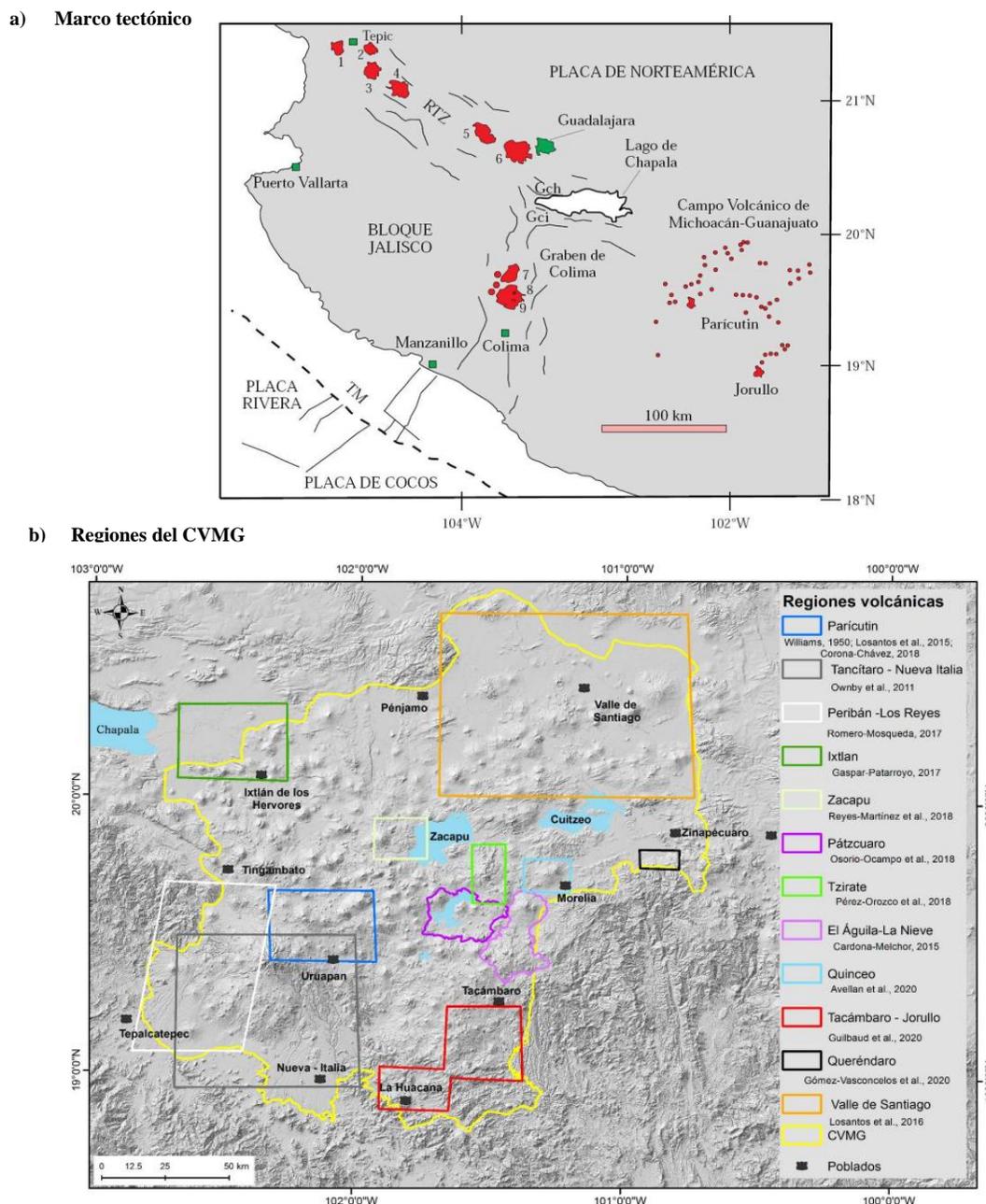


**Figura 1.2- CVTM.** Mapa de elevación del límite del Cinturón Volcánico Mexicano y la zona de estudio está dentro del recuadro rojo. Lo limitado de negro es el CVTM, el límite rojo corresponde al CVMG y los triángulos rojos son los principales volcanes de la faja Transmexicana.

En el mapa de elevación se aprecia el efecto del cambio en el ángulo de subducción que presentan las placas de Rivera y Cocos a lo largo de la costa mexicana, ya que a menor ángulo, mayor es la distancia que se necesita para que la placa subducida alcance profundidades de fusión, generando todo el CVTM a una distancia variante de la costa.

Específicamente la tectónica y geología del CVMG es que el campo presenta una característica estructural conocida como la *Depresión del Bajo Balsas* que es una amplia región de tierras bajas que están situadas entre el CVTM y la Sierra Madre del Sur (Figura 1.2a), ubicado en el centro meridional de México (Figura 1.2). Es una cuenca en dirección Noroeste-Suroeste que se extiende desde el sureste del estado de Jalisco hasta el límite entre Michoacán y Guerrero para desembocar en las costas de Lázaro Cárdenas. Una sección de la zona sureste (suroriental) también es denominada Región Volcánica Graben del Cahulote de Santa Ana-Pedernales (RV-GCP).

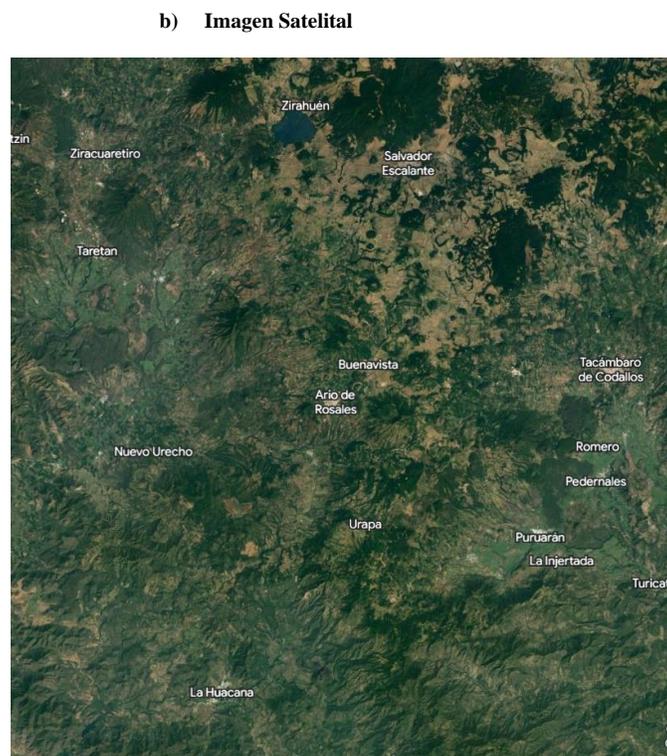
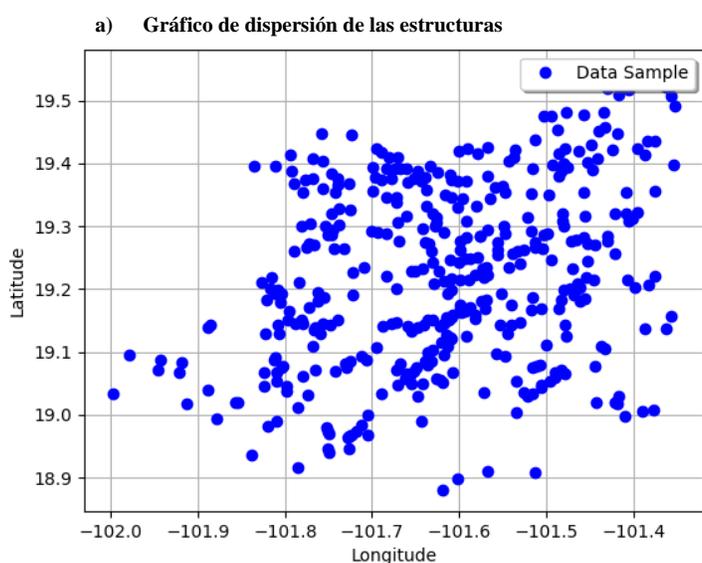
Dentro de la RV-GCP se identificaron 5 dominios morfoestructurales: i) Alto occidental; ii) Escalón occidental; iii) Bajo Central, iv) Alto oriental y v) Bajo oriental. Dichos dominios se pueden considerar como unidades del paisaje que delimitan y condicionan la disposición de los edificios volcánicos.



**Figura 1.3- Modelo digital del terreno del centro de México que muestra:** a) Marco tectónico zona oeste. Depresión del Bajo Balsas. Esquema de la porción oeste de la FVTM (modificado de “Geología e historia eruptiva de algunos de los grandes volcanes activos de México” J.L.Macías,2005; y Luhr et al., 1989). Las estructuras volcánicas son: 1) San Juan, 2) Sangangüey, 3) Tepetitlic, 4) Ceberuco, 5) Tequila, 6) Caldera de la Primavera. Complejo volcánico de Colima: 7) Cántaro, 8) Nevado de Colima y 9) volcán de Colima. Las abreviaciones son: RTZ-rift Tepic-Zacoalco, GCh-graben de Chapala, Gci graben de Citala; b) Regiones volcánicas que se han estudiado dentro del CVMG. Las referencias están insertadas en la imagen. Fuente: Gómez-Cálderón (2021)

El área de estudio es la zona sureste donde se encuentran cientos de estructuras volcánicas del CVMG. Los dominios morfoestructurales que la componen son denominados: Complejo Volcánico La Nieve, Domos postcaldericos Villa Madero, Graben de Tanicua, Graben del Cahulote, Meseta Sta. Clara y Semigraben de Tacambaro.

Se trabajará con un total de 409 estructuras volcánicas, que son los datos de Conos, Domos y Spatter disponibles de nuestra zona (Figura 1.4).



**Figura 1.4- Localización** a) Las estructuras volcánicas con datos disponibles en la zona sureste del campo. Imagen de autoría propia. b) Imagen Satelital del área de estudio

## Capítulo 2- Fundamentos de Vulcanología

En esta sección se dará una explicación de los términos vulcanológicos y variables que emplearemos, empezando por la definición de vulcanología.

*"La vulcanología es el estudio del origen, distribución, estructura, evolución y actividad de los volcanes, y de los procesos relacionados con ellos, incluyendo la génesis y el emplazamiento de las rocas ígneas"<sup>24</sup>*

El enfoque central de la vulcanología es el estudio integral de los volcanes, abarcando su origen, actividad, estructura y distribución. Esta rama especializada dentro de la geología, combina disciplinas como la física, la química y las matemáticas para interpretar tanto la naturaleza como las dinámicas que rigen el comportamiento de los volcanes y los fenómenos vinculados a ellos.

Desarrollada como ciencia a partir de las observaciones del Vesubio de parte del padre de la vulcanología, William Hamilton. El primer observatorio vulcanológico se funda en Italia <sup>25</sup>en 1841, ya desde entonces se considera que nace esta ciencia, sin embargo, cobra importancia hasta principios del siglo XX con el desarrollo de técnicas de análisis químico y tecnologías de medición. Además del impulso en la investigación después de erupciones importantes como la del monte Pelée (1903), St.Helena (1980)<sup>26</sup> o el Pinatubo (1991).<sup>27</sup>

En México, para el manejo y la mitigación de riesgos volcánicos se han usado Instituciones como el Servicio Meteorológico Nacional (SMN), el Centro Nacional de Prevención de Desastres (CENAPRED) y la Universidad Nacional Autónoma de México (UNAM) donde realizan investigaciones y monitoreo constante de la actividad de los volcanes mexicanos, con el fin de prepararse y prevenir desastres. Así, dicha actividad también ha sido objeto de estudios por parte de investigadores nacionales e internacionales. Se han llevado a cabo investigaciones sobre la historia eruptiva de los volcanes en México, la caracterización de los depósitos volcánicos y la evaluación del riesgo volcánico para las poblaciones cercanas a los volcanes. Es por esto que es una ciencia de importancia particular para nuestro país. Por ello, es que la presente investigación tiene como objeto medular la aplicación de modelos estocásticos y estadísticos a información relativa a los volcanes de la zona sureste del CVMG.

Para ello, se define un **volcán** como un punto o superficie terrestre por dónde sale al exterior material fundido (magma). Estos materiales se acumulan en el centro emisor, dando lugar a relieves positivos con morfologías diversas. Se le dice centro emisor

Estas emisiones de magma forman cuerpos o **conos volcánicos** que se refieren al emplazamiento de magma y la estructura de roca ígnea generada por una erupción. Consideraremos que las erupciones tienen pulsos. Y una erupción no termina hasta que el cuerpo volcánico entra en reposo y, estratigráficamente, encontramos suelo generado en la parte superior del depósito de dicha erupción. Por tanto, una erupción puede durar varias horas, días, meses o años. <sup>28</sup>En cada uno de sus pulsos emite material con ciertas diferencias, aunque preservando un tipo de mayor cantidad (que es con lo que clasificaremos los conos).

## Magmas

El **magma**, es una masa silicatada que contiene una fase sólida, una líquida y una gaseosa que se encuentra en el interior de la Tierra y que puede ascender a la superficie a través de fisuras y donde al salir y ser desgasificado se convierte en lava<sup>29</sup>. Este puede almacenarse en reservorio que se conoce como **Cámara magmática**<sup>30</sup> que es un cuerpo subterráneo que alberga magma proveniente del manto (u otra cámara) donde puede permanecer el magma hasta ser excitado por otro pulso entrante, reaccionando, generando calor y presión suficiente para ser expulsados en una erupción; o generando vesiculación (acumulación de sus propios gases en burbujas) hasta que estas burbujas superen la presión litostática del conducto volcánico y generando así una erupción.

El porcentaje de sílice que forme al magma definirá gran parte del estilo eruptivo y posterior depósito, pudiendo así clasificarlos dependiendo de este porcentaje. Dichos tipos de roca están definidos en la serie de reacción de Bowen. Hablaremos de las extrusivas (aquellos magmas que se solidifican en superficie):<sup>31</sup>

### i. Ultramáficos ( $\text{SiO}_2 < 47\%$ )

Provenientes de erupciones poco comunes (dado el calor necesario para esos magmas, datan mayormente del precámbrico) son sumamente líquidas y no forman conos. Con bajo contenido de sílice y baja cristalización. Principalmente Olivinos. Genera Komatitas y Peridotitas.

### ii. Máficos ( $48\% < \text{SiO}_2 < 52\%$ )

Con un contenido mayor de sílice, ya puede generar una pequeña fragmentación (erupciones estrombolianas) pero su efecto principal sigue siendo flujos de lava. Mayormente producido en las dorsales oceánicas. Genera Gabros o Basaltos.

### iii. Intermedio ( $53\% < \text{SiO}_2 < 62\%$ )

Tienen una viscosidad más considerable que los anteriores dada su mayor nivel de cristalización, nucleación (retención de gas y creación de burbujas) genera erupciones más violentas. Genera Dioritas y Andesitas.

### iv. Félsico ( $63\% < \text{SiO}_2$ )

Con un nivel de sílice superior al 70%, que genera una mayor retención de gases ocasiona las erupciones más violentas por su gran retención de gas y cristalización generando una mayor fragmentación y velocidad de ascenso por el conducto. Genera Granodiorita y Dacita.

## Clasificación por frecuencia eruptiva

Estos magmas forman volcanes y en función de la frecuencia de las erupciones los volcanes se clasifican en Monogenéticos (Estructuras volcánicas creadas en una misma erupción y sin otra posterior conocida. Suelen generar volcanes de menor tamaño) y en Poligenéticos (Estructuras creadas en múltiples erupciones. Más común en magmas evolucionados por la presencia de una cámara magmática y la viscosidad de estos).

Así, al conjunto de estos cuerpos en un área geográfica definida en la cual ha habido actividad volcánica recurrente en el tiempo y que se caracteriza por la presencia de rocas volcánicas extrusivas, explosivas e intrusivas relacionadas se le llama **Campo Volcánico**<sup>32</sup>

Se utiliza el término “estructuras volcánicas” y no “volcán” porque un mismo volcán puede generar varias estructuras (cómo El Jorullo, volcán histórico mexicano, que posee cuatro estructuras <sup>33</sup> ).

### Clasificación morfológica

Una última forma de clasificar los cuerpos volcánicos sería por su **forma**<sup>34</sup> (Figura 2.1) y en particular en el campo monogenético estudiado encontramos una gran variedad:

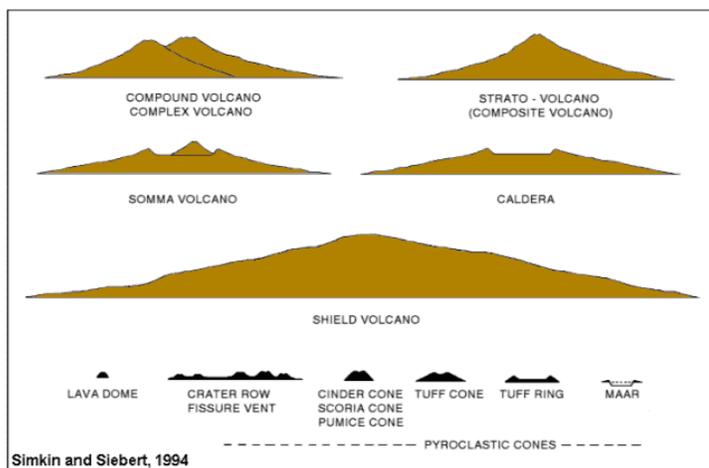


Figura 2.1- Morfologías 1. Imagen por Simkin. T- and Siebert, L. 1994.

Las primeras a abordar son los **conos**: existen conos de escoria que son producidos en erupciones estrombolianas con mucho gas resultando en rocas vesiculadas ; conos de toba, que son más sólidas y uniformes, creadas en erupciones freatomagmáticas (con presencia de agua) que ocasiona morfologías particulares como esta y los maares; **anillos** de toba, son como los conos pero de mayores dimensiones y lados más empinados (Figura 2.2);

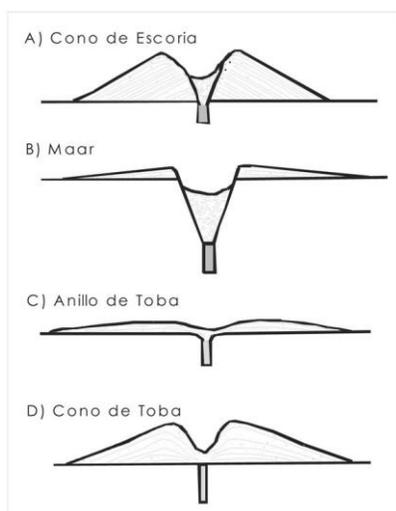


Figura 2.2- Morfologías 2 “Estructuras y morfologías de formas volcánicas explosivas pequeñas.” Por INSUGEO (Instituto Superior de Correlación Geológica)

**maares**, se forman cuando hay contacto con agua superficial y son más violentos que los anteriores, se caracterizan por una estructura bajo la superficie llamada Diatrema y la presencia de kimberlita. <sup>35</sup> Están también los conos de ceniza, formado de piroclastos (fragmentos) de menor tamaño a las escorias.

Los **Spatter** o cono soldado que se originan cuando el material expulsado por el volcán cae todavía parcialmente fundido, los fragmentos se aplastan y sueldan entre sí.

Los **Domos**<sup>36</sup>: montículos circulares que se originan de lava de alto contenido en sílice, cuya viscosidad no permite que la lava fluya demasiado lejos de su chimenea antes de solidificarse. Los domos pueden alcanzar alturas de varios cientos de metros, y pueden crecer lentamente y en forma continua durante meses e incluso años. Los lados de estas estructuras están formados de trozos inestables de roca. Debido a la posibilidad de acumulación de presión de gas, el domo puede a lo largo de su historia sufrir erupciones explosivas.

Finalmente, en los mayores tamaños tenemos a los **estratovolcanes** (Figura 2.1), que son originados por vulcanismo poligenético, se forman por la alternancia de momentos de actividad explosiva con otros de emisión de coladas de lava fluida. La lava no llega lejos por su viscosidad, formando conos de gran altura. La mayoría de los volcanes famosos en el mundo (presentes en corteza continental) son de este tipo. Un ejemplo de estos es el Popocatépetl.

Un segundo volcán poligenético son los **Escudo** (Figura 2.1) que son edificios formados esencialmente por acumulación de coladas basálticas, tienen también gran tamaño, hasta el centenar de kilómetros de diámetro, y pendientes de menos de 10 grados.

Por último, los volcanes de mayor tamaño son las **Calderas** (Figura 2.1), depresiones asociadas a las zonas centrales de edificios volcánicos poligenéticos cuyo origen se encuentra en una inestabilidad del volcán, causada normalmente por una subsidencia del terreno relacionada con un vaciado parcial de una cámara magmática somera, o como consecuencia de deslizamientos a gran escala de una parte del edificio volcánico.<sup>3</sup>

Todas estas diferencias suelen depender del magma, cada tipo de magma suele tener un tipo de erupción asociada. Los tipos de erupciones de menos a más explosivos: Hawaiiana, Estromboliana, Vulcaniana, Pliniana y aquellas con presencia de agua: Surtseyana, Freática, Freatomagmática y Freatoplina.

La cercanía de cuerpos generados por erupciones crea conglomerados de estructuras volcánicas llamados campos. En el caso particular de los volcanes monogenéticos, el dique que atraviesa el pulso de magma para llegar a la superficie colapsa una vez este sale. Ocasionalmente que cada pulso de magma deba buscar una nueva vía a la superficie (dique o grieta en la corteza) generando varios volcanes monogenéticos, cerca unos de otros. Formando así un **campo monogenético**.

Los campos volcánicos comprenden pequeños volcanes (por lo general con volúmenes inferiores a 1 km<sup>3</sup>), tales como conos cineríticos, maares, anillos de toba, volcanes escudo y flujos de lava.<sup>37</sup> Es dentro de estos campos monogenéticos, en particular el CVMG, que nos centraremos.

## **Cronometría**

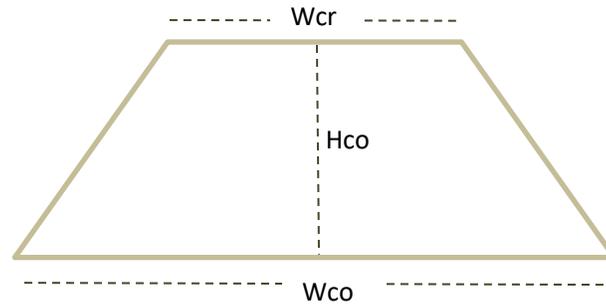
*“Los parámetros morfométricos de conos de escoria han sido utilizados ampliamente para determinar edades relativas”<sup>38</sup>*

Es una herramienta útil cuando no es posible determinar la edad radiométrica (por métodos físicos o químicos). La medida geométrica de los conos (Figura 2.3) se puede usar para determinar los efectos de degradación morfométrica (erosión) de un volcán. Para ello Wood (1980) propone ciertos parámetros para estimar su edad:

---

<sup>3</sup> La mayoría de conceptos de esta sección fueron obtenidos de: Sparks, R. S. J. (2015). Domes. In H. Sigurdsson, B. Houghton, S. McNutt, H. Rymer, & J. Stix (Eds.), *The Encyclopedia of Volcanoes*. Salvo aquellos que se especifique otra fuente

Uno de los principios de la cronometría es que los diámetros del cráter o la base no cambian, o la variación es poca conforme pasa el tiempo. Así, conforme los volcanes sean más antiguos, presentaran una relación  $Hco/Wco$  (o también llamado Índice de aspecto <sup>4</sup>o juventud) más baja, en comparación con conos más recientes. Sin embargo, esta relación no es exactamente lineal como se verá en el desarrollo del presente trabajo.



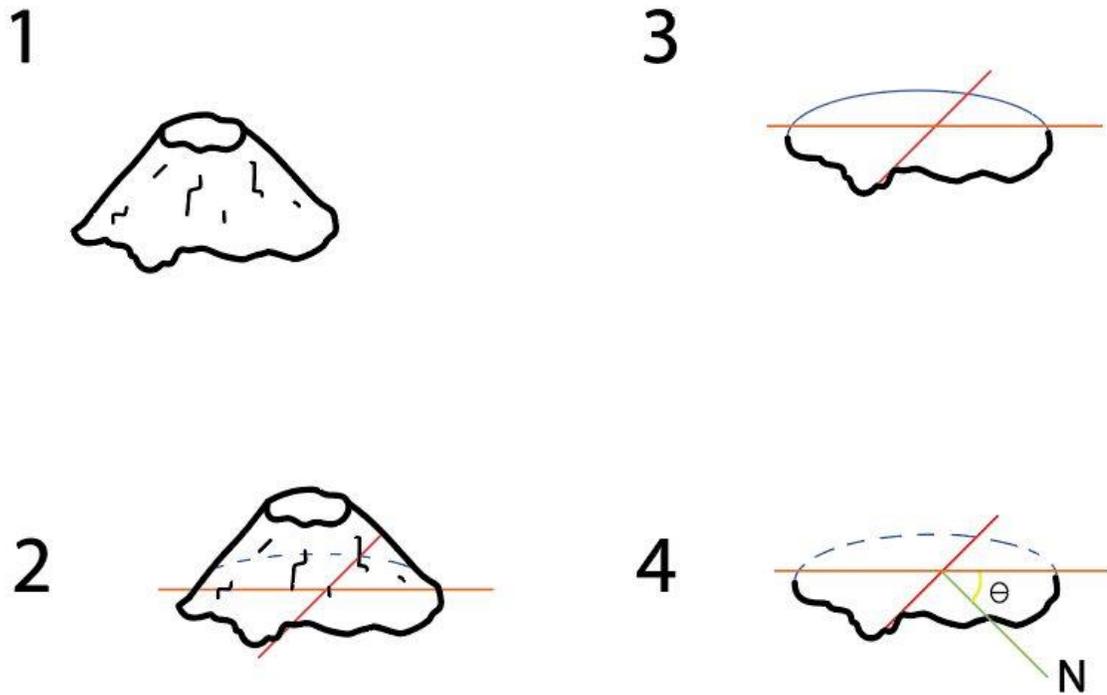
**Figura 2.3- Medida H/W:** Wcr: Diámetro del cráter. Wco: Diámetro basal del cono. Hco: altura del cono

Otra variable que tener en cuenta es la **Elongación** (Figura 2.4) que es el valor en grados de la orientación del eje mayor de la elipse basal del volcán y se mide en grados respecto al norte. La elongación de un volcán puede proporcionar información principalmente sobre la tectónica subyacente, considerándose que las fallas subyacentes son paralelas a la elongación de las estructuras en superficie. De forma que la elongación resulta paralela a la dirección de máximo esfuerzo.

La información de la elongación contribuye a conocer la historia eruptiva y comportamiento en el pasado de las estructuras. Por ejemplo, un volcán con una elongación marcada puede indicar que ha sufrido erupciones efusivas prolongadas en el pasado, mientras que un volcán con una elongación más difícil de determinar puede indicar que ha sufrido erupciones explosivas más cortas y violentas.

---

<sup>4</sup> Si el Índice de Aspecto ( $Hco/Wco$ ) es mayor a 0.20 se considera un cono del Holoceno. Si es menor, pertenece al Pleistoceno.



**Figura 2.4- Elongación:** Son los grados respecto al norte que se encuentra el eje mayor de la base. 1) La estructura volcánica 2) La base y sus diámetros, 3) La base tiende a tener un diámetro más grande que el resto, 4) Respecto a este diámetro (o eje mayor) se toman los grados conforme al Norte en sentido horario. Figura hecha por Esquivel A. Adriana para el presente trabajo.

## Capítulo 3- Marco Teórico

La teoría de las copulas es de suma relevancia para poder modelar dependencias complejas en distribuciones de probabilidad multivariadas de manera sistemática y flexible, teniendo importancia ya que son funciones que permiten caracterizar la relación de dependencia entre variables aleatorias de manera desacoplada de su comportamiento marginal individual. Pero para introducirlas será necesario mencionar primero otras herramientas estadísticas usadas en el presente trabajo.<sup>5</sup> Las siguientes definiciones provienen de “Estadística Elemental” de Johnson-Kuby:

### 3.1- Inferencia Estadística<sup>39</sup>

Las definiciones de Estadística Descriptiva como Media, Varianza, entre otras se encuentran en el Capítulo 9 (Anexo).

#### a. Asimetría (Skewness)

Las medidas de asimetría (o sesgo) son indicadores que permiten establecer el grado de simetría que presenta una distribución de probabilidad de una variable aleatoria.

Como eje de simetría consideramos una recta paralela al eje de ordenadas que pasa por la media de la distribución.

$$\text{Asimetría} = \frac{\sum_i^N (X_i - \bar{X})^3}{(N-1) \sigma^3}$$

#### b. Kurtosis

La curtosis (Kurtosis en inglés) es una característica de forma de su distribución de probabilidad. Una curtosis grande (alejada del 3 que es la asociada a una distribución normal) implica una mayor concentración de valores de la variable tanto muy cerca de la media de la distribución (pico) como muy lejos de ella (colas), al tiempo que existe una relativamente menor frecuencia de valores intermedios.

$$\text{Curtosis} = \sum_{i=1}^n \frac{(X_i - \mu)^4}{n \sigma^4}$$

#### c. Covarianza

Valor que indica el grado de variación conjunta de dos variables aleatorias respecto a sus medias.

$$\text{Cov}(X, Y) = E(X * Y) - E(X)E(Y)$$

Sin embargo, no es una medida fiable de correlación, ya que si dos variables X,Y son independientes, entonces su covarianza es cero. Aunque, si la covarianza es cero, no implica que sean independientes.

Además, varía valores en los reales, mientras que la correlación lo hace entre -1 y 1.

---

<sup>5</sup> Para consultar un concepto de probabilidad o estadística descriptiva, revisar el Anexo correspondiente al final del trabajo.

**d. Coeficientes de correlación:**

Son medidas de dependencia entre dos variables. Son adimensionales o absolutas y por ello son interpretables.

Dan valores en el intervalo [-1,1] y una correlación negativa se interpreta como que, si una variable disminuye, la otra aumenta. Una positiva implica que si una sube la otra también. Y si vale cero implica que no hay correlación.

**i. Pearson**

Es la más conocida de las medidas de correlación, su símbolo es una rho, y satisface las características de un coeficiente de correlación, sin embargo, rho es una variable aleatoria que depende de la distribución bivariada de (X, Y).

Las medidas de correlación poseen distribuciones que son independientes de la distribución bivariada de (X, Y) si X y Y son independientes y continuas.

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

**ii. Kendall**

Una medida de asociación para variables de escala ordinal y superior es la llamada tau de Kendall, representada con la letra griega Tau. Los datos son una muestra bivariada de tamaño n de forma  $(X_i, Y_i)$ , y se elabora de la siguiente manera:

Se toma el primero de los datos y clasifica el resto respecto a este en tres grupos:

1. Concordantes: si ambos miembros de una observación son mayores que sus miembros respectivos de la otra observación.  
 $N_c$  es el número de parejas concordantes.
2. Discordantes: si los dos números en una observación difieren en dirección opuesta (una negativa y otra positiva) de los respectivos miembros de la otra observación.  
 $N_d$  es el número de parejas discordantes.
3. Las parejas con empates (alguno de ambas entradas igual en ambas observaciones) no son ni discordantes ni concordantes.

La suma de los tres debe dar el número total de posibles parejas  $\frac{n(n-1)}{2}$ .

Así:

$$\tau = \frac{N_c - N_d}{\frac{n(n-1)}{2}}$$

Esta medida también se encuentra entre -1 y 1.

**iii. Spearman**

Una medida de asociación no paramétrica basada en rangos, puntajes del ordenamiento dado a

los datos, y que se aplica a variables de escala ordinal y superior es la rho de Spearman. Con los datos de una muestra bivariada de tamaño n, se crea una medida llamada  $R(Y_i)$  que es el *Rango de*  $Y_i$  cuando es comparada con otros valores de Y, no contando los empates.

Es decir, R es el lugar que ocuparía  $Y_i$  en caso de que los datos estuvieran ordenados de menor a mayor. De modo que la fórmula de la rho de Spearman sin considerar la posibilidad de empates es la siguiente:

$$\rho_{Spearman} = 1 - \frac{6 \sum_{i=1}^n (R(X_i) - R(Y_i))^2}{n(n^2 - 1)}$$

o bien

$$\rho_{Spearman} = 1 - \frac{6T}{n(n^2 - 1)}$$

de forma que la rho de Spearman te marca directamente un valor entre  $[-1,1]$ , sin embargo, al trabajarse en una prueba de hipótesis donde hay hipótesis nula y alternativa lo que interesa es solamente la dependencia o independencia, de modo que basta con calcular la estadística T y comparar con los valores de la tabla de Spearman.

La prueba estadística que será usada es sobre el valor del coeficiente de correlación, de modo que la hipótesis nula es  $H_0: \rho=0$  y la hipótesis alternativa es  $H_a: \rho \neq 0$ . En inferencia clásica se determina la zona de rechazo para una prueba, utilizando la información de la muestra para obtener un estadístico de prueba del cual se conoce su distribución bajo  $H_0$ . La inferencia correspondiente a la prueba, la podemos hacer a partir de la comparación del valor del estadístico en la muestra y un nivel crítico para saber cuándo rechazar o no la prueba. El p-valor tiene una utilidad similar.

#### iv.P-value

El p-valor se define como la probabilidad de obtener un resultado igual o más extremo que el observado, suponiendo que la hipótesis nula es cierta.<sup>40</sup>

Para determinar si el coeficiente de correlación es estadísticamente significativo, se calcula el p value y se compara con la distribución t. Primero se obtiene un t-score de un coeficiente de correlación r.

$$t = r \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$$

Donde r es el coeficiente de correlación muestral entre las muestras de X e Y, que se calcula como la covarianza dividida entre el producto de las desviaciones, y n es el tamaño de la muestra.<sup>41</sup>

Entonces se calcula el p-value como el resultado de la distribución t con n-2 grados de libertad y usando la prueba de dos colas. Dos colas significa que la Hipótesis nula es rechazada ya sea que  $p < 0$  o  $p > 0$

Es esencial tener presente que el nivel de confianza  $\alpha$  es la probabilidad de tener un error de tipo 1, es decir, rechazar una hipótesis nula dado que era verdadera Aquí se considerará el resultado como significativo si su p-value es menor a 0.05, dado que nuestro nivel de confianza será 5%.

Cuando se mencione un p-value de aquí en adelante, será referente al coeficiente de Pearson.

### 3.2- Teoría de Cópulas

#### Estocástico

Se denomina estocástico al sistema cuyo comportamiento intrínseco es no determinista. Un proceso estocástico es aquel que, en la medida en que el subsiguiente estado del sistema se determina tanto por las acciones predecibles del proceso como por elementos aleatorios.

#### Cópulas

Para el estudio y búsqueda, principalmente, de dichas funciones de distribución conjuntas, se han creado las cópulas: funciones de  $R^2$  en  $R$ , y que, dadas dos funciones de distribución marginales, genera una función de distribución conjunta.

Es sencillo ver que hay muchas formas de relacionar dos variables entre sí, por ello existen muchas cópulas diferentes, toda aquella que, dadas dos marginales, genera una función de distribución con las características matemáticas correctas, sin embargo, eso no asegura que dicha función obtenida sea la que mejor se adecua a los datos. Por lo tanto, en el presente trabajo se usará el p-value para medir el ajuste de la cópula a los datos.

Para la teoría nos apoyaremos en “Construcción de medidas de dependencia por medio de cópulas” de Erdely<sup>43</sup>

#### Definición 1

Sea  $H$  una función, se dice que es real de doble entrada si el  $\text{Dom}(H)=A \times B$  y el  $\text{Ran}(H)=C$  tal que  $A, B$  y  $C$  son subconjuntos no vacíos de los reales extendidos  $\bar{R}=[-\infty, \infty]$

#### Definición 1.1

Dada  $H$  función real de doble entrada. Diremos que  $H$  es 2-crecientes si  $0 \leq V_H(B)$  para todo rectángulo  $B$  cuyos vértices están en  $\text{Dom}(H)$ . Donde  $V_H(B)$  se refiere al  $H$ -volumen del conjunto  $B$  que, suponiendo que  $B$  es el conjunto  $[x_1, y_1] \times [x_2, y_2]$ , se define como:

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1).$$

Esta característica ayuda a diferenciar entre cópulas y funciones de doble entrada.

#### Ejemplo

Sean  $[u_1, u_2] \times [v_1, v_2]$  un rectángulo en  $[0,1] \times [0,1]$  las siguientes funciones son 2- crecientes considerando la función Indicadora  $1_A=1$  para  $a$  en  $A$  y 0 en otro caso (i)

1.  $\Pi(u,v):=uv$
2.  $M(u,v):= \min\{u,v\} = (u-v) 1_A + v$  para los  $(u,v)$  tal que  $u \leq v$
3.  $W(u,v):= \max\{u+v-1, 0\} = (u+v-1) 1_B$  con  $B=\{ (u,v) \text{ en } [0,1] \times [0,1]: u+v-1 \geq 0\}$

Aunque otra forma para comprobar si es 2 creciente es que sus derivadas parciales de segundo orden sean no negativas en su dominio.<sup>42</sup>

**Definición 1.2**

La función real de doble entrada H está fijada si dados  $a_1 = \inf(A)$  y  $a_2 = \inf(B)$  se da que  $H(x, a_2) = 0 = H(a_1, y)$  para todo  $(x, y)$  en  $\text{Dom}(H)$

**Definición 1.3**

F es función de distribución si es no decreciente y  $F(-\infty) = 0$  y  $F(\infty) = 1$ .

**Definición 1.4**

Si H es función real de doble entrada, es 2-creciente y está fijada para  $a = -\infty$  y  $H(\infty, \infty) = 1$  entonces H es función de distribución conjunta.

**Definición 1.5**

Sea  $H: S_1 \times S_2$  subconjunto de  $R^2$  y  $b_1 := \sup(S_1)$  y  $b_2 := \sup(S_2)$  entonces la función H tiene marginales:

$$F: S_1 \rightarrow R \text{ con } F(x) := H(x, b_2)$$
$$G: S_2 \rightarrow R \text{ con } G(y) := H(b_1, y)$$

**Definición 2**

Sea C función real de doble entrada, entonces C es una **2-cópula** o cópula bivariada si cumple con:

- 1.  $\text{Dom}(C) = [0,1] \times [0,1]$
- 2. C está fijada en cero y es 2-creciente
- 3. Para todo  $u, v$  se tiene  $C(u,1) = u$  y  $C(1,v) = v$

Nota: es fácil ver que las funciones definidas en (i) cumplen con ser cópulas bivariadas.

**Lema 2.1**

Sea C una cópula, se tiene:

$$W(u,v) \leq C(u,v) \leq M(u,v)$$

con W, M funciones de la definición 1.2

más aún:

$$W(u,v) \leq \Pi(u,v) \leq M(u,v)$$

A  $\Pi$  se le llama la cópula producto y caracteriza completamente la independencia.

Observación:

Sí  $C = M$  entonces la dependencia es directa es decir

$$P[X > x, Y \leq y] = 0, \text{ o bien } P[X \leq x, Y > y] = 0$$

Sí  $C = W$  entonces la dependencia es inversa es decir

$$P[X > x, Y > y] = 0, \text{ o bien } P[X \leq x, Y \leq y] = 0$$

**Teorema de Sklar**

Si bien aquí sólo lo enunciaremos, es el teorema fundamental para el uso de cópulas en estadística, puesto que nos garantiza su existencia. Su demostración, como muchas de las definiciones aquí vistas, pueden encontrarse en Erdely.

Sea  $H$  una función de distribución conjunta con marginales  $F$  y  $G$ . Entonces existe una cópula  $C$  tal que:

$$H(x,y) = C(F(x), G(y))$$

Si  $F$  y  $G$  son continuas entonces  $C$  es única, de otro modo  $C$  es única en  $\text{Ran}(F) \times \text{Ran}(G)$ . Por otra parte, si  $C$  es una cópula y  $F$  y  $G$  son funciones de distribución entonces  $H$  como se define es una función de distribución conjunta con marginales  $F$  y  $G$ .

Es decir, el Teorema de Sklar nos garantiza que dadas cualesquiera dos marginales existe una cópula. Las cópulas se dividen en paramétricas y no-paramétricas. La información que mencionamos a continuación sobre teoría de Cópulas puede ser encontrada en el Nelsen<sup>44</sup>.

La copula paramétrica es aquella que pertenece a una familia de distribuciones donde la dependencia entre las variables aleatorias depende de uno o más parámetros. Ejemplos son la cópula Gaussiana, cópula de Clayton, cópula de Student-t, cópula de Gumbel, cópula de Frank, etc. Por un lado, tienen una forma funcional conocida, pero la dependencia debe especificarse por los parámetros.

En cambio, la cópula no paramétrica es donde la dependencia no se define por parámetros, sino que se estima directamente a partir de los datos. Algunos ejemplos son la cópula de Kendall, cópula de Spearman, cópula de Kernel Smoothing, cópula de Bernstein, entre otros. Estas resultan más flexibles y no imponen una estructura funcional rígida, pero son más difíciles de especificar y estimar.

### 3.3- Criterios de Selección

Hay varios criterios para elegir la mejor cópula:

- Bayesian Information Criterion (BIC): Mide la calidad de ajuste del modelo penalizando la complejidad del mismo. El modelo con menor BIC es preferible.
- Akaike Information Criterion (AIC): Similar al BIC pero con una penalización menor por el número de parámetros. También se elige el modelo con menor AIC.
- Log-likelihood: Mide el ajuste del modelo a los datos observados. Cuanto mayor es el log-likelihood, mejor es el ajuste.
- Prueba de Bondia-Gallant: Compara si una copula propuesta es un buen ajuste o debe ser rechazada estadísticamente.
- Prueba de uniforme: Revisa si los residuales de una estimación copulista siguen realmente una distribución uniforme.
- Contrastes de Hosking: Comparan si una copula es significativamente más flexible que otra más restrictiva.

- Gráficos de QQ: Representan los percentiles empíricos (de la muestra) contra los teóricos (del ajuste), mostrando el ajuste de la cola.
- Gráfico de R-vínculo: Muestra si la dependencia estimada replica bien la observada en los datos.

### 3.3.1-BIC Best estimated<sup>45</sup>

Criterio de Información Bayesiano (BIC) es una medida de bondad de ajuste de un modelo estadístico, y es a menudo utilizado como un criterio para para la selección de modelos entre un conjunto finito de modelos

El BIC se calcula con:

$$\text{BIC} = -2 \log(L) + k \log(n)$$

donde L es la función de verosimilitud del modelo, k es el número de parámetros estimados en el modelo y n es el número de observaciones en los datos.

El BIC proporciona una medida relativa de la calidad de ajuste de un modelo, donde valores más bajos de BIC indican un mejor ajuste y valores más altos de BIC indican un peor ajuste. Además, el BIC penaliza la complejidad del modelo al incluir un término adicional que aumenta a medida que aumenta el número de parámetros estimados.

## 3.4- Tipos de cópulas

### 3.4.1- Cópulas no paramétricas

Se estiman directamente de las marginales o de los datos observados y nos centraremos en la cópula de Kernel y la de Bernstein.

Específicamente, lo que se conoce como 'Cópula de Kernel' es una familia de funciones, donde asignamos en este trabajo ese nombre a la función que se basa en la función Epanechnikov.<sup>46</sup>

Del mismo modo, denominamos 'Cópula de Bernstein' a una de las funciones pertenecientes a la familia de cópulas de kernel, en este caso basada en una función de aproximación de Bernstein.

#### a. Cópula Kernel smoothing<sup>47</sup>

Para calcular la copula de Kernel Smoothing, se debe seleccionar un kernel de suavizado. Un Kernel (o núcleo) es una función de densidad de probabilidad que debe ser simétrica respecto al origen y se usa para suavizar la estimación empírica de la cópula. Los kernels más comunes son el kernel Gaussiano, el kernel Epanechnikov y el kernel uniforme. En este trabajo se usará el Epanechnikov, por la simplicidad de su expresión matemática, lo que mejora su eficiencia computacional siendo

más rápida de calcular que otros kernels. Además que minimiza la varianza integrada de la estimación. <sup>48</sup>

Primero se determina el parámetro (h) que es cantidad de suavizado que se aplica a los datos. Este parámetro se puede ajustar mediante técnicas de validación cruzada o mediante otras técnicas de selección de modelos.

La validación cruzada consiste en dividir el conjunto de datos en k subconjuntos, tomar k-1 de los subconjuntos para ajustar el parámetro que varía en un rango h y finalmente se toma aquel parámetro que minimiza el error de predicción al compararlo con el subconjunto no elegido.

Luego, de no tenerse las funciones de distribución teóricas, se estima la función de densidad marginal para cada variable aleatoria utilizando la técnica de suavizado de kernel obteniendo  $F_x$  y  $F_y$  funciones de distribución.

La función que resultante es:

$$K(x,y) = \left(\frac{1}{h^2}\right) f\left(\frac{F_x}{h}, \frac{F_y}{h}\right)$$

donde:

- x,y son las variables independientes con  $F_i$  su función de distribución respectiva.
- h es el parámetro de suavizado o ancho de banda (bandwidth en inglés)
- f es una función de núcleo o kernel, que es una función simétrica y no negativa que integra a 1.

La función de núcleo f(x) puede ser una función gaussiana, uniforme, triangular o Epanechnikov (la que se usará en este caso).

Es importante destacar que la copula de Kernel Smoothing es útil para estimar copulas a partir de datos observados, lo que permite modelar la dependencia entre variables aleatorias sin hacer supuestos sobre la forma funcional de la copula subyacente.

#### i. Kernel Epanechnikov

Es una copula utilizada comúnmente en estadística para estimar la densidad de probabilidad y para suavizar datos. Es una función de densidad simétrica, no negativa y tiene soporte en el intervalo [-1, 1].

Se define matemáticamente como:

$$f(x,y) = \begin{cases} \left(\frac{3}{4}\right) (1 - x^2 - y^2) & \text{si } -1 \leq x \leq 1 \\ f(x,y) = 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, la función de Kernel suavizada con parámetro h es:

$$\text{Kernel}(x,y) = \left(\frac{1}{h^2}\right) \left(\frac{3}{4}\right) \left(1 - \left(\frac{F_x}{h}\right)^2 - \left(\frac{F_y}{h}\right)^2\right)$$

#### b. Cópula Bernstein<sup>49</sup>

La copula de Bernstein se define como:

$$C(u, v) = \sum_{k=0}^n \left[ \frac{n!}{(k!(n-k)!)} \right] u^k (1-u)^{(n-k)} v^k (1-v)^{(n-k)}$$

donde n es un número entero positivo y k es un número entero.

Para calcular la copula de Bernstein, se deben seleccionar un valor para n. Este valor determina la precisión de la aproximación de la copula. Suele depender del número de datos con los que se cuenta.

Como se observa en la definición de C(u, v) los coeficientes de Bernstein para cada valor de k se obtienen utilizando la fórmula de combinaciones de k en n:

$$\frac{n!}{(k!(n-k)!)}$$

### 3.4.2- Cópulas paramétricas

Se calcula un parámetro (o varios) utilizando las distribuciones marginales y un supuesto de distribución conjunta subyacente. A partir de este parámetro se obtiene la cópula. Entre las cópulas paramétricas destacan:

- Cópula normal: Se basa en la distribución normal multivariante.
- Cópula de Student: Extiende la copula normal para incorporar más curtosis.
- Cópula de Clayton: Modela dependencias asimétricas. Apropia para colas inferiores más dependientes.
- Cópula de Gumbel: Modela dependencias asimétricas. Apropia para colas superiores más dependientes.
- Cópula Frank: Permite diferentes tipos de dependencia, incluyendo asimetrías.
- Cópula Joe: Generaliza las copulas de Gumbel y Clayton.

#### a. Cópula de Frank<sup>50</sup>

Dentro de las cópulas paramétricas, la de Frank es la más versátil en términos de tener menos suposiciones previas sobre la distribución conjunta subyacente. Por ello, es la única de las paramétricas que explicaremos.

Esta copula se define como:

$$C(u, v) = \frac{-1}{\alpha} \ln \left[ 1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{(e^{-\alpha} - 1)} \right]$$

donde  $\alpha$  es un parámetro de forma que controla la dependencia entre las dos variables. El cual se puede estimar a partir de los datos utilizando técnicas de estimación de máxima verosimilitud. Para  $\alpha=0$  representa independencia completa. Cuando  $u$  tiende a 0 representa la dependencia en la cola inferior y cuando  $v$  tiende a 0 representa la dependencia de la cola superior, es así como permite asimetrías. Fue una de las primeras cópulas definidas por Murray Frank al buscar una función bivariada con propiedades de concavidad.

## Capítulo 4- Metodología

Se guardan los registros en un archivo csv de forma que cada columna es una de las categorías en las que los datos están agrupados. Se limpian y filtran los registros a fin de quedarse con valores numéricos únicamente y con las dos columnas que se usarán para la comparación de variables en curso.

Una vez filtrados los datos se usa la implementación computacional para ver el resumen estadístico de nuestras variables de forma separada. Se estiman y ajustan funciones de probabilidad a cada variable (marginales) con sus respectivos parámetros, para posteriormente calcular el BIC respectivo a cada cópula paramétrica y elegir la de mejor ajuste. Para las cópulas no paramétricas no es necesario el uso de las marginales, por lo que estas se ajustarán directamente de los datos originales.

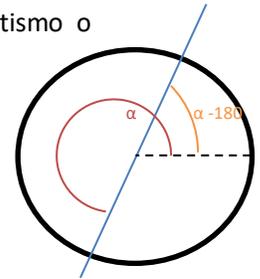
Se calcula el parámetro conforme a el comportamiento las funciones marginales y posteriormente se calculan las cópulas paramétricas (Frank) y no-paramétricas (Bernstein y Kernel Smoothing – Epanechnikov), además de sus parámetros por medio de validación cruzada. Una vez obtenidas las expresiones de las cópulas, se hacen simulaciones de los datos con las tres cópulas obtenidas y, a partir de los datos simulados, se calculan los coeficientes de correlación de Kendall, Spearman y Pearson, así como el p-value correspondiente para medir el ajuste de cada una de las cópulas, siendo este el procedimiento usual al utilizar cópulas en otros campos del conocimiento.

## Capítulo 5- Aplicación

Los datos que se toman son:

- Coordenadas (Longitud y latitud),
- Tipología de la estructura volcánica
- la elongación: anotada en un rango de 0 a 360° ( $\alpha$ ) y corregida a  $\alpha - 180$  (Figura 5.1).
- las edades (Por métodos radiométricos, químicos, paleomagnetismo o morfológicos)
- Wco: Diámetro basal.
- Wcr: Diámetro del cráter.
- Hco: Altura de la estructura volcánica.

con lo que se obtuvo el Hco/Wco



**Figura 5.1- Elongación:**  
Esquema de corrección del ángulo de elongación

Se hace una filtración y limpieza de la base de datos a fin de quedarse con registros numéricos. Posteriormente aplicamos la implementación computacional (Capítulo 9) para la estimación de las marginales y las cúpulas utilizando el BIC como criterio de selección.

### Caso de estudio:

La edad es tomada en cantidad de Millones de años antes del presente. Se pasa de los registros en decimal a números enteros utilizando que 0.1 Ma = 100,000 años, y se ponen todas las cantidades en años antes del presente (AP o BP por sus siglas en inglés).

Las edades “antes del presente” inician a partir del año 1950. Esta convención se debe al uso generalizado de la datación por radiocarbono, un método comúnmente empleado para determinar la antigüedad de los restos orgánicos antiguos. El radiocarbono, un isótopo del carbono presenta en pequeñas cantidades en la atmósfera terrestre, es absorbido por los seres vivos durante su existencia y dejan de hacerlo después de morir. A partir de ese momento, la cantidad de radiocarbono en el organismo comienza a disminuir debido a su desintegración radioactiva. La datación por radiocarbono es factible gracias al conocimiento de la tasa de desintegración del radiocarbono. Sin embargo, existen variaciones en la cantidad de radiocarbono presente en la atmósfera a lo largo del tiempo debido a factores como los cambios en la producción de radiación solar y la circulación oceánica. Estas fluctuaciones hacen que la datación precisa sea un desafío. Para abordar esta complejidad, los científicos establecieron un punto de referencia conocido como "año 1950" para la datación por radiocarbono. En ese año, se llevó a cabo un estudio detallado del contenido de radiocarbono en la atmósfera y se estableció como el punto de partida estándar.

La Latitud en el caso de México son los grados entre el ecuador y nuestra posición hacia el Norte

La Longitud, que son los grados a partir del meridiano 0 o de Greenwich, son negativos por estar del lado Este del meridiano 0

El Diámetro Basal ( $W_{co}$ ) se considera el promedio entre el máximo y el mínimo de los diámetros de la base de una estructura. Análogamente con el Diámetro del cráter.

Las tipologías encontradas dentro de la zona sureste del CVMG fueron Spatter, Cono, Escudo, Domo, Maar, Cono compuesto y Mesa de Lava. Con un total de 434 cuerpos volcánicos.

De los cuales 409 son Conos, Spatter o Domos. Para obtener cada conjunto de parejas de variables de cada comparación se filtró primero por tipología, de forma que se hará un análisis por tipo, utilizando formato csv. Se excluyeron columnas de los datos que no se usarían en cada comparación y, se eliminan celdas vacías o no-numéricas.

Con celdas no-numéricas nos referimos principalmente a algunos registros en las edades obtenidas por sobreposición de depósitos, que es la obtención de la edad radiométrica  $x$  de un depósito ajeno a la estructura estudiada en aquellos casos en que el depósito hubiese cubierto a la estructura volcánica, infiriendo que esta estructura es previa y por tanto su edad es " $< x$ " (menor que  $x$ ). Ya que, en todos los casos de este tipo,  $x$  toma los valores 0.4, 0.5 y 1 (en millones de años), lo cual implica un margen lo suficientemente grande (comparando con el resto de edades de la localidad) como para que suponer asignarles cualquier edad exacta a estos registros de " $< x$ " podría llevar a un error que afecte a las estimaciones. Por ejemplo: en los Spatter, 15 estructuras están datadas " $<1$ " y 23 estructuras datadas " $<0.5$ ", con la mayoría estando debajo de 1. Esto, sumado a aquellas estructuras sin edad asignada, explica por qué de 409 datos disponibles trabajaremos con menos de 200.

Una vez filtrados los datos se usa la implementación computacional en Python (la cual se encuentra en el Capítulo 9) para ver el resumen estadístico de nuestras variables de forma separada y después obtener las funciones de distribución marginales con sus respectivos parámetros, para posteriormente calcular el BIC respectivo a cada marginal y posteriormente cópula paramétrica, y elegir la de mejor ajuste (por la naturaleza aparentemente aleatoria de nuestros datos, dicha cópula elegida siempre es la de Frank. Capítulo 3) y calcular su parámetro conforme a el comportamiento las funciones marginales. Después se calculan las cópulas no-paramétricas de Bernstein y Kernel Smoothing, se hacen simulaciones de los datos con las tres cópulas obtenidas y finalmente se calculan los coeficientes de correlación y un p-value correspondiente para cada una.

Posteriormente se hace el resumen y muestran las gráficas y resultados estadísticos obtenidos para discutirlos. No obstante, la cópula de Frank quedará relegada a ser de uso auxiliar para los casos en las que las cópulas no-paramétricas no puedan dar resultados concluyentes, debido a que por ser paramétrica su parámetro ya induce una dependencia y es esta dependencia ( $\alpha$ ) la que genera la cópula (Capítulo 3). Es necesario aclarar que si bien en la literatura el parámetro de la cópula de Frank es  $\alpha$ , en la librería de Python utilizada está nombrada como  $\theta$  (theta) por lo que a partir de ahora se utilizará  $\theta$  para referirse al parámetro de esta cópula.

**Nota para el Lector:**

Es esencial tener presente que el nivel de confianza  $\alpha$  este caso es 5% (0.05). Al llevar a cabo cada prueba, adoptamos un enfoque formal, estableciendo una Hipótesis nula que postula que no hay dependencia y una Hipótesis alternativa de dos colas, es decir, sólo le pedimos que sea diferente de la no-dependencia (sin especificar si positiva o negativa). Y, de rechazarse la hipótesis nula, (e.d. tener p-value menor a 0.05) podemos decir que hay dependencia y a partir del signo del coeficiente de correlación deducir si es positiva o negativa dicha dependencia.

Además, mencionar que las pruebas estadísticas de este tipo lo que señalan es relación estadística (numérica) entre las variables, para saber si dicha relación es causal se necesita además un sustento propio del área al que las variables pertenecen (como es en este caso la vulcanología). En este caso, las pruebas surgen a partir de dudas sobre variables que sí se piensa que podrían tener una relación causal, es por ello que de encontrar la relación estadística diremos directamente que lo que hay es dependencia.

En resumen:

$$H_0: \rho=0$$

$$H_a: \rho \neq 0$$

$$\text{Si } p\text{-value} < 0.05 = \alpha$$

Entonces rechazamos  $H_0$

Al quedarnos con  $H_a: \rho \neq 0$  entonces hay dependencia

Como  $\rho$  es coeficiente de correlación:

Si  $\rho > 0$  la dependencia es positiva

Si  $\rho < 0$  la dependencia es negativa

Si  $|\rho| < 0.1$  consideraremos que no es suficientemente significativo para afirmar la existencia de dependencia (por estar cercano a 0) a pesar de haber obtenido un p-value  $< \alpha$

En otro caso, si p-value  $> \alpha$  directamente decimos que no hay evidencia suficiente en contra de  $H_0$ , en otras palabras, no podemos afirmar que haya dependencia entre las variables.

## Capítulo 6- Desarrollo y Resultados

Se emplearon los datos de las 434 entradas disponibles, dividiéndose en Spatter, Conos y Domos; habiendo 251, 115 y 43 respectivamente. Con un total de 409 entradas utilizables.

Hay 173 entradas con edades asignadas: 138 de Spatter, 22 de Domos, 13 de Conos. Dichas edades se obtuvieron por métodos químicos  $^{40}\text{Ar}/^{39}\text{Ar}$ ,  $^{230}\text{Th}$ - $^{238}\text{U}$  o  $^{14}\text{C}$ . Si consideramos la edad relativa (aquella estimada a partir de que la estructura este por encima o cubierta por un flujo de lava con Edad radiométrica conocida) con base a la morfología de la estructura volcánica, obtendríamos un total de 46 edades para los Conos. Sin embargo, primero buscaremos la dependencia entre la morfometría y la edad para ello iniciamos considerando sólo las edades absolutas obtenidas por métodos químicos y comparándolo con su Índice de Aspecto (Altura/diámetro basal) correspondiente.

También un análisis detallado de los resultados de las Cópulas no paramétricas (Cópula de Bersntein y Cópula de Kernel) y la Cópula paramétrica de Frank será realizado.

### 6.1 Conos de Escoria

De aquí en adelante empleo el término “Conos” para referir a los Conos de escoria. Se utiliza el índice de aspecto definido en el capítulo 2 para ver su relación con la Edad.

#### 6.1.1 Edad vs Índice de aspecto

##### Análisis Exploratorio de datos

De las edades de la base de datos, hay 13 datos de conos. Hay 4 de ellos que están en miles de años, el resto está en cientos de miles o millones de años. Por lo tanto, mostramos los resultados del análisis con y sin esos 4 datos llamando al conjunto de todos (en cantidades enteras) “Conos de Edad Ajustada”, y al otro conjunto sin esos 4 datos menores a 20,000 años lo llamaremos simplemente “Conos B”. (Tabla 6.1)

El nombre de “Conos de Edad Ajustada” hace referencia a que se pasa de los registros en decimal a números enteros (Capítulo 5)

De modo que se aplica el método basado en la teoría de cópulas a las variables Edad e Índice de aspecto (H/D) para dos casos: uno en el conjunto “Conos de edad ajustada” (13 datos) y otra en el conjunto “Conos B” (con 9 datos). El objetivo en esta sección es dar una conclusión en sentido de los resultados de Wood, C. A. (1980), por medio de representar la dependencia entre estas dos variables.

Las estadísticas de Conos con edad ajustada y Conos B son mostradas en la Tabla. (Tabla 6.1) .

**Tabla 6.1 – Estadísticas de Conos** Resultados de medidas estadísticas con la lista original de datos de Conos: “Conos Edad Ajustada” (todos los datos) y “Conos B” (sin datos menores a 20,000 años). En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

Conos Edad Ajustada	Conos B
Tamaño= 13	Tamaño= 9
Nombre= [Índice de Aspecto,Edad]	Nombre= [Índice de Aspecto,Edad]
Mínimo= [0.124774,2218]	Mínimo= [0.124774,80000]
Quartil 1= [0.147434,11789.5]	Quartil 1= [0.142932,612500]
Promedio= [0.212687,525165]	Promedio= [0.219257,755556]
Mediana= [0.197909,620000]	Mediana= [0.197909,780000]
Quartil 3= [0.245634,780000]	Quartil 3= [0.251095,817500]
Máximo= [0.429744,1.51e+06]	Máximo= [0.429744,1.51e+06]
Rango= [0.30497,1.50778e+06]	Rango= [0.30497,1.43e+06]
Varianza= [0.00659936,2.21406e+11]	Varianza= [0.0089946,1.38028e+11]
Desviación Estándar= [0.0812365,470538]	Desviación Estándar= [0.0948399,371521]
Skewness= [1.59511,0.411592]	Skewness= [1.41573,0.374266]
Kurtosis= [6.63186,2.75815]	Kurtosis= [5.54241,5.94012]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.00659936 -5399.33 ] [ -5399.33 2.21406e+11 ]]	CoVarianza= [[ 0.0089946 -13565.6 ] [ -13565.6 1.38028e+11 ]]
Desviación Estándar= [0.0812365,470538]	Desviación Estándar= [0.0948399,371521]
Matriz de Correlación de Pearson= [[ 1 -0.141252 ] [ -0.141252 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.385003 ] [ -0.385003 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.196116 ] [ -0.196116 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.377168 ] [ -0.377168 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.287297 ] [ -0.287297 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.491596 ] [ -0.491596 1 ]]

En la tabla 6.1 se observa que los valores de la varianza del conjunto Conos Edad Ajustada correspondiente a la Edad ( $2.21 \times 10^{11}$ ) resulta el doble de la varianza respectiva en Conos B, conjunto que tiene únicamente 4 datos menos.

El conjunto de Conos exhibe un rango significativo de edades ajustadas, desde un mínimo de 2218 años (correspondiente a "El Zoyate") hasta un máximo de 1.5 millones de años (relacionado con "C. Colorado"). En contraste, el conjunto de Conos B muestra un rango más restringido, con un mínimo de 80,000 años (asociado a "C. La Laguna") y la misma edad máxima de 1.5 Ma.

El análisis detallado revela una proporción sorprendente entre el mínimo y el máximo en cada conjunto, destacando un cambio de 676.28 veces la Edad mínima en comparación con la máxima en Conos Edad Ajustada a 18.75 veces en Conos B. Esta discrepancia sugiere implicaciones significativas para la comprensión de la historia geológica y la evolución temporal de estas formaciones.

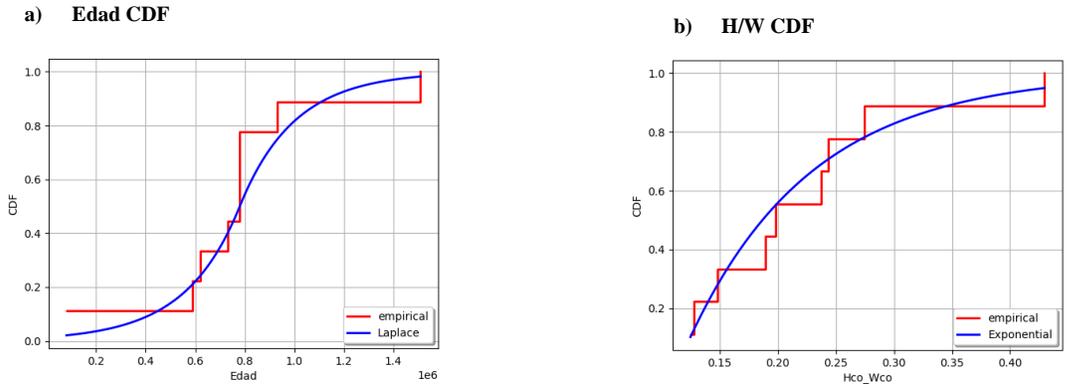
### Conos B: Marginales

Las marginales (funciones de probabilidad ajustadas a cada variable) en el caso del conjunto “Conos B” fueron:

$$F_{Edad} = \text{Laplace} (\mu = 780000, \lambda = 4.54545 \times 10^{-6})$$

$$F_{H/W} = \text{Exponential} (\lambda = 9.44946, \text{gamma} = 0.113431)$$

Y sus gráficas son (Figura 6.1):



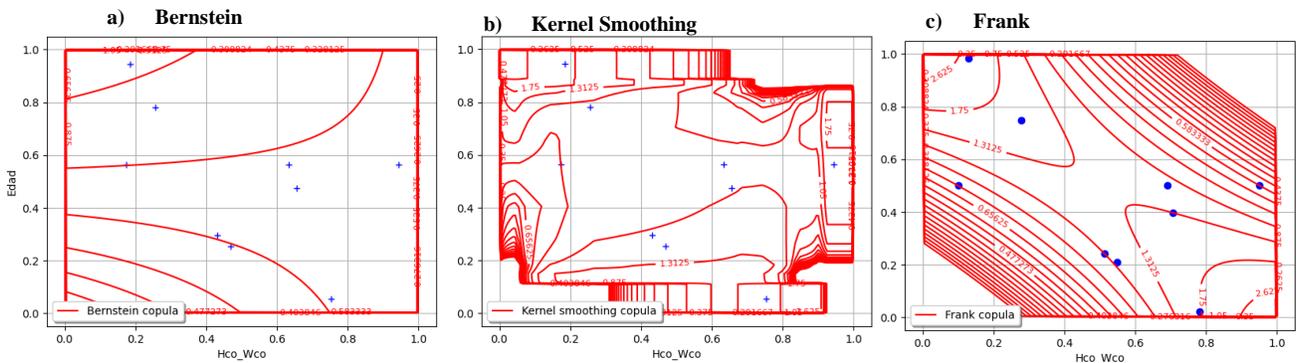
**Figura 6.1** | Funciones de Distribución Acumulativa (CDF) de las variables de “Conos B” (sin datos menores a 20,000 años):  
 | a)Edad b)Índice de Aspecto

### Conos B: Cópulas

Ahora se calculan y muestran las gráficas de cada cópula: Bernstein, Kernel Smoothing y Frank. Al finalizar el análisis se pondrá la expresión matemática de aquellas que resulten las mejor ajustadas.

La cópula se estima (Capítulo 3) y esta asigna a cada par de valores en  $R^2$ , con entradas en el intervalo  $[0,1]$ , un único valor también en el intervalo  $[0,1]$ . Dado que a una pareja de valores le asigna un tercer valor, su representación gráfica en dos dimensiones consiste en una serie de subconjuntos del espacio  $[0,1] \times [0,1]$ , que indican cual es la medida de los puntos en ese conjunto, similar a las curvas de nivel. Se utiliza el término “medida” puesto que las curvas no van en el intervalo  $[0,1]$  como lo hace la medida de la probabilidad usual, para mejorar su visualización se duplicó la diferencia entre cada contorno de nivel (Capítulo 9: Anexo, Figura 9.1) por lo que estas, en cambio, varían en el intervalo  $[0,2]$ .

Una vez estimados los parámetros de las cópulas, se grafican. Siendo las resultantes de “Conos B” las siguientes (Figura 6.2):



**Figura 6.2** | Gráfica de la Cópula de “Conos B” (Sin edades menores a 80,000): a)Bernstein b)Kernel Smoothing c)Frank

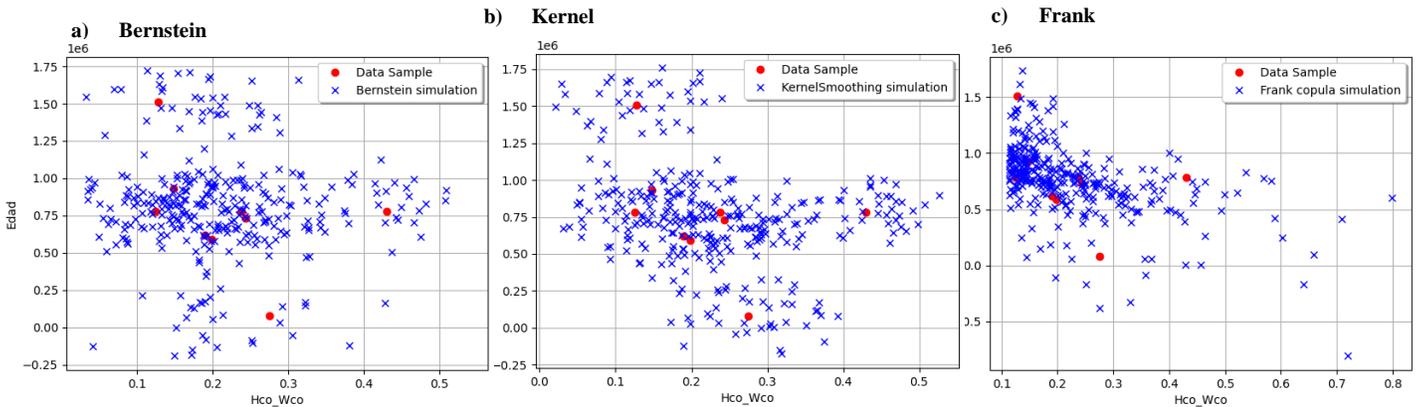
Es similar a una gráfica de curvas de nivel, con cada nivel siendo una medida de probabilidad en el intervalo  $[0,2]$ . Teniendo en cuenta que lo que nos interesa es el comportamiento general, lo primero a denotar es la diferencia de la Cópula Kernel con las otras dos: ya que la cópula paramétrica proviene de un parámetro, esto ocasiona que las curvas de la gráfica sean más suavizadas que las pertenecientes a las cópulas no-paramétricas que se estiman directamente de las marginales. (Observar gráficas b) y c) de la Figura 6.4). Los ejes son valores de las funciones de distribución, de forma que las rectas rojas nos indican qué ocurre cuando una de las variables aumenta en probabilidad y de ahí nos fijamos qué ocurre con la otra variable.

Por ejemplo, al examinar la cópula de Kernel, tomando el punto  $(0.4,0)$  se aprecia que este punto es una de las rectas rojas, en cambio, si disminuimos el eje x y nos acercamos al 0, siguiendo la curva roja localizada, la otra variable aumentará llegando al  $(0,0.2)$  también sobre una recta roja. De esta forma se puede apreciar que la gráfica señala una relación inversa entre ambas variables.

Este comportamiento no es tan evidente en la representación gráfica de la cópula de Bernstein. Las rectas rojas "exteriores" o cercanas a los ejes, que forman un cuadrado, sugieren independencia entre las variables. No obstante, esta interpretación puede tener excepciones según el comportamiento de las curvas rojas internas, como se abordará en comparaciones posteriores.

En tanto a la Cópula de Frank, se evidencia una dependencia negativa más evidente que en la Cópula de Kernel, es por ello que la cópula de Frank será utilizada de forma auxiliar para evidenciar los tipos de dependencia.

Teniendo las cópulas, se realizan las simulaciones y muestran sus gráficos de dispersión (Figura 6.3):



**Figura 6.3|** Dispersión de Edad contra H/W de “Conos B” (sin edades menores a 20,000) usando cada ajuste de cópula (a)Bernstein, b)Kernel, (c)Frank)

Resaltando que la simulación de la cópula de Frank exhibió un mayor número de edades negativas en comparación con las demás. Efectuando una comparación entre las simulaciones de las Cópulas de Kernel y Bernstein, se observa que la cópula de Kernel resulta la mejor aproximación. Observando los datos simulados más cercanos a los datos originales (puntos rojos). Para determinar la significancia o no de la

asociación entre los datos reales y simulados se observa el p-value del resumen estadístico de los datos simulados. (Tabla 6.2) Recordando que el p-value corresponde al ajuste del coeficiente de correlación de Spearman y se revisará el promedio de los tres coeficientes para determinar si está o no lejano al 0.

**Tabla 6.2 – Estadísticas de los datos Simulados** a con la cópula de Kernel y la Cópula de Bernstein del conjunto Conos B. En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

----- Cópula de KernelSmoothing	----- Cópula de Bernstein
Conos B	Conos B
Tamaño= 337	Tamaño= 337
Nombre= [Índice de Aspecto,Edad]	Nombre= [Índice de Aspecto,Edad]
Mínimo= [0.0202866,-175648]	Mínimo= [0.0316382,-192305]
Quartil 1= [0.15646,587257]	Quartil 1= [0.147384,663551]
Promedio= [0.229642,750232]	Promedio= [0.214542,827161]
Mediana= [0.213719,741944]	Mediana= [0.196236,818611]
Quartil 3= [0.292376,893503]	Quartil 3= [0.264929,963734]
Máximo= [0.526883,1.75806e+06]	Máximo= [0.538296,1.72001e+06]
Rango= [0.506597,1.9337e+06]	Rango= [0.506658,1.91232e+06]
Varianza= [0.0118049,1.40675e+11]	Varianza= [0.0106419,1.43213e+11]
Desviación Estándar= [0.10865,375066]	Desviación Estándar= [0.10316,378435]
Skewness= [0.565112,0.336123]	Skewness= [0.844028,0.0421823]
Kurtosis= [2.86262,3.78865]	Kurtosis= [3.59121,3.88708]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.0118049 -12113 ] [ -12113 1.40675e+11 ]]	CoVarianza= [[ 0.0106419 1245.61 ] [ 1245.61 1.43213e+11 ]]
Desviación Estándar= [0.10865,375066]	Desviación Estándar= [0.10316,378435]
Matriz de Correlación de Pearson= [[ 1 -0.297244 ] [ -0.297244 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.0319065 ] [ 0.0319065 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.20065 ] [ -0.20065 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.00907871 ] [ -0.00907871 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.305888 ] [ -0.305888 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.0160079 ] [ -0.0160079 1 ]]
----- pvalue=2.657469x10^-8	----- pvalue=0.559423

Por ende, se muestra la expresión de la cópula de Kernel y su Bandwidth (Capítulo 3):

$$\text{Bandwidth} = 0.120195$$

$$\frac{3}{4} * \text{Bandwidth}^2 = \frac{3}{0.0577873521}$$

$$\text{Kernel}(F_x, F_y) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{F_x}{0.120195} \right)^2 - \left( \frac{F_y}{0.120195} \right)^2 \right)$$

Y las funciones de distribución marginales asociadas a cada variable de este conjunto (Conos B) fueron:

$$F_{\text{Edad}} = \text{Laplace} (\mu= 780000, \lambda= 4.54545 \times 10^{-6})$$

$$F_{H/W} = \text{Exponential} (\lambda = 9.44946, \text{gamma} = 0.113431)$$

La Cópula de Kernel que es la función que relaciona la Edad con el Índice de Aspecto por medio de sus funciones de distribución (marginales estimadas al inicio de esta sección. Es importante subrayar que esta ecuación no es resoluble, es decir, no puede despejarse directamente. En consecuencia, no es posible presentar una expresión del tipo  $x = f(y)$  o  $\text{Edad} = f(H/W)$ . La justificación matemática de la imposibilidad de obtener una expresión despejada se aborda en el Capítulo 7.

Por todo lo anterior, la forma de obtener un valor para la Edad y valor en índice de aspecto, por medio de nuestra cópula, es utilizando métodos numéricos.

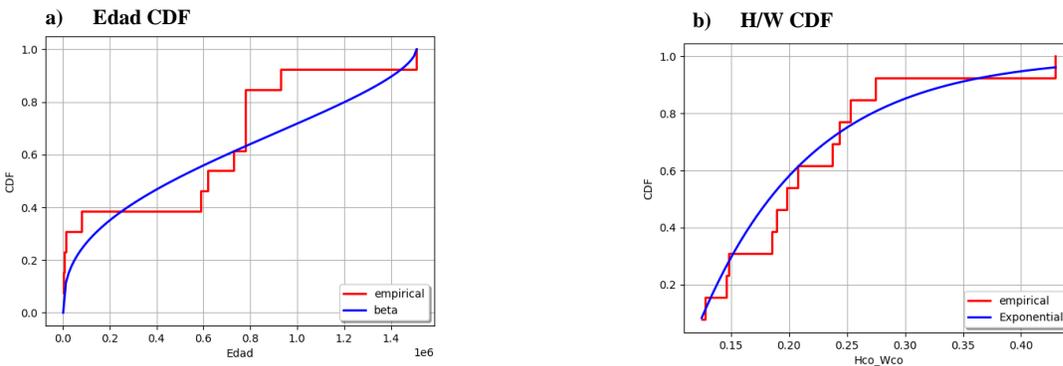
### Conos Edad Ajustada: Marginales

Las funciones ajustadas a cada variable en el conjunto “Conos Edad Ajustada” fueron (Figura 6.4):

$$F_{\text{Edad}} = \text{Beta}(\alpha = 0.39, \beta = 0.6, a = 2218, b = 1.51e+06)$$

$$F_{\text{H/W}} = \text{Exponential}(\lambda = 10.3916, \gamma = 0.116456)$$

Con gráficas:



**Figura 6.4** | Funciones de Distribución Acumulativa (CDF) de las variables de “Conos Edad Ajustada: a) Edad b) Índice de Aspecto

El análisis previo (Análisis exploratorio de los datos) explica por qué difieren las funciones marginales ajustadas a la Edad en cada conjunto, pasando de una función de distribución Laplace a una Beta, mientras que para el Índice de Aspecto (que tiene una variación en valores menor al de la Edad) su marginal se mantiene la misma para ambos conjuntos.

### Conos Edad Ajustada: Cópulas

Se estiman las cópulas no paramétricas y las paramétricas (Capítulo 3) y las gráficas respectivas para “Conos Edad Ajustada” (Figura 6.5). La Cópula de Bernstein muestra una correlación negativa principalmente en los puntos extremos (cerca de 0 y 1). Y la Cópula paramétrica (Frank) mantiene esa tendencia negativa mostrada en sus curvas internas.

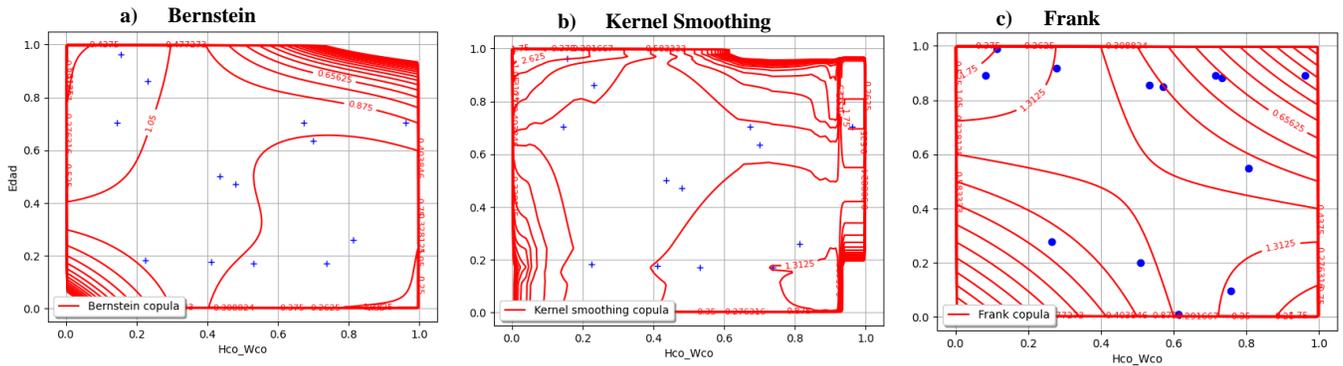


Figura 6.5| Gráfica de la Cópula de “Conos Edad Ajustada”: a)Bernstein b)Kernel Smoothing c)Frank

Las simulaciones resultantes de las Cópulas para “Conos Edad Ajustada” son (Figura 6.6) :

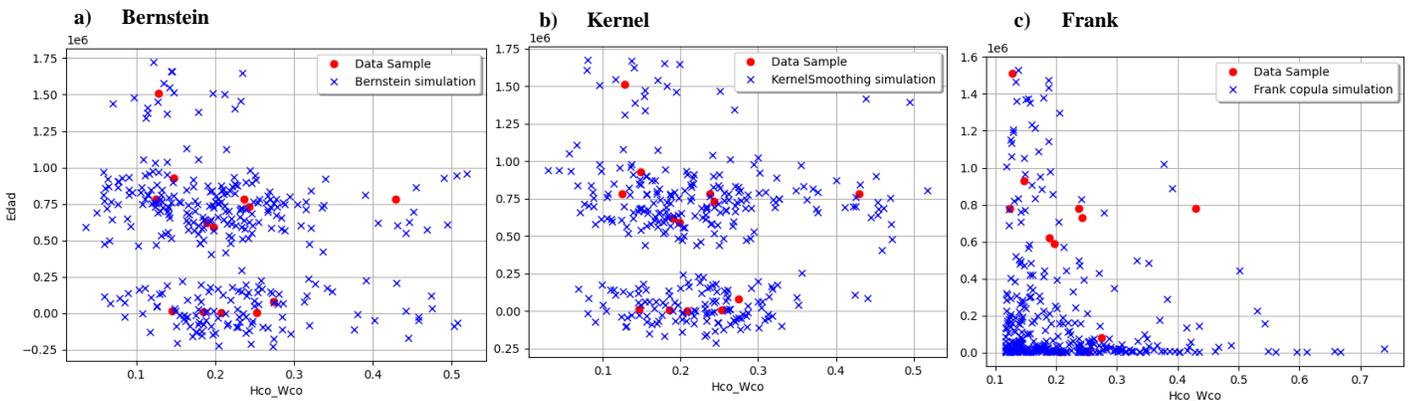


Figura 6.6 | Dispersión de Edad contra H/W de “Conos edad Ajustada” usando el ajuste de cópula ( a) Bernstein, b) Kernel, c) Frank)

Se observa que la cópula de Frank no parece funcionar con estos datos, generando datos con Índice de aspecto superior al 0.5 e incluso en 0.8, lo cual no es posible ya que hablamos de una pendiente (la pendiente máxima posible en un montículo es de 0.6<sup>51</sup>). Por otro lado, viendo la similitud entre los gráficos de a) y b) e la Figura 6.6, se diferencian por tener los datos simulados más cercanos a los originales en el gráfico a) correspondiente a la simulación proveniente de la Cópula de Bernstein. Mientras que la de Kernel nos da datos simulados más aleatorios. Nuevamente, esto queda confirmado al observar el p-value respectivo a cada cópula en la Tabla 6.3

**Tabla 6.3 – Estadísticas de los datos Simulados** a con la cópula de Kernel y la cópula de Bernstein del conjunto Conos Edad Ajustada. En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

----- Cópula de KernelSmoothing	----- Cópula de Bernstein
Conos Edad Ajustada	Conos Edad Ajustada
Tamaño= 337	Tamaño= 337
Nombre= [Índice de Aspecto,Edad]	Nombre= [Índice de Aspecto,Edad]
Mínimo= [0.028362,-215506]	Mínimo= [0.0354682,-230742]
Quartil 1= [0.155409,79583.3]	Quartil 1= [0.140858,120617]
Promedio= [0.22125,517014]	Promedio= [0.209682,530987]
Mediana= [0.21279,619325]	Mediana= [0.198213,630573]
Quartil 3= [0.269424,787956]	Quartil 3= [0.250126,806124]
Máximo= [0.5184,1.67337e+06]	Máximo= [0.520006,1.72119e+06]
Rango= [0.490038,1.88888e+06]	Rango= [0.484537,1.95194e+06]
Varianza= [0.00847039,1.91174e+11]	Varianza= [0.00911828,1.86538e+11]
Desviación Estándar= [0.0920347,437234]	Desviación Estándar= [0.0954897,431900]
Skewness= [0.749679,0.306999]	Skewness= [1.0054,0.150013]
Kurtosis= [3.57999,2.61912]	Kurtosis= [4.25563,2.59924]
-----	-----
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
-----	-----
CoVarianza= [[ 0.00847039 -2357.22 ] [ -2357.22 1.91174e+11 ]]	CoVarianza= [[ 0.00911828 -9284.23 ] [ -9284.23 1.86538e+11 ]]
Desviación Estándar= [0.0920347,437234]	Desviación Estándar= [0.0954897,431900]
Matriz de Correlación de Pearson= [[ 1 -0.058578 ] [ -0.058578 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.225116 ] [ -0.225116 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0683552 ] [ -0.0683552 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.166879 ] [ -0.166879 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.101785 ] [ -0.101785 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.252735 ] [ -0.252735 1 ]]
-----	-----
pvalue=0.28359710	pvalue=3.0341875x10^-5

Si bien, como lo visto en el Marco teórico (capítulo 3) y el análisis exhaustivo de la cópula de Bernstein por Martín Díaz<sup>52</sup> y el Van Huong<sup>53</sup>, dicha cópula funciona con aproximación numérica, por lo que carece de una expresión exacta de función bivariada  $y$ , sobre todo, abreviada que podamos otorgar (como aquella de Kernel), ya que proviene de una suma de  $n$  términos donde  $n$  depende del número de observaciones o datos originales (capítulo 3). Por tanto, nos limitaremos a dar los dos parámetros (de cada variable) que el método de ajustar dicha cópula nos otorgó (siendo  $p_i$  el parámetro de la variable  $i$ ), parámetros que utiliza la librería de Python para facilitar el cómputo.

$$\text{Bernstein}(x, y, p_x, p_y) = \text{Bernstein}(x, y, 0.0978587, 240916)$$

Y las funciones de distribución asociadas a cada variable del conjunto Conos Edad Ajustada fueron:

$$F_{\text{Edad}} = \text{LogUniform}(a_{\text{Log}} = 7.63537, b_{\text{Log}} = 14.2922)$$

$$F_{H/W} = \text{Exponential}(\lambda = 10.3916, \text{gamma} = 0.116456)$$

### Comparación

Los resultados de los estadísticos de los datos simulados, limitándonos a comparar las cópulas no paramétricas, resultaron (Tabla 6.4 y 6.5):

**Tabla 6.4 – Estadísticas de Conos con la cópula de Bernstein ajustada a datos de “Conos edad ajustada” (todos los datos) y “Conos B” (sin datos menores a 20,000 años). En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]**

----- Cópula de Bernstein Conos Edad Ajustada -----	----- Cópula de Bernstein Conos B -----
Tamaño= 337	Tamaño= 337
Nombre= [Índice de Aspecto,Edad]	Nombre= [Índice de Aspecto,Edad]
Mínimo= [0.0354682,-230742]	Mínimo= [0.0316382,-192305]
Quartil 1= [0.140858,120617]	Quartil 1= [0.147384,663551]
Promedio= [0.209682,530987]	Promedio= [0.214542,827161]
Mediana= [0.198213,630573]	Mediana= [0.196236,818611]
Quartil 3= [0.250126,806124]	Quartil 3= [0.264929,963734]
Máximo= [0.520006,1.72119e+06]	Máximo= [0.538296,1.72001e+06]
Rango= [0.484537,1.95194e+06]	Rango= [0.506658,1.91232e+06]
Varianza= [0.00911828,1.86538e+11]	Varianza= [0.0106419,1.43213e+11]
Desviación Estándar= [0.0954897,431900]	Desviación Estándar= [0.10316,378435]
Skewness= [1.0054,0.150013]	Skewness= [0.844028,0.0421823]
Kurtosis= [4.25563,2.59924]	Kurtosis= [3.59121,3.88708]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 0.00911828 -9284.23 ] [ -9284.23 1.86538e+11 ]]	CoVarianza= [[ 0.0106419 1245.61 ] [ 1245.61 1.43213e+11 ]]
Desviación Estándar= [0.0954897,431900]	Desviación Estándar= [0.10316,378435]
Matriz de Correlación de Pearson= [[ 1 -0.225116 ] [ -0.225116 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.0319065 ] [ 0.0319065 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.166879 ] [ -0.166879 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.00907871 ] [ -0.00907871 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.252735 ] [ -0.252735 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.0160079 ] [ -0.0160079 1 ]]
----- pvalue=3.0341875x10^-5	----- pvalue=0.559423

----- Cópula de KernelSmoothing Conos Edad Ajustada -----	----- Cópula de KernelSmoothing Conos B -----
Tamaño= 337	Tamaño= 337
Nombre= [Índice de Aspecto,Edad]	Nombre= [Índice de Aspecto,Edad]
Mínimo= [0.028362,-215506]	Mínimo= [0.0202866,-175648]
Quartil 1= [0.155409,79583.3]	Quartil 1= [0.15646,587257]
Promedio= [0.22125,517014]	Promedio= [0.229642,750232]
Mediana= [0.21279,619325]	Mediana= [0.213719,741944]
Quartil 3= [0.269424,787956]	Quartil 3= [0.292376,893503]
Máximo= [0.5184,1.67337e+06]	Máximo= [0.526883,1.75806e+06]
Rango= [0.490038,1.88888e+06]	Rango= [0.506597,1.9337e+06]
Varianza= [0.00847039,1.91174e+11]	Varianza= [0.0118049,1.40675e+11]
Desviación Estándar= [0.0920347,437234]	Desviación Estándar= [0.10865,375066]
Skewness= [0.749679,0.306999]	Skewness= [0.565112,0.336123]
Kurtosis= [3.57999,2.61912]	Kurtosis= [2.86262,3.78865]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 0.00847039 -2357.22 ] [ -2357.22 1.91174e+11 ]]	CoVarianza= [[ 0.0118049 -12113 ] [ -12113 1.40675e+11 ]]
Desviación Estándar= [0.0920347,437234]	Desviación Estándar= [0.10865,375066]
Matriz de Correlación de Pearson= [[ 1 -0.058578 ] [ -0.058578 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.297244 ] [ -0.297244 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0683552 ] [ -0.0683552 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.20065 ] [ -0.20065 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.101785 ] [ -0.101785 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.305888 ] [ -0.305888 1 ]]
----- pvalue=0.28359710	----- pvalue=2.657469x10^-8

**Tabla 6.5 – Estadísticas de Conos con la cópula de Kernel ajustada a datos de “Conos edad ajustada” (todos los datos) y “Conos B” (sin datos menores a 20,000 años). En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]**

De los análisis anteriores, se deduce que al trabajar exclusivamente con los datos de millones ("Conos B"), la cópula de Kernel proporciona la mejor aproximación, mientras que la de Bernstein arroja resultados no concluyentes. Esta conclusión se basa en el p-value obtenido de los datos simulados con la cópula de Bernstein, el cual indica "no tener suficiente evidencia en contra de la hipótesis nula" (Tabla 6.4). En contraste, la cópula de Kernel revela coeficientes de correlación negativos y significativamente diferentes de cero (en el rango de -1 a

1), siendo el p-value resultante de  $2.657 \times 10^{-8}$ . Este resultado sugiere que la cópula de Kernel se ajusta mejor a este conjunto de datos y que existe una dependencia negativa entre las variables.

En ambos casos: Conos de Edad Ajustada y Conos B, con sus respectivas copulas mejor ajustadas, nos resultan los coeficientes de correlación negativos, implicando que hay una dependencia.

### 6.1.2: Edad vs Localización

Se procede con las comparaciones de la edad con la localización (latitud y longitud) de forma separada para posteriormente hacer un análisis conjunto. Esto tiene una justificación matemática: si se intentase convertir la localización (un conjunto en dos dimensiones: un plano) en una lista de valores únicos para cada localización (una recta) no habría forma de hacerlo sin perder información, ya que no existe función biyectiva de  $\mathbb{R}^2$  en  $\mathbb{R}$  (por el teorema de la dimensión invariante de Giuseppe Peano). Ello nos deja con las únicas opciones de aumentar las dimensiones (lo cual complicaría lecturas como la de la gráfica de la cópula, además de que los coeficientes de correlación funcionan en parejas de variables, para más de dos dimensiones corresponden matrices de estos valores, que en el fondo es hacer las comparaciones 2 a 2) o, la otra opción es ver ambas relaciones por separado para hacer un análisis conjunto.

Para obtener resultados más estadísticamente significativos, se incluyen las edades obtenidas morfológicamente por Gómez-Calderón<sup>54</sup>, teniendo un total de 46 Conos con edad asignada.

#### 6.1.2.1 Edad vs Longitud

##### Análisis Exploratorio de datos

De observar el gráfico (Figura 6.7) es notable que las edades más antiguas se concentran de forma aparente alrededor de los  $-101.50^\circ$ . Sin embargo, hay pocos datos en dichas edades. Mientras que las edades más recientes no parecen tener alguna tendencia respecto a su longitud.

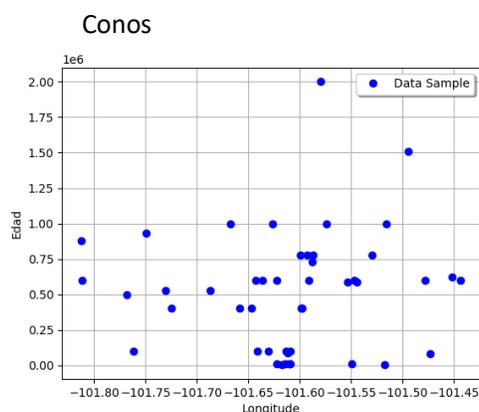


Figura 6.7| Gráfico de dispersión Edad - Longitud

Las estadísticas descriptivas de estos datos son (Tabla 6.6) :

-----	
Conos Edad vs Longitud	
-----	
Tamaño= 46	
Nombre= [Longitud,Edad]	
Mínimo= [-101.813,2218]	
Cuartil 1= [-101.643,100000]	
Promedio= [-101.612,522982]	
Mediana= [-101.61,590000]	
Cuartil 3= [-101.553,780000]	
Máximo= [-101.443,2e+06]	
Rango= [0.369467,1.99778e+06]	
Varianza= [0.00753227,1.76487e+11]	
Desviación Estándar= [0.0867887,420103]	
Skewness= [-0.393784,1.03461]	
Kurtosis= [3.24372,5.25592]	
-----	
Resumen de Estadísticas Bivariadas	
-----	
CoVarianza= [[ 0.00753227 2463.13 ]	
[ 2463.13 1.76487e+11 ]]	
Desviación Estándar= [0.0867887,420103]	
Matriz de Correlación de Pearson= [[ 1 0.0675568 ]	
[ 0.0675568 1 ]]	
Matriz de Correlación de Kendall= [[ 1 0.0798425 ]	
[ 0.0798425 1 ]]	
Matriz de Correlación de Spearman= [[ 1 0.130733 ]	
[ 0.130733 1 ]]	
-----	

Tabla 6.6 | Estadísticas de Conos vs longitud de datos originales

### Marginales

Las funciones mejor ajustadas a cada variable fueron (Figura 6.8):

$$F_{\text{Edad}} = \text{WeibullMin}(\beta = 550973, \alpha = 0.862872, \text{gamma} = 2218)$$

$$F_{\text{Longitud}} = \text{Laplace}(\mu = -101.61, \lambda = 15.9633)$$

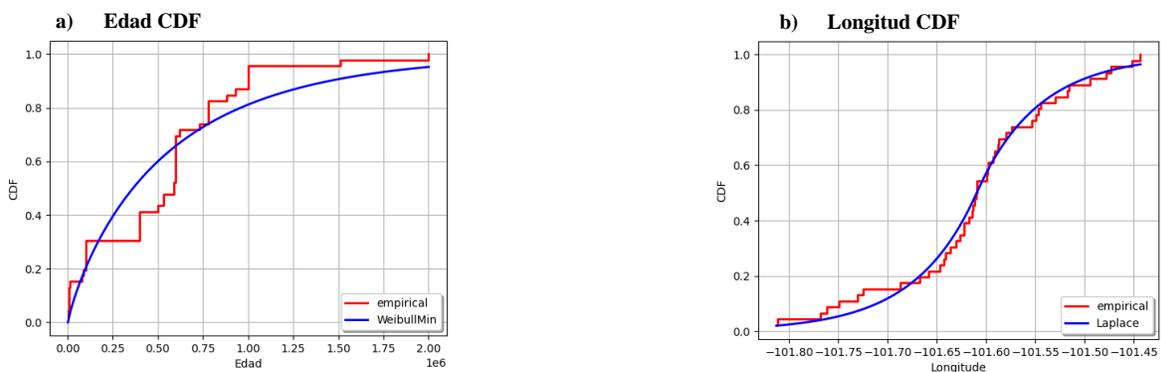


Figura 6.8 | Funciones de Distribución Acumulativa (CDF) de las variables de “Conos Edad vs Longitud”: a)Edad b)Longitud

## Cóputas

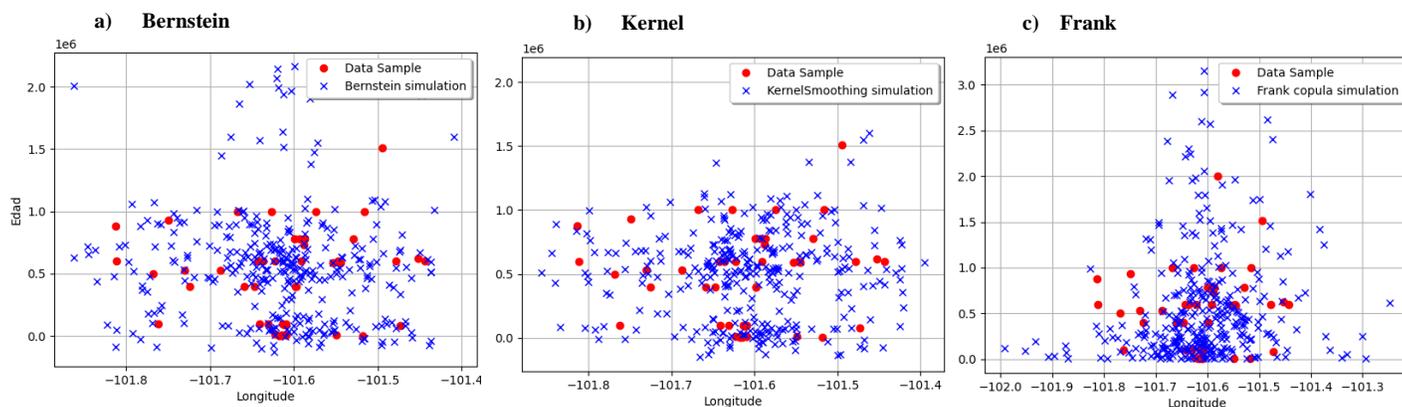
Repitiendo el procedimiento utilizado en Edad vs Índice de aspecto, se obtienen las nuevas gráficas de las cópulas, su posterior simulación y resumen estadístico. Dado que las conclusiones no parten de las gráficas sino del análisis de los resultados numéricos de la simulación proveniente de las cópulas, se muestran estos primero, buscando posteriormente añadir únicamente la gráfica y simulación respectiva de la cópula mejor ajustada como ilustración de lo obtenido cuantitativamente. Las variaciones entre cópulas quedan capturadas en el resumen estadístico de cada una (Tabla 6.7)

----- Cópula de KernelSmoothing Conos Edad vs Longitud	----- Cópula de Bernstein Conos Edad vs Longitud	----- Cópula de Frank Conos Edad vs Longitud
Tamaño= 337	Tamaño= 337	Tamaño= 337
Nombre= [Longitud,Edad]	Nombre= [Longitud,Edad]	Nombre= [Longitud,Edad]
Mínimo= [-101.857,-150334]	Mínimo= [-101.863,-131044]	Mínimo= [-101.993,2652.04]
Cuartil 1= [-101.649,167380]	Cuartil 1= [-101.659,178018]	Cuartil 1= [-101.659,129203]
Promedio= [-101.605,528978]	Promedio= [-101.616,568376]	Promedio= [-101.611,550213]
Mediana= [-101.603,536293]	Mediana= [-101.611,549432]	Mediana= [-101.611,340974]
Cuartil 3= [-101.553,731298]	Cuartil 3= [-101.563,732785]	Cuartil 3= [-101.564,728815]
Máximo= [-101.395,2.08477e+06]	Máximo= [-101.405,2.16248e+06]	Máximo= [-101.247,3.15024e+06]
Rango= [0.46183,2.23511e+06]	Rango= [0.458146,2.29352e+06]	Rango= [0.74554,3.14759e+06]
Varianza= [0.00787318,1.70198e+11]	Varianza= [0.00817501,2.12659e+11]	Varianza= [0.0097664,3.55317e+11]
Desviación Estándar= [0.088731,412551]	Desviación Estándar= [0.0904158,461149]	Desviación Estándar= [0.0988251,596084]
Skewness= [-0.332741,0.920925]	Skewness= [-0.330128,1.19917]	Skewness= [-0.108318,1.80861]
Kurtosis= [3.25217,5.00849]	Kurtosis= [3.10416,5.1275]	Kurtosis= [5.36507,6.35015]
----- Resumen de Estadísticas Bivariadas	----- Resumen de Estadísticas Bivariadas	----- Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.00787318 2693.59 ] [ 2693.59 1.70198e+11 ]]	CoVarianza= [[ 0.00817501 -2236.47 ] [ -2236.47 2.12659e+11 ]]	CoVarianza= [[ 0.0097664 7953.97 ] [ 7953.97 3.55317e+11 ]]
Desviación Estándar= [0.088731,412551]	Desviación Estándar= [0.0904158,461149]	Desviación Estándar= [0.0988251,596084]
Matriz de Correlación de Pearson= [[ 1 0.0735831 ] [ 0.0735831 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.0536385 ] [ -0.0536385 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.135023 ] [ 0.135023 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.0254698 ] [ 0.0254698 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.0973576 ] [ -0.0973576 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.116963 ] [ 0.116963 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.0381511 ] [ 0.0381511 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.139643 ] [ -0.139643 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.178811 ] [ 0.178811 1 ]]
----- pvalue=0.17778	----- pvalue=0.326237	----- theta = 0.722318

**Tabla 6.7 | Estadísticas de Conos vs longitud con la cópula de Kernel, Bernstein y Frank.**

De todo lo anterior se destaca que todas las cópulas no paramétricas tienen pvalores mayores a 0.05, además de que todos los coeficientes de correlación -los provenientes de datos originales (Tabla 6.6) y los de las cópulas- están cercanos a 0, lo cual ya indica independencia a pesar de tener coeficientes de correlación cercanos a 0.

De las cópulas no paramétricas la del p-value más pequeño es Kernel, sin embargo, dado que ambos p-value fueron mayores a 0.05, por esta ocasión se muestran las tres simulaciones para ilustrar gráficamente que un menor p-value sigue significando un mejor ajuste (Figura 6.9)



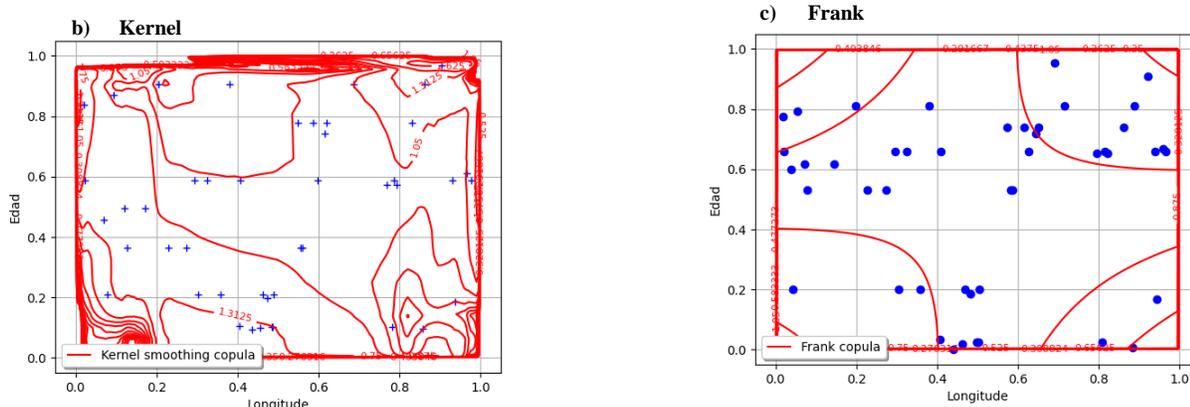
**Figura 6.9]** Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing, c) Frank en el conjunto de datos "Conos vs Longitud".

De forma que la cópula de Kernel es la mejor ajustada, ya que no genera aquellos en la región alrededor del punto (-101.6,2.0) que aparecen en Bernstein y Frank.

En cuanto a la cópula paramétrica de Frank, en vez de poseer un p-value, tiene su parámetro theta que induce dependencia de forma análoga a un coeficiente de correlación, aunque el que la cópula provenga de este parámetro ocasiona que su simulación tenga tendencia de acumulación, a comparación de la dispersión presente en las simulaciones de cópulas no paramétricas. Además, su valor theta no tiene un rango definido como los coeficientes de correlación lo tienen entre el 0 y 1, por eso mismo, un parámetro theta de 0.7 sigue considerándose cercano al 0 mostrando independencia.

Es por esto que este trabajo le dará prioridad en las siguientes comparaciones a las cópulas no paramétricas, ya que los gráficos de dispersión de la cópula de Frank se ajustan más al parámetro theta del que provienen que a los datos originales, y dicho parámetro puede ser reemplazado en el análisis por los coeficientes de correlación de las no paramétricas.

En cuanto a las gráficas de las Cópulas se muestran las de la cópula no-paramétrica mejor ajustada y la de Frank (Figura 6.10):



**Figura 6. 10|** Gráfica de la Cópula de a)Kernel (no-paramétrica) ajustada a los datos “Conos: Edad vs Longitud”;  
| b) La cópula de Frank (paramétrica) ajustada al mismo conjunto de datos.

En la gráfica de la cópula se aprecian los márgenes rectos (además del comportamiento de las curvas internas) indican independencia ya que su lectura es: entre más aumente la probabilidad de un valor de la Longitud, la Edad sigue pudiendo tomar cualquier valor. Igualmente, las curvas internas no reflejan una tendencia (o dirección) específica como aquellas de la Figura 6.5.

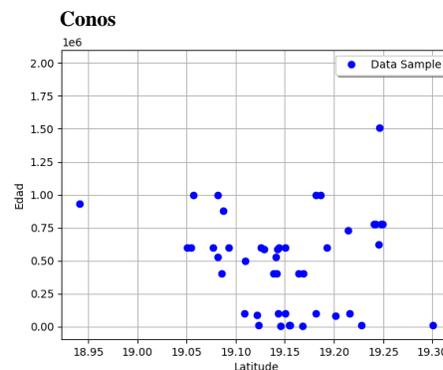
### 6.1.2.2 Edad vs Latitud

#### Análisis Exploratorio de datos

En la gráfica de dispersión de los datos (Figura 6.11) se observa:

La edad más antigua de conos está en la latitud 19.25°. Mientras que el resto de las edades (concentradas en el rango de un millón de años) aparentan moverse de una latitud menor a una mayor conforme la edad disminuye, siendo la edad más antigua la que hace intuir independencia. De ello se establece la hipótesis nula de que no hay dependencia entre Latitud y Edad, con su correspondiente Hipótesis alternativa de que sí hay dependencia.

Reiterando la noción de que las gráficas no prueban nada, son meramente ilustrativas, por tanto, para hacer una conclusión hace falta ver los datos numéricos (Tabla 6.8) y las cópulas.



**Figura 6. 11|** Gráfico de dispersión Edad - Latitud

Conos Edad vs Latitud	
Tamaño=	46
Nombre=	[ Latitud,Edad]
Mínimo=	[18.9409,2218]
Quartil 1=	[19.1099,100000]
Promedio=	[19.1534,522982]
Mediana=	[19.148,590000]
Quartil 3=	[19.2017,780000]
Máximo=	[19.3009,2e+06]
Rango=	[0.35993,1.99778e+06]
Varianza=	[0.00474041,1.76487e+11]
Desviación Estándar=	[0.0688507,420103]
Skewness=	[-0.35128,1.03461]
Kurtosis=	[3.78866,5.25592]
Resumen de Estadísticas Bivariadas	
CoVarianza=	[[ 0.00474041 981.116 ] [ 981.116 1.76487e+11 ]]
Desviación Estándar=	[0.0688507,420103]
Matriz de Correlación de Pearson=	[[ 1 0.03392 ] [ 0.03392 1 ]]
Matriz de Correlación de Kendall=	[[ 1 -0.0139724 ] [ -0.0139724 1 ]]
Matriz de Correlación de Spearman=	[[ 1 0.0182158 ] [ 0.0182158 1 ]]

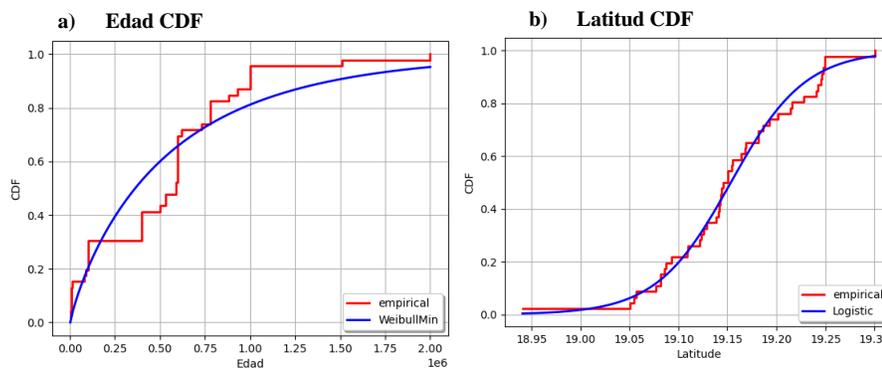
**Tabla 6.8** | Estadísticas de los Conos en Edad vs Latitud: univariado

### Marginales

Tras obtener las estadísticas descriptivas (Tabla 6.8), se ajustan las funciones marginales a cada variable (Figura 6.12) resultando:

$$F_{\text{Edad}} = \text{WeibullMin}(\beta = 550973, \alpha = 0.862872, \text{gamma} = 2218)$$

$$F_{\text{Latitud}} = \text{Logistic}(\mu = 19.1534, \beta = 0.0379594)$$



**Figura 6.12** | Funciones de Distribución Acumulativa (CDF) de las variables de "Conos Edad vs Latitud": a) Edad b) Latitud

A partir de las marginales estimadas se obtienen las cópulas y sus respectivas estadísticas.

## Cóputas

Nuevamente se obtienen las cóputas no paramétricas, se simulan los datos y obtienen las estadísticas de los datos simulados (Tabla 6.9):

----- Cópula de KernelSmoothing	----- Cópula de Bernstein
Conos Edad vs Latitud	Conos Edad vs Latitud
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Edad]	Nombre= [Latitud,Edad]
Mínimo= [18.9572,-150334]	Mínimo= [18.887,-134230]
Quartil 1= [19.1081,167380]	Quartil 1= [19.1093,160466]
Promedio= [19.1553,528978]	Promedio= [19.1545,536811]
Mediana= [19.1523,536293]	Mediana= [19.1533,580231]
Quartil 3= [19.2021,731298]	Quartil 3= [19.2096,744267]
Máximo= [19.3464,2.08477e+06]	Máximo= [19.3298,2.02577e+06]
Rango= [0.38921,2.23511e+06]	Rango= [0.442797,2.16e+06]
Varianza= [0.00479168,1.70198e+11]	Varianza= [0.00671309,1.63972e+11]
Desviación Estándar= [0.069222,412551]	Desviación Estándar= [0.0819335,404934]
Skewness= [0.00335729,0.920925]	Skewness= [-0.551232,0.645438]
Kurtosis= [2.98917,5.00849]	Kurtosis= [3.95018,4.23057]
-----	-----
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
-----	-----
CoVarianza= [[ 0.00479168 666.512 ] [ 666.512 1.70198e+11 ]]	CoVarianza= [[ 0.00671309 -3700.45 ] [ -3700.45 1.63972e+11 ]]
Desviación Estándar= [0.069222,412551]	Desviación Estándar= [0.0819335,404934]
Matriz de Correlación de Pearson= [[ 1 0.0233392 ] [ 0.0233392 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.111534 ] [ -0.111534 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0529179 ] [ -0.0529179 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.0559559 ] [ -0.0559559 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.0669731 ] [ -0.0669731 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.0760314 ] [ -0.0760314 1 ]]
----- pvalue=0.669439	----- pvalue=0.04072878

**Tabla 6. 9)** Estadísticas de los Conos en Edad vs Latitud: con cóputas no paramétricas

Se observa de las estadísticas que la cóputa mejor ajustada fue la de Bernstein, por tener un p-value menor a 0.05, sin embargo, todos sus coeficientes de correlación están cercanos a 0 (recordando que se considera cercano a 0 si el promedio de los coeficientes de correlación en valor absoluto es menor a 0.1), por tanto, no hay suficiente evidencia en contra de la Hipótesis nula y no se puede concluir que haya dependencia entre la Edad y la Latitud.

Esto queda plasmado en la forma de la gráfica, y simulación, de la cóputa mejor ajustada (Figura 6.13) que fue Bernstein en este caso:

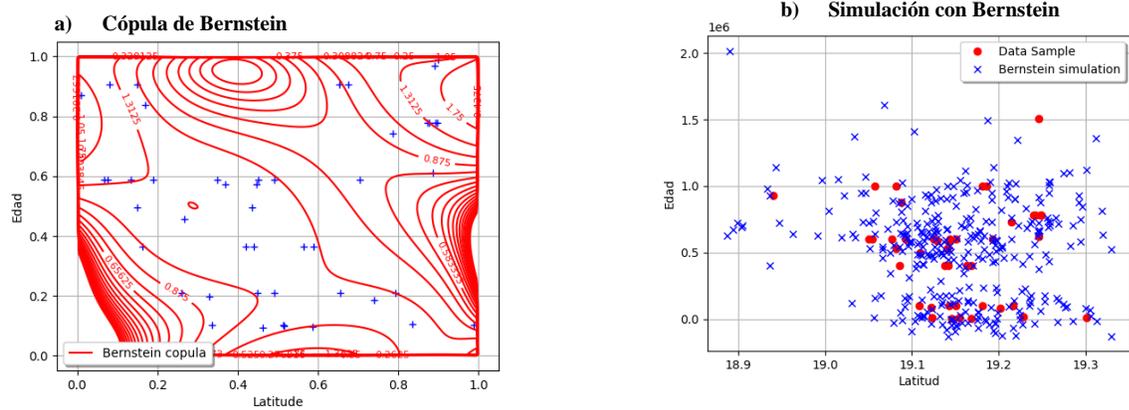


Figura 6. 13| Gráfica de: a) la Cópula de Bernstein ajustada a los datos "Conos: Edad vs Latitud"; b) Simulación respectiva

Se observa nuevamente la tendencia a la acumulación de las curvas rojas sobre los márgenes y la falta de tendencia en la mayoría de curvas internas, *indicando independencia*.

## 6.2 Spatter y Domos

Dado que las comparaciones en parejas de variables serán las mismas para estas dos tipologías (Localización vs Edad, Diámetro Basal vs Edad, Altura vs Edad) iremos haciendo el análisis de cada pareja de variables a la par en ambos grupos: Spatter y Domos. Hay 138 edades disponibles para Spatter y 22 para Domos, todas obtenidas por métodos radiométricos o químicos.

### 6.2.1 Edad vs Localización

#### 6.2.1.1 Longitud vs Edad

##### Análisis Exploratorio de datos

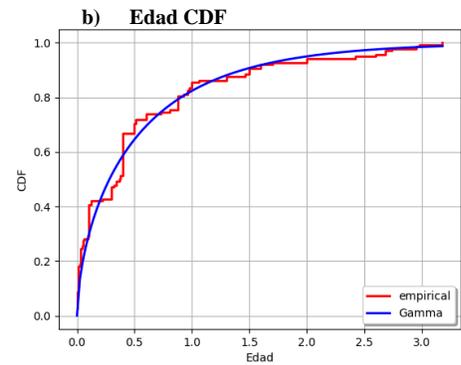
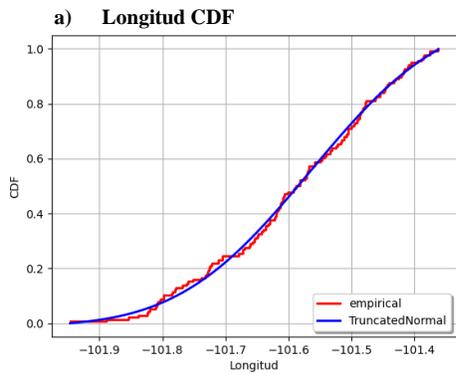
Sus resultados estadísticos iniciales son (Tabla 6.10):

Spatter: Longitud vs Edad	Domos: Longitud vs Edad
Tamaño= 138	Tamaño= 22
Nombre= [Longitud,Edad]	Nombre= [Longitud,Edad]
Mínimo= [-101.946,0.000191]	Mínimo= [-101.944,0.06]
Quartil 1= [-101.678,0.045833]	Quartil 1= [-101.572,0.14]
Promedio= [-101.594,0.549067]	Promedio= [-101.531,0.897636]
Mediana= [-101.585,0.375]	Mediana= [-101.463,0.65]
Quartil 3= [-101.487,0.81]	Quartil 3= [-101.432,1.47]
Máximo= [-101.362,3.18]	Máximo= [-101.352,2.95]
Rango= [0.584421,3.17981]	Rango= [0.591983,2.89]
Varianza= [0.0174803,0.518498]	Varianza= [0.0302989,0.718582]
Desviación Estándar= [0.132213,0.720068]	Desviación Estándar= [0.174066,0.847692]
Skewness= [-0.328015,1.95883]	Skewness= [-1.25936,1.04741]
Kurtosis= [2.31162,6.49907]	Kurtosis= [3.47913,3.33472]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.0174803 0.00346087 ] [ 0.00346087 0.518498 ]]	CoVarianza= [[ 0.0302989 0.00280119 ] [ 0.00280119 0.718582 ]]
Desviación Estándar= [0.132213,0.720068]	Desviación Estándar= [0.174066,0.847692]
Matriz de Correlación de Pearson= [[ 1 0.0363528 ] [ 0.0363528 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.0189841 ] [ 0.0189841 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.0764099 ] [ 0.0764099 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.0742386 ] [ 0.0742386 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.0948906 ] [ 0.0948906 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.0876202 ] [ 0.0876202 1 ]]

**Tabla 6.10| Estadísticas de Spatter y Domos para el conjunto de Longitud vs Edad**

### Marginales

Con las siguientes funciones marginales ajustadas para Spatter (Figura 6.14):

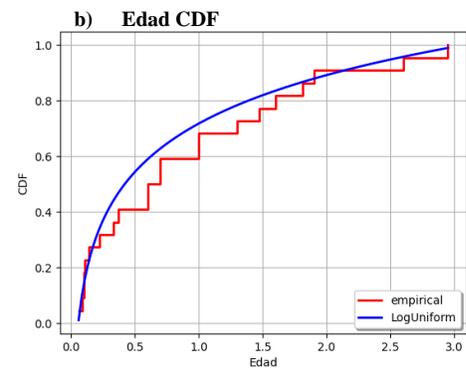
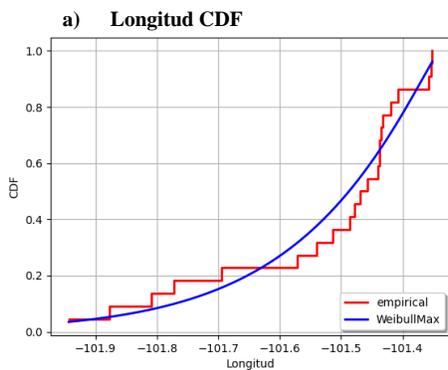


**Figura 6.14** Funciones de Distribución Acumulativa (CDF) de las variables de “Spatter Longitud vs Edad”: a) Longitud; b) Edad

$$F_{Edad} = \text{Gamma}(k = 0.581036, \lambda = 1.05859, \text{gamma} = 0.000189636)$$

$$F_{Longitud} = \text{TruncatedNormal}(\mu = -101.565, \sigma = 0.164805, a = -101.949, b = -101.36)$$

En cuanto a las de Domos (Figura 6.15):



**Figura 6.15** Funciones de Distribución Acumulativa (CDF) de las variables de “Domos Longitud vs Edad”: a) Longitud; b) Edad

$$F_{Edad} = \text{LogUniform}(a_{\text{Log}} = -2.85597, b_{\text{Log}} = 1.12263)$$

$$F_{Longitud} = \text{WeibullMax}(\beta = 0.203763, \alpha = 1.11654, \text{gamma} = -101.34)$$

De lo anterior se observa la mayor variedad en localización que tienen los Spatter y Domos a comparación de los conos. Con sus respectivas reservas ya que los datos de conos con edad son el doble de los Domos con edad, pero son un tercio de los Spatter con edad.

## Cóputas

Se obtienen las cóputas, sus simulaciones y de esta forma los resultados estadísticos de la Cópula de Kernel (Tabla 6.11):

----- Cópula de KernelSmoothing	----- Cópula de KernelSmoothing
Spatter: Longitud vs Edad	Domos: Longitud vs Edad
Tamaño= 337	Tamaño= 337
Nombre= [Longitud,Edad]	Nombre= [Longitud,Edad]
Mínimo= [-101.95,-0.093649]	Mínimo= [-102.04,-0.51693]
Quartil 1= [-101.687,0.0597728]	Quartil 1= [-101.664,0.302033]
Promedio= [-101.589,0.571098]	Promedio= [-101.543,0.96441]
Mediana= [-101.58,0.345958]	Mediana= [-101.474,0.728831]
Quartil 3= [-101.484,0.844614]	Quartil 3= [-101.407,1.582]
Máximo= [-101.306,3.25309]	Máximo= [-101.265,3.55174]
Rango= [0.64317,3.34674]	Rango= [0.774976,4.06867]
Varianza= [0.0191984,0.571709]	Varianza= [0.0360706,0.755802]
Desviación Estándar= [0.138558,0.756114]	Desviación Estándar= [0.189923,0.869369]
Skewness= [-0.233647,1.96723]	Skewness= [-0.898604,0.714935]
Kurtosis= [2.41192,6.49465]	Kurtosis= [2.69116,2.86885]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.0191984 -0.0085135 ] [ -0.0085135 0.571709 ]]	CoVarianza= [[ 0.0360706 -0.00287822 ] [ -0.00287822 0.755802 ]]
Desviación Estándar= [0.138558,0.756114]	Desviación Estándar= [0.189923,0.869369]
Matriz de Correlación de Pearson= [[ 1 -0.0812623 ] [ -0.0812623 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.0174318 ] [ -0.0174318 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.0340893 ] [ 0.0340893 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.00597004 ] [ -0.00597004 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.0440169 ] [ 0.0440169 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.00746512 ] [ -0.00746512 1 ]]
pvalue=0.136567	pvalue=0.7498469

**Tabla 6.11| Estadísticas de Spatter y Domos con la cóputa de Kernel ajustada a datos de Longitud-Edad**

Los coeficientes de correlación son cercanos a 0, cayendo todos en una vecindad de 0.045 de radio, y los p-values mayores a 0.05. En cuanto a la Cópula de Bernstein los resultados estadísticos (Tabla 6.12) son:

----- Cópula de Bernstein	----- Cópula de Bernstein
Spatter: Longitud vs Edad	Domos: Longitud vs Edad
Tamaño= 337	Tamaño= 337
Nombre= [Longitud,Edad]	Nombre= [Longitud,Edad]
Mínimo= [-101.981,-0.0931439]	Mínimo= [-102.051,-0.463172]
Quartil 1= [-101.674,0.0703902]	Quartil 1= [-101.634,0.279038]
Promedio= [-101.583,0.574456]	Promedio= [-101.54,0.958489]
Mediana= [-101.564,0.332127]	Mediana= [-101.483,0.726858]
Quartil 3= [-101.474,0.854906]	Quartil 3= [-101.412,1.42431]
Máximo= [-101.291,3.23917]	Máximo= [-101.252,3.57822]
Rango= [0.689838,3.33232]	Rango= [0.79937,4.04139]
Varianza= [0.0200984,0.568266]	Varianza= [0.0329157,0.822768]
Desviación Estándar= [0.141769,0.753834]	Desviación Estándar= [0.181427,0.907066]
Skewness= [-0.425181,1.90458]	Skewness= [-0.97026,0.90494]
Kurtosis= [2.604,6.03753]	Kurtosis= [2.999,3.19777]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.0200984 -0.00425471 ] [ -0.00425471 0.568266 ]]	CoVarianza= [[ 0.0329157 -0.0132823 ] [ -0.0132823 0.822768 ]]
Desviación Estándar= [0.141769,0.753834]	Desviación Estándar= [0.181427,0.907066]
Matriz de Correlación de Pearson= [[ 1 -0.039812 ] [ -0.039812 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.0807108 ] [ -0.0807108 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.0315459 ] [ 0.0315459 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.0973576 ] [ -0.0973576 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.0406432 ] [ 0.0406432 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.139643 ] [ -0.139643 1 ]]
pvalue=0.46635431	pvalue=0.139257

**Tabla 6.12| Estadísticas de Spatter y Domos con la cóputa de Bernstein ajustada a datos de Longitud-Edad**

Donde vuelven a promediar cercano a 0 los coeficientes de correlación. Tomamos los p-valores más pequeños como las cópulas no paramétricas mejor ajustadas de forma que para Spatter es la cópula de Kernel y para Domos la de Frank.

La gráfica de la cópula no paramétrica mejor ajustada y simulación respectiva para Spatter: (Figura 6.16)

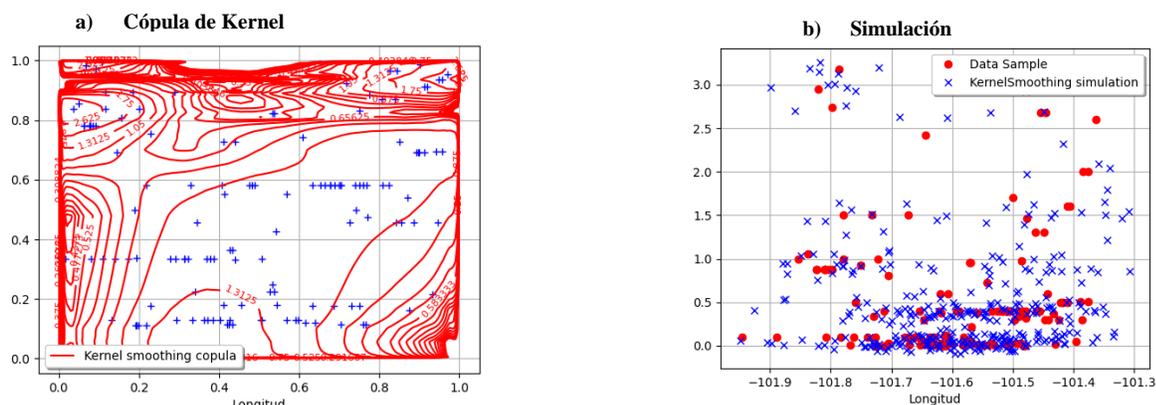


Figura 6.16| Spatter Edad vs Longitud". a) Gráfica de la cópula de Kernel; b) Datos originales contra los simulados con la Cópula de Kernel Smoothing

Las simulaciones muestran una tendencia a la edad 0. La gráfica de la cópula vuelve a mostrar una acumulación de curvas en los márgenes, sin una tendencia clara en las curvas internas, indicando un rechazo a la dependencia entre las variables.

Para domos, al momento de generar la gráfica de la cópula de Bernstein la estimación del polinomio genera valores fuera de los márgenes preestablecidos y la versión de la librería openturns utilizada (1.20) no re-ajusta los márgenes de forma que no imprime las curvas rojas, por lo que añadimos las otras dos gráficas (Figura 6.17) y las simulaciones respectivas (Figura 6.18):

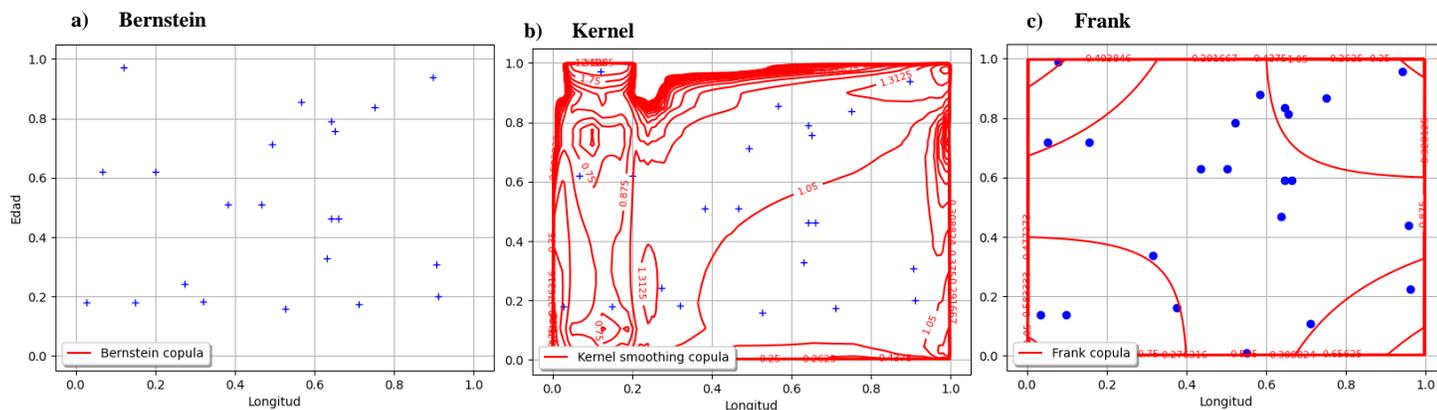
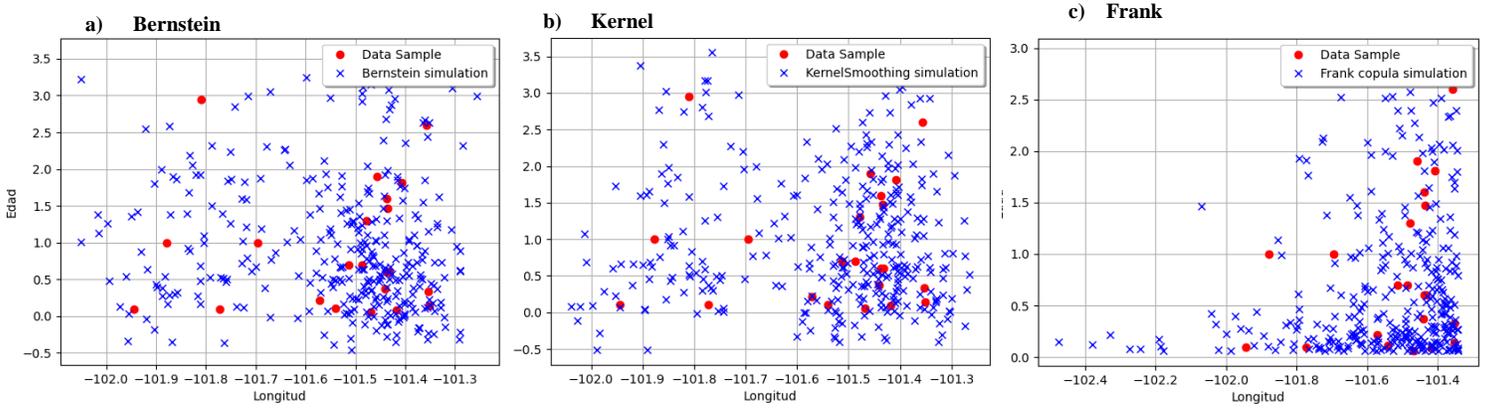


Figura 6.17 | Gráfica de las Cópulas de: a) Bernstein; b) Kernel; c) Frank; en el conjunto "Domos Longitud vs Edad"



**Figura 6.18** Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing c) Frank en el conjunto de datos "Domos Edad vs Longitud".

Nótese que, al cambiar la función ajustada para la Edad, no parece limitar la generación de Edades negativas en la simulación en las cópulas no paramétricas (Figura 6.18). Además, que en las gráficas de las cópulas se muestra la ausencia de dependencia por las mismas razones que en la sección 6.1.2

En resumen, *no existe dependencia* entre la Longitud y la Edad tanto en Domos como en Spatter.

### 6.2.1.2 Latitud vs Edad

#### Análisis Exploratorio de datos

Los resultados estadísticos son (Tabla 6.13):

Spatter: Latitud vs Edad	Domos: Latitud vs Edad
Tamaño= 138	Tamaño= 22
Nombre= [Latitud,Edad]	Nombre= [Latitud,Edad]
Mínimo= [18.8798,0.000191]	Mínimo= [18.9898,0.06]
Quartil 1= [19.0476,0.045833]	Quartil 1= [19.1097,0.14]
Promedio= [19.1455,0.549067]	Promedio= [19.292,0.897636]
Mediana= [19.1079,0.375]	Mediana= [19.3082,0.65]
Quartil 3= [19.1993,0.81]	Quartil 3= [19.4515,1.47]
Máximo= [19.548,3.18]	Máximo= [19.5265,2.95]
Rango= [0.668256,3.17981]	Rango= [0.536741,2.89]
Varianza= [0.0236126,0.518498]	Varianza= [0.0333661,0.718582]
Desviación Estándar= [0.153664,0.720068]	Desviación Estándar= [0.182664,0.847692]
Skewness= [0.95885,1.95883]	Skewness= [-0.228145,1.04741]
Kurtosis= [3.31652,6.49907]	Kurtosis= [1.43776,3.33472]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.0236126 0.0194013 ] [ 0.0194013 0.518498 ]]	CoVarianza= [[ 0.0333661 0.0124418 ] [ 0.0124418 0.718582 ]]
Desviación Estándar= [0.153664,0.720068]	Desviación Estándar= [0.182664,0.847692]
Matriz de Correlación de Pearson= [[ 1 0.175341 ] [ 0.175341 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.0803516 ] [ 0.0803516 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.011939 ] [ -0.011939 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.14411 ] [ 0.14411 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.00506547 ] [ 0.00506547 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.186546 ] [ 0.186546 1 ]]

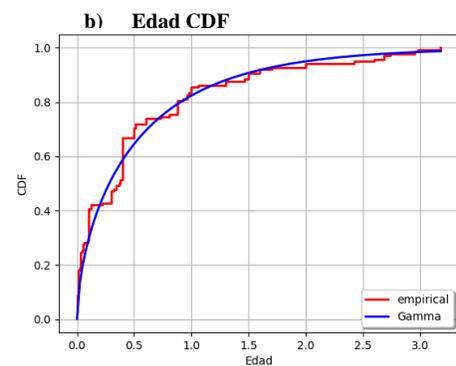
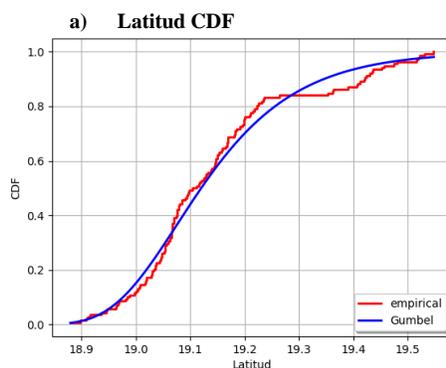
**Tabla 6.13| Estadísticas de Spatter y Domos para el conjunto de Latitud vs Edad**

#### Marginales

Las funciones univariadas ajustadas a los Spatter (Figura 6.19):

$$F_{Edad} = \text{Gamma}(k = 0.581036, \lambda = 1.05859, \text{gamma} = 0.000189636)$$

$$F_{Latitud} = \text{Gumbel}(\beta = 0.119811, \text{gamma} = 19.0764)$$



**Figura 6.19| Funciones de Distribución Acumulativa (CDF) de las variables de “Spatter Latitud vs Edad”: a) Latitud; b) Edad**

Mientras que las de los Domos (Figura 6.20) fueron:

$$F_{\text{Edad}} = \text{LogUniform}(a_{\text{Log}} = -2.85597, b_{\text{Log}} = 1.12263)$$

$$F_{\text{Latitud}} = \text{Uniform}(a = 18.9674, b = 19.5489)$$

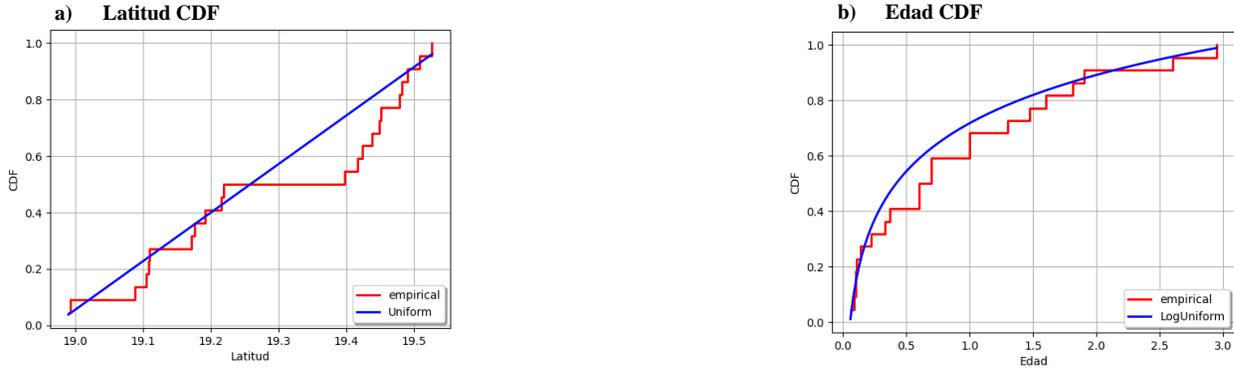


Figura 6.20| Funciones de Distribución Acumulativa (CDF) de las variables de “Domos Latitud vs Edad”: a) Latitud; b) Edad

De las dos gráficas para edades se observa cómo el cambio de 22 a 138 datos afecta la función que se le ajusta a la Edad, sin embargo, no parece cambiar drásticamente la forma general de la curva ajustada.

### Cóputas

Posteriormente se obtienen las cópulas, se simulan los datos y las estadísticas descriptivas de dichas simulaciones, primero los resultados provenientes de la Cópula de Kernel (Tabla 6.14):

----- Cópula de KernelSmoothing Spatter: Latitud vs Edad -----	----- Cópula de KernelSmoothing Domos: Latitud vs Edad -----
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Edad]	Nombre= [Latitud,Edad]
Mínimo= [18.8272,-0.093649]	Mínimo= [18.9168,-0.51693]
Quartil 1= [19.0338,0.0597728]	Quartil 1= [19.1282,0.302033]
Promedio= [19.1391,0.571098]	Promedio= [19.2921,0.96441]
Mediana= [19.1073,0.345958]	Mediana= [19.3083,0.728831]
Quartil 3= [19.2008,0.844614]	Quartil 3= [19.4703,1.582]
Máximo= [19.6046,3.25309]	Máximo= [19.5873,3.55174]
Rango= [0.777438,3.34674]	Rango= [0.67045,4.06867]
Varianza= [0.0250123,0.571709]	Varianza= [0.035425,0.755802]
Desviación Estándar= [0.158153,0.756114]	Desviación Estándar= [0.188215,0.869369]
Skewness= [0.91802,1.96723]	Skewness= [-0.134278,0.714935]
Kurtosis= [3.40678,6.49465]	Kurtosis= [1.57682,2.86885]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 0.0250123 0.00949435 ] [ 0.00949435 0.571709 ]]	CoVarianza= [[ 0.035425 -0.00199199 ] [ -0.00199199 0.755802 ]]
Desviación Estándar= [0.158153,0.756114]	Desviación Estándar= [0.188215,0.869369]
Matriz de Correlación de Pearson= [[ 1 0.0793964 ] [ 0.0793964 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.0121739 ] [ -0.0121739 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.00777165 ] [ 0.00777165 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.00219019 ] [ -0.00219019 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.0184927 ] [ 0.0184927 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.00539386 ] [ -0.00539386 1 ]]
----- pvalue=0.145834	----- pvalue=0.823800

Tabla 6.14| Estadísticas de Spatter y Domos con la cópula de Kernel ajustada a datos de Latitud-Edad

De donde conviene observar cómo todos los coeficientes de correlación son muy cercanos a 0 (basta con una vecindad de 0.08 de radio para incluirlos a todos), mientras que, por otro lado, los p-values no alcanzan a ser significativos por ser mayores a 0.05. Esto se repite en la cópula de Bernstein (Tabla 6.15):

----- Cópula de Bernstein	----- Cópula de Bernstein
Spatter: Latitud vs Edad	Domos: Latitud vs Edad
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Edad]	Nombre= [Latitud,Edad]
Mínimo= [18.8372,-0.0774917]	Mínimo= [18.9013,-0.557193]
Quartil 1= [19.0373,0.0507808]	Quartil 1= [19.1252,0.216482]
Promedio= [19.1462,0.541847]	Promedio= [19.2782,0.911606]
Mediana= [19.1152,0.316808]	Mediana= [19.2501,0.689225]
Quartil 3= [19.2167,0.796964]	Quartil 3= [19.4497,1.45499]
Máximo= [19.5913,3.24721]	Máximo= [19.6155,3.58914]
Rango= [0.754102,3.3247]	Rango= [0.714187,4.14633]
Varianza= [0.0248432,0.544399]	Varianza= [0.0334273,0.83297]
Desviación Estándar= [0.157617,0.737834]	Desviación Estándar= [0.182831,0.912672]
Skewness= [0.799861,1.97672]	Skewness= [-0.0881095,0.880521]
Kurtosis= [3.10979,6.39647]	Kurtosis= [1.7301,3.16296]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.0248432 -0.00266182 ] [ -0.00266182 0.544399 ]]	CoVarianza= [[ 0.0334273 0.0157184 ] [ 0.0157184 0.83297 ]]
Desviación Estándar= [0.157617,0.737834]	Desviación Estándar= [0.182831,0.912672]
Matriz de Correlación de Pearson= [[ 1 -0.0228885 ] [ -0.0228885 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.0941981 ] [ 0.0941981 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0936131 ] [ -0.0936131 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.0807546 ] [ 0.0807546 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.138069 ] [ -0.138069 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.125596 ] [ 0.125596 1 ]]
pvalue=0.675457	pvalue=0.084227

**Tabla 6.15| Estadísticas de Spatter y Domos con la cópula de Bernstein ajustada a datos de Latitud-Edad**

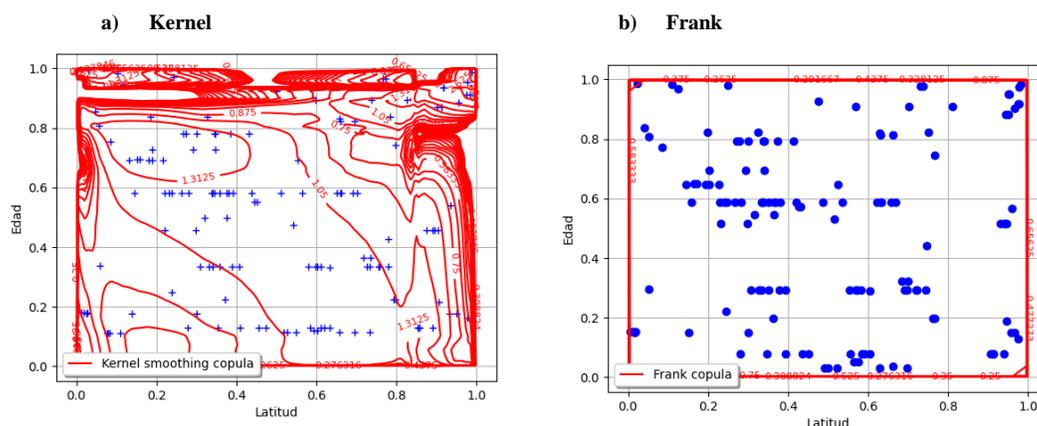
De forma que no hay dependencia por tener p-values mayores al 0.05, y al tomar los menores de cada conjunto de datos vemos que la cópula no paramétrica mejor ajustada para los Spatter fue Kernel mientras que para los Domos fue Bernstein, al no tener dependencia (o no haber suficiente evidencia en contra de la hipótesis nula) habremos de recurrir a los resultados estadísticos de la cópula de Frank (Tabla 6.16), sólo para confirmar lo que inducen estos resultados numéricos.

----- Cópula de Frank Spatter: Latitud vs Edad	----- Cópula de Frank Domos: Latitud vs Edad
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Edad]	Nombre= [Latitud,Edad]
Mínimo= [18.8466,0.000198679]	Mínimo= [18.968,0.0581814]
Quartil 1= [19.029,0.0885948]	Quartil 1= [19.099,0.154588]
Promedio= [19.148,0.498466]	Promedio= [19.256,0.720621]
Mediana= [19.1179,0.261352]	Mediana= [19.254,0.381276]
Quartil 3= [19.2306,0.657442]	Quartil 3= [19.4086,1.02414]
Máximo= [19.8526,3.26843]	Máximo= [19.548,2.93519]
Rango= [1.00603,3.26823]	Rango= [0.579935,2.87701]
Varianza= [0.0278165,0.386037]	Varianza= [0.0305798,0.563308]
Desviación Estándar= [0.166783,0.621319]	Desviación Estándar= [0.174871,0.750539]
Skewness= [1.12181,2.015]	Skewness= [0.0378436,1.32769]
Kurtosis= [4.77678,7.19998]	Kurtosis= [1.72183,3.76834]
----- Resumen de Estadísticas Bivariadas	----- Resumen de Estadísticas Bivariadas
-----	-----
CoVarianza= [[ 0.0278165 -0.00151437 ] [ -0.00151437 0.386037 ]]	CoVarianza= [[ 0.0305798 0.0283703 ] [ 0.0283703 0.563308 ]]
Desviación Estándar= [0.166783,0.621319]	Desviación Estándar= [0.174871,0.750539]
Matriz de Correlación de Pearson= [[ 1 -0.0146139 ] [ -0.0146139 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.216159 ] [ 0.216159 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.0187933 ] [ 0.0187933 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.18433 ] [ 0.18433 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.0322095 ] [ 0.0322095 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.276451 ] [ 0.276451 1 ]]
-----theta = -0.107	-----theta = 1.3193

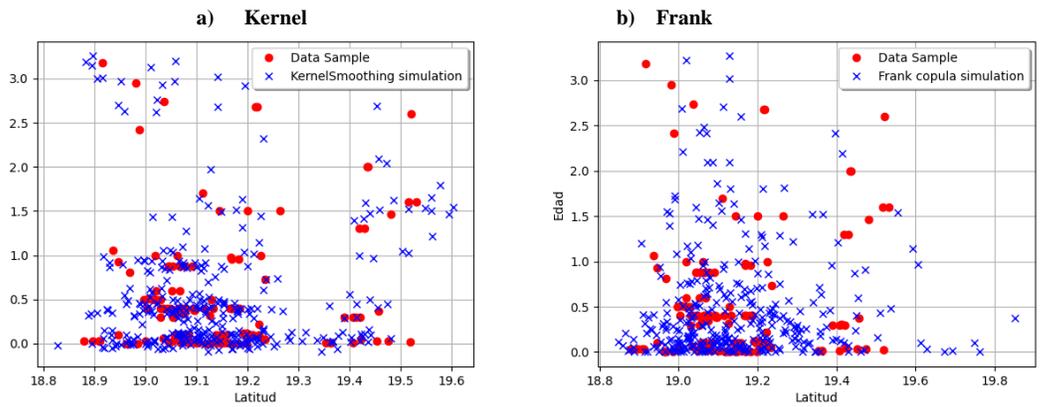
**Tabla 6.16| Estadísticas de Spatter y Domos con la cópula de Frank ajustada a datos de Latitud-Edad**

Con su parámetro (theta) que proviene de las estadísticas univariadas, induce dependencia y se puede ver que en Spatter el valor vuelve a rondar el 0, mientras que en Domos alcanza 1.31. Valor que no resulta suficiente para asegurar dependencia, ya que el parámetro de la cópula no está en el rango de -1 a 1, sino en todos los reales y por ende un valor de 0.1 y 1.3 son suficientemente cercanos a 0 para indicar independencia.

Por último, mostramos la gráfica y simulación de la cópula no paramétrica mejor ajustada para cada grupo: Spatter (Figura 6.21 y 6.22) y Domos (Figura 6.23), junto con la de Frank en los Spatter por tener, en esta ocasión, una particularidad en su gráfica que conviene resaltar.



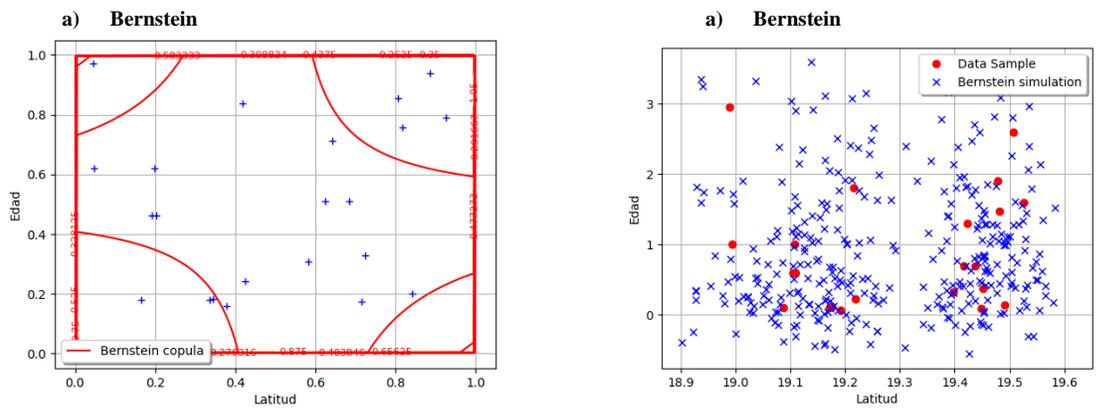
**Figura 6.21| Gráfica de las Cópulas de: a) Bernstein; b) Kernel; c) Frank; en el conjunto "Spatter Latitud vs Edad"**



**Figura 6.22** Resultado de los datos originales contra los simulados con la Cópula de: a) Kernel [Smoothing b) Frank en el conjunto de datos “Spatter Edad vs Latitud”.

En la gráfica de las cópulas se vuelve a apreciar la tendencia de acumulación en los márgenes, lo cual resalta en particular en la de Frank que muestra una contundente ausencia de dependencia.

Veamos ahora el caso de los Domos:



**Figura 6.23** Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing c) Frank en el conjunto de datos “Domos Edad vs Latitud”.

Donde se mantiene la acumulación en los márgenes para la gráfica de la cópula, indicando independencia, y se refleja en lo disperso de los datos simulados.

Por lo anterior, en general, *no se observa dependencia* entre la localización y la Edad en ninguna de las tres tipologías de nuestra área de estudio.

## 6.2.2 Altura vs Edad

En la presente sección se analizan la altura y posteriormente (Sección 6.2.3) el diámetro basal para buscar formas morfométricas de deducir la edad de un Spatter o Domo. De los datos disponibles: 133 tienen alturas registradas de los Spatter, mientras que 21 de los Domos tienen altura registrada.

### Análisis Exploratorio de datos

Los resultados estadísticos de estos datos son (Tabla 6.17):

----- Spatter: Altura vs Edad -----	----- Domos: Altura vs Edad -----
Tamaño= 133	Tamaño= 21
Nombre= [Altura,Edad]	Nombre= [Altura,Edad]
Mínimo= [19,0,000191]	Mínimo= [91,0,06]
Quartil 1= [91,5,0,031329]	Quartil 1= [193,5,0,1325]
Promedio= [149,158,0,513618]	Promedio= [303,857,0,870381]
Mediana= [122,0,34]	Mediana= [289,0,6]
Quartil 3= [178,0,6325]	Quartil 3= [368,1,375]
Máximo= [453,3,18]	Máximo= [662,2,95]
Rango= [434,3,17981]	Rango= [571,2,89]
Varianza= [7918,06,0,466801]	Varianza= [21332,0,737351]
Desviación Estándar= [88,9835,0,683229]	Desviación Estándar= [146,055,0,858691]
Skewness= [1,2175,2,10729]	Skewness= [0,871085,1,15203]
Kurtosis= [4,12439,7,47208]	Kurtosis= [3,51929,3,50936]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
-----	-----
CoVarianza= [[ 7918.06   -3.15427 ] [ -3.15427   0.466801 ]]	CoVarianza= [[ 21332   -32.2634 ] [ -32.2634   0.737351 ]]
Desviación Estándar= [88.9835,0.683229]	Desviación Estándar= [146.055,0.858691]
Matriz de Correlación de Pearson= [[ 1   -0.0518828 ] [ -0.0518828   1   ]]	Matriz de Correlación de Pearson= [[ 1   -0.257251 ] [ -0.257251   1   ]]
Matriz de Correlación de Kendall= [[ 1   -0.0503672 ] [ -0.0503672   1   ]]	Matriz de Correlación de Kendall= [[ 1   -0.120485 ] [ -0.120485   1   ]]
Matriz de Correlación de Spearman= [[ 1   -0.0724662 ] [ -0.0724662   1   ]]	Matriz de Correlación de Spearman= [[ 1   -0.168781 ] [ -0.168781   1   ]]
-----	-----

**Tabla 6.17| Estadísticas de Spatter y Domos para el conjunto de Altura vs Edad**

### Marginales

Cuyas funciones mejor ajustadas para los Spatter fueron (Figura 6.24):



**Figura 6.24| Funciones de Distribución Acumulativa (CDF) de las variables de "Spatter Altura vs Edad": a) Altura; b)Edad**

$$F_{\text{Edad}} = \text{Gamma}(k = 0.564713, \lambda = 1.09989, \text{gamma} = 0.000189585)$$

$$F_{\text{Altura}} = \text{LogNormal}(\mu\text{-Log} = 5.06255, \sigma\text{-Log} = 0.479032, \text{gamma} = -28.02)$$

Y las de los Domos son (Figura 6.25):



Figura 6.25| Funciones de Distribución Acumulativa (CDF) de las variables de “Domo Altura vs Edad”: a) Altura; b)Edad

$$F_{\text{Edad}} = \text{LogUniform}(a\text{Log} = -2.85786, b\text{Log} = 1.12436)$$

$$F_{\text{Altura}} = \text{InverseNormal}(\mu = 303.857, \lambda = 1168.78)$$

### Cóputas

Siguiendo la metodología se obtienen las cóputas, sus simulaciones y resultados estadísticos, aquellos de la cóputa Kernel son (Tabla 6.18):

----- Cópula de KernelSmoothing Spatter: Altura vs Edad ----- Tamaño= 337	----- Cópula de KernelSmoothing Domos: Altura vs Edad ----- Tamaño= 337
Nombre= [Altura,Edad]	Nombre= [Altura,Edad]
Mínimo= [-5.69657,-0.0920115]	Mínimo= [-35.3486,-0.487382]
Quartil 1= [90.6646,0.0696814]	Quartil 1= [213.718,0.235762]
Promedio= [159.437,0.524352]	Promedio= [310.513,0.826031]
Mediana= [133.004,0.341074]	Mediana= [294.463,0.595937]
Quartil 3= [195.314,0.667847]	Quartil 3= [393.961,1.1973]
Máximo= [487.347,3.13911]	Máximo= [782.071,3.54876]
Rango= [493.044,3.23112]	Rango= [817.42,4.03614]
Varianza= [9953.77,0.470613]	Varianza= [24013.4,0.747355]
Desviación Estándar= [99.7686,0.686012]	Desviación Estándar= [154.963,0.864497]
Skewness= [1.1275,2.091]	Skewness= [0.474823,1.1889]
Kurtosis= [3.73031,7.2528]	Kurtosis= [3.25943,3.89802]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 9953.77 -2.51796 ] [ -2.51796 0.470613 ]]	CoVarianza= [[ 24013.4 -25.6439 ] [ -25.6439 0.747355 ]]
Desviación Estándar= [99.7686,0.686012]	Desviación Estándar= [154.963,0.864497]
Matriz de Correlación de Pearson= [[ 1 -0.0367894 ] [ -0.0367894 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.191423 ] [ -0.191423 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0117635 ] [ -0.0117635 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.140172 ] [ -0.140172 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.0182854 ] [ -0.0182854 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.212739 ] [ -0.212739 1 ]]
----- pvalue=0.50089	----- pvalue=0.00040973

Tabla 6.18| Estadísticas de Spatter y Domos con la cóputa de Kernel Smoothing para el conjunto de Altura vs Edad

En la cópula de Kernel para el conjunto de datos de Domos tiene un p-value significativo y sus coeficientes de correlación son negativos, así como lejanos a 0 (menores a -0.1).

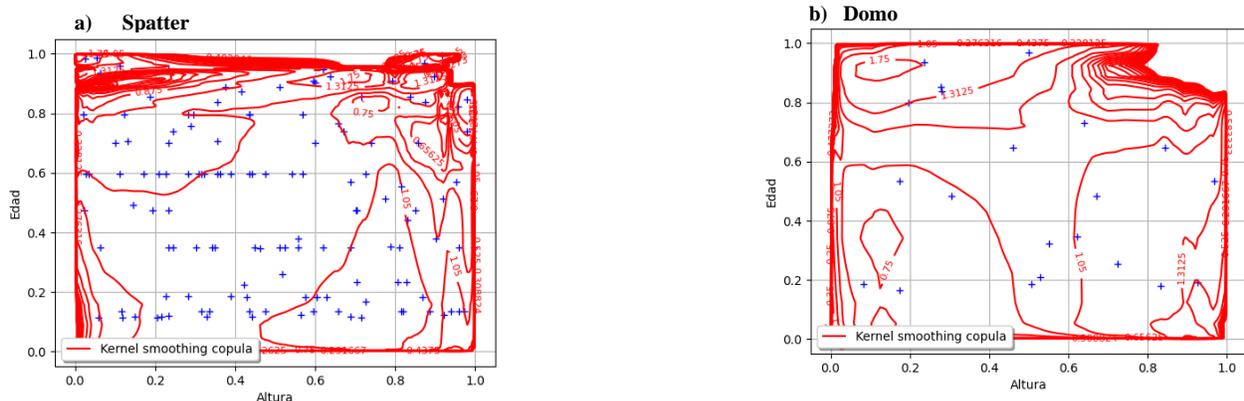
Con Bernstein (Tabla 6.19) las estadísticas resultan:

----- Cópula de Bernstein Spatter: Altura vs Edad -----	----- Cópula de Bernstein Domos: Altura vs Edad -----
Tamaño= 337	Tamaño= 337
Nombre= [Altura,Edad]	Nombre= [Altura,Edad]
Mínimo= [-12.5489,-0.102452]	Mínimo= [-81.6614,-0.46197]
Quartil 1= [88.2537,0.0667522]	Quartil 1= [176.262,0.257286]
Promedio= [148.155,0.547615]	Promedio= [295.844,0.929585]
Mediana= [126.001,0.334918]	Mediana= [277.476,0.673164]
Quartil 3= [178.549,0.845305]	Quartil 3= [393.996,1.36192]
Máximo= [480.727,3.26542]	Máximo= [808.089,3.57486]
Rango= [493.276,3.36787]	Rango= [889.75,4.03683]
Varianza= [8334.01,0.524219]	Varianza= [27503.2,0.84032]
Desviación Estándar= [91.2908,0.724029]	Desviación Estándar= [165.841,0.91669]
Skewness= [1.25105,2.06134]	Skewness= [0.533847,0.980283]
Kurtosis= [4.26626,7.17427]	Kurtosis= [3.07993,3.2723]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 8334.01 1.63101 ] [ 1.63101 0.524219 ]]	CoVarianza= [[ 27503.2 -10.8817 ] [ -10.8817 0.84032 ]]
Desviación Estándar= [91.2908,0.724029]	Desviación Estándar= [165.841,0.91669]
Matriz de Correlación de Pearson= [[ 1 0.0246759 ] [ 0.0246759 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.0715784 ] [ -0.0715784 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0221492 ] [ -0.0221492 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.0973576 ] [ -0.0973576 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.0365709 ] [ -0.0365709 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.139643 ] [ -0.139643 1 ]]
----- pvalue=0.651719	----- pvalue=0.18992

**Tabla 6.19| Estadísticas de Spatter y Domos con la cópula de Bernstein para el conjunto de Altura vs Edad**

Para ambos conjuntos (Spatter y Domos) Bernstein no parece tener un buen ajuste dado su p-value mayor a 0.05. Sin embargo, para Domos, los coeficientes de correlación siguen siendo negativos y rondan el límite de 0.1 que se estableció para considerarlos cercanos al 0. Por otro lado, para los Spatter los coeficientes de correlación caen en una vecindad de 0.04 sin importar la cópula aplicada. Dada la experiencia previa en el análisis de Localización vs Edad, podemos deducir que esto es suficiente para afirmar que en Spatter *no hay evidencia de dependencia* entre Altura y Edad.

Para ambos conjuntos (Domos y Spatter) la cópula de Kernel fue la mejor ajustada. (Figura 6.26)



**Figura 6.26| Grafica de la cópula de Kernel en los conjuntos: a) Spatter b) Domo; en "Altura vs Edad"**

Con sus respectivas simulaciones (Figura 6.27):

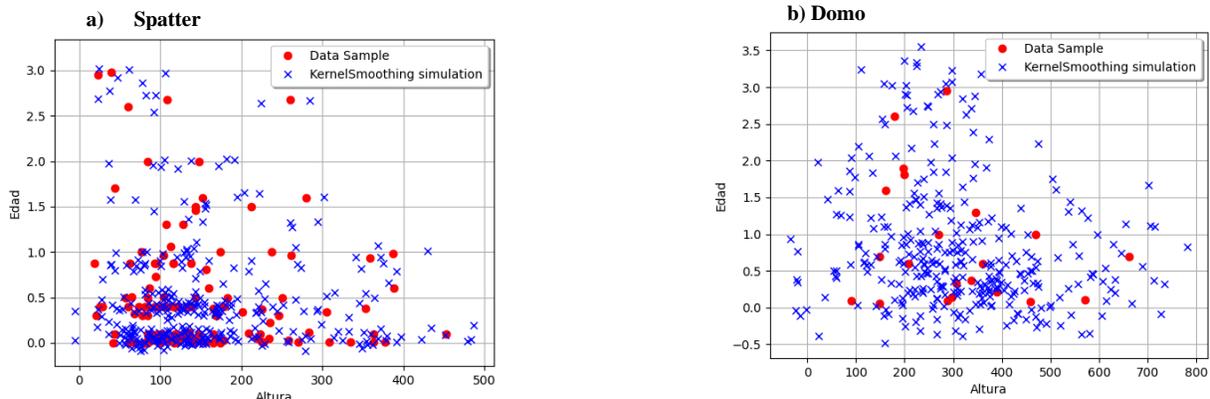


Figura 6.27| Resultado de los datos originales contra los simulados con la Cópula de Kernel Smoothing en el conjunto de : a)Spatter, |b)Domos; en “ Edad vs Longitud”.

La simulación resultante en Domo resulta desalentadora, pareciendo ser más aleatorios los datos de lo que se esperaba. Esto probablemente se deba a los pocos datos, para apreciar dicha aleatoriedad basta ver la gráfica de la cópula (Figura 6.27).

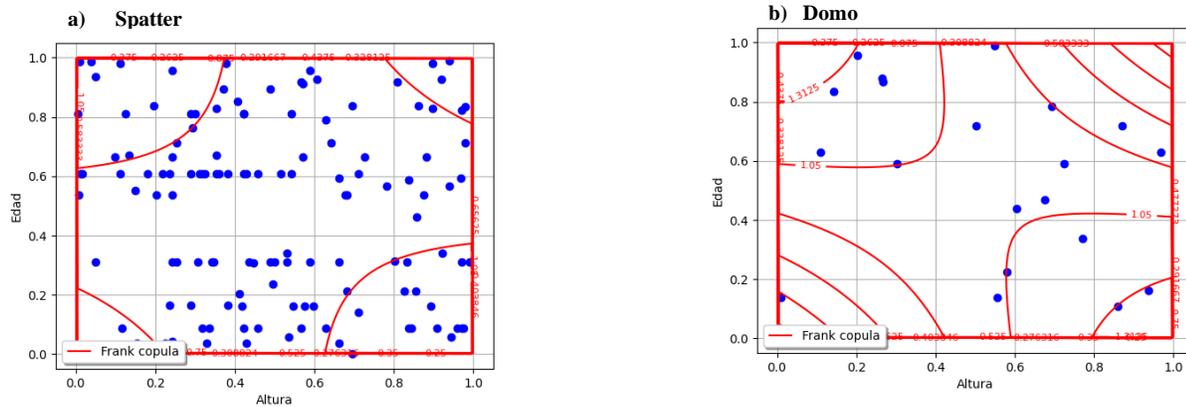
Al observar las gráficas de las cópulas se encuentra una zona de evidente diferencia entre la de Spatter y la de Domo alrededor del (1,1), y puede que sea dicha zona la que llevó a las diferencias en los resultados numéricos que señalan que en Domo hay dependencia negativa entre la Altura y la Edad, más no para los Spatter.

Cómo se encontró dependencia, se muestra la expresión de la cópula que relaciona Altura y Edad para Domo :

$$\text{Kernel}(Fx, Fy) = \left( \frac{3}{1.3802 \cdot 10^{11}} \right) \left( 1 - \left( \frac{Fx}{185.757} \right)^2 - \left( \frac{Fy}{185.757} \right)^2 \right)$$

Con Bandwith = 185.757

Si se busca reafirmar esto, es posible apoyarse de las gráficas de las cópulas de Frank (Figura 6.28) que por ser paramétricas son más suavizadas y más sencillas para leer.



**Figura 6.28** | Gráfica de la cópula de Frank en los conjuntos: a) Spatter b) Domo; en “Altura vs Edad”

Aquí se aprecia la diferencia entre las tipologías, donde las curvas internas de los domos tienen más marcada una tendencia de pendiente negativa que la de los Spatter.

### 6.2.3 Diámetro Basal vs Edad

Ahora buscaremos posible dependencia entre el Diámetro basal, considerado como el promedio de los diámetros máximo y mínimo de la base de una estructura (y que abreviaremos como DiamBasal), y la Edad radiométrica de Spatter y Domos. De los datos disponibles, 126 Spatter y 19 Domos tienen información suficiente para este análisis

#### Análisis Exploratorio de datos

Las estadísticas descriptivas respectivas son: (Tabla 6.20)

Spatter: Diámetro Basal vs Edad	Domos Diámetro Basal vs Edad
Tamaño= 126	Tamaño= 19
Nombre= [DiamBasal,Edad]	Nombre= [DiamBasal,Edad]
Mínimo= [313,0.000191]	Mínimo= [678,0.06]
Quartil 1= [562.5,0.045833]	Quartil 1= [1079.25,0.1375]
Promedio= [852.46,0.543563]	Promedio= [1701.24,0.854632]
Mediana= [814.25,0.355]	Mediana= [1528,0.6]
Quartil 3= [1089.5,0.81]	Quartil 3= [2224.62,1.225]
Máximo= [3085,3.18]	Máximo= [4176,2.95]
Rango= [2772,3.17981]	Rango= [3498,2.89]
Varianza= [151600,0.511288]	Varianza= [725356,0.730485]
Desviación Estándar= [389.358,0.715044]	Desviación Estándar= [851.678,0.854684]
Skewness= [1.89193,1.97687]	Skewness= [1.41631,1.30466]
Kurtosis= [10.6509,6.6441]	Kurtosis= [5.65805,4.04339]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 151600    52.8 ] [ 52.8    0.511288 ]]	CoVarianza= [[ 725356    -87.6947 ] [ -87.6947    0.730485 ]]
Desviación Estándar= [389.358,0.715044]	Desviación Estándar= [851.678,0.854684]
Matriz de Correlación de Pearson= [[ 1    0.18965 ] [ 0.18965    1 ]]	Matriz de Correlación de Pearson= [[ 1    -0.120474 ] [ -0.120474    1 ]]
Matriz de Correlación de Kendall= [[ 1    0.0902306 ] [ 0.0902306    1 ]]	Matriz de Correlación de Kendall= [[ 1    -0.0532582 ] [ -0.0532582    1 ]]
Matriz de Correlación de Spearman= [[ 1    0.122965 ] [ 0.122965    1 ]]	Matriz de Correlación de Spearman= [[ 1    -0.0430581 ] [ -0.0430581    1 ]]

Tabla 6.20) Estadísticas de Spatter y Domos para el conjunto de DiamBasal vs Edad

#### Marginales

Las funciones marginales mejor ajustadas para los Spatter fueron (Figura 6.29):

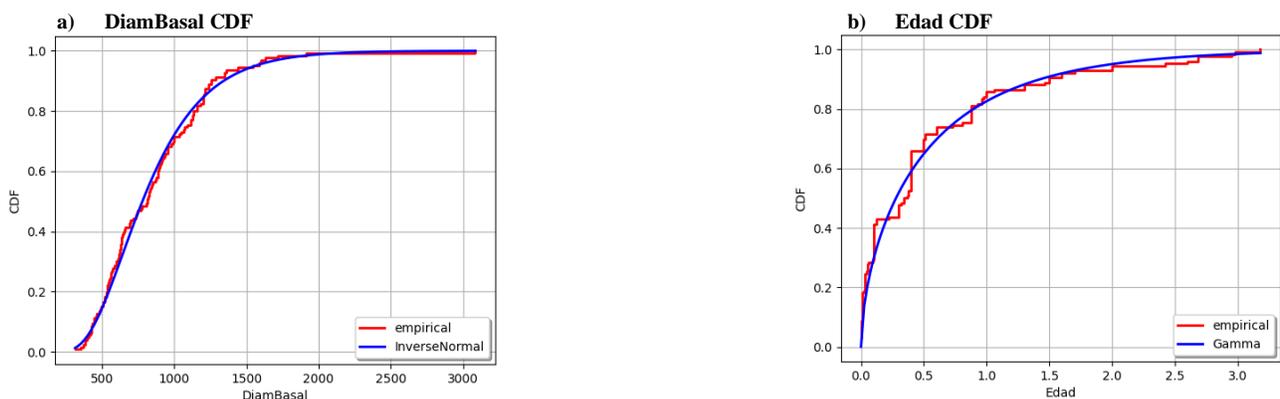
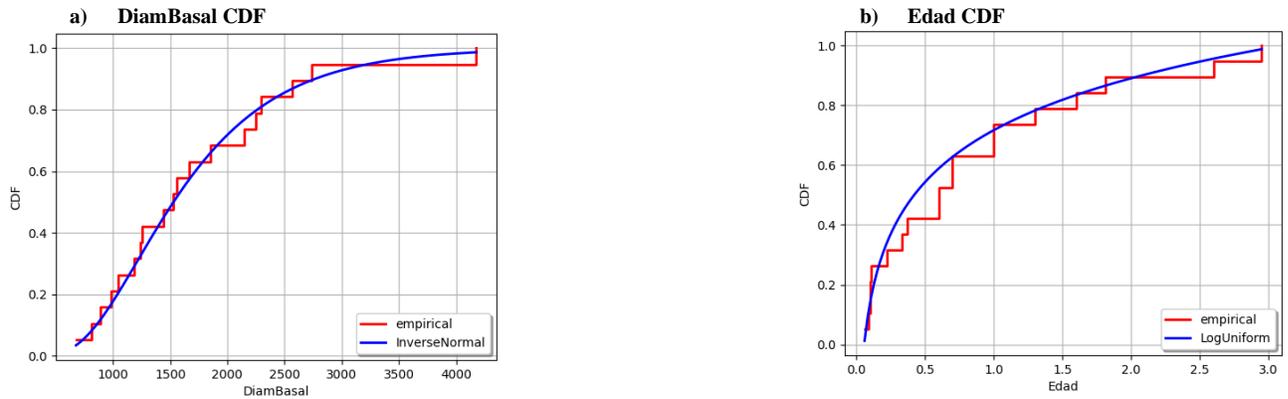


Figura 6.29) Funciones de Distribución Acumulativa (CDF) de las variables de "Spatter Diámetro Basal vs Edad": a) DiamBasal; b) Edad

$$F_{\text{Edad}} = \text{Gamma}(k = 0.577474, \lambda = 1.06276, \text{gamma} = 0.000189)$$

$$F_{\text{DiamBasal}} = \text{InverseNormal}(\mu = 852.46, \lambda = 4492.57)$$

Y para los Doms (Figura 6.30):



**Figura 6.30** Funciones de Distribución Acumulativa (CDF) de las variables de “Doms Diámetro Basal vs Edad”: a) DiamBasal; b) Edad

$$F_{\text{Edad}} = \text{LogUniform}(a_{\text{Log}} = -2.8622, b_{\text{Log}} = 1.12833)$$

$$F_{\text{DiamBasal}} = \text{InverseNormal}(\mu = 1701.24, \lambda = 7473.93)$$

## Cóputas

Se aplica las cóputas, se simulan los datos y obtenemos los resultados estadísticos que para la cóputa de Kernel son (Tabla 6.21):

----- Cópula de KernelSmoothing Spatter: Diámetro Basal vs Edad -----	----- Cópula de KernelSmoothing Domos Diámetro Basal vs Edad -----
Tamaño= 337	Tamaño= 337
Nombre= [DiamBasal,Edad]	Nombre= [DiamBasal,Edad]
Mínimo= [147.374,-0.0958373]	Mínimo= [73.0676,-0.578975]
Quartil 1= [584.417,0.0691083]	Quartil 1= [1138.87,0.285946]
Promedio= [860.327,0.587895]	Promedio= [1798.41,0.974616]
Mediana= [811.356,0.378214]	Mediana= [1630.82,0.762695]
Quartil 3= [1087.8,0.82388]	Quartil 3= [2284.75,1.34207]
Máximo= [3125.94,3.2731]	Máximo= [4885.1,3.5769]
Rango= [2978.57,3.36893]	Rango= [4812.04,4.15587]
Varianza= [159824,0.539585]	Varianza= [916766,0.865259]
Desviación Estándar= [399.78,0.734564]	Desviación Estándar= [957.479,0.930193]
Skewness= [1.72446,1.84098]	Skewness= [1.02138,0.935702]
Kurtosis= [9.54982,5.93014]	Kurtosis= [4.13168,3.12846]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 159824    65.0896 ] [ 65.0896    0.539585 ]]	CoVarianza= [[ 916766    -144.593 ] [ -144.593    0.865259 ]]
Desviación Estándar= [399.78,0.734564]	Desviación Estándar= [957.479,0.930193]
Matriz de Correlación de Pearson= [[ 1    0.221646 ] [ 0.221646    1 ]]	Matriz de Correlación de Pearson= [[ 1    -0.162347 ] [ -0.162347    1 ]]
Matriz de Correlación de Kendall= [[ 1    0.102727 ] [ 0.102727    1 ]]	Matriz de Correlación de Kendall= [[ 1    -0.0800834 ] [ -0.0800834    1 ]]
Matriz de Correlación de Spearman= [[ 1    0.149275 ] [ 0.149275    1 ]]	Matriz de Correlación de Spearman= [[ 1    -0.120538 ] [ -0.120538    1 ]]
----- pvalue=4.0442887 x10^-5	----- pvalue=0.00279827

**Tabla 6.21| Estadísticas de Spatter y Domos con la cópula de Kernel Smoothing para el conjunto de DiamBasal vs Edad**

Donde resaltan los p-values significativos en ambos conjuntos al usar la cópula de Kernel, y observando los coeficientes de correlación resulta que: en los domos promedian menores a -0.1 y en los Spatter promedian mayores a 0.1, por lo que ambos están lejanos al cero. Siempre guardando el respectivo cuidado ya que los coeficientes de correlación están más claramente lejanos del cero en Spatter, volviéndose claro que entre mayor sea la cantidad de datos con los que trabajamos, más contundentes serán los resultados (más marcadamente cercanos o lejanos al 0 estarán estos coeficientes), ya que en las últimas comparaciones resultaron los Domos (los de menos datos) aquellos que más rondan el límite de aceptación o rechazo de la dependencia.

Los resultados estadísticos al aplicar la cópula de Bernstein son (Tabla 6.22):

----- Cópula de Bernstein	----- Cópula de Bernstein
Spatter: Diámetro Basal vs Edad	Domos Diámetro Basal vs Edad
Tamaño= 337	Tamaño= 337
Nombre= [DiamBasal,Edad]	Nombre= [DiamBasal,Edad]
Mínimo= [219.128,-0.0911806]	Mínimo= [-148.07,-0.521014]
Quartil 1= [555.528,0.0328548]	Quartil 1= [1003.12,0.260454]
Promedio= [843.811,0.512854]	Promedio= [1657.16,0.915321]
Mediana= [773.592,0.292955]	Mediana= [1517.43,0.677629]
Quartil 3= [1044.75,0.597268]	Quartil 3= [2140.12,1.27025]
Máximo= [3182.29,3.18445]	Máximo= [4881.91,3.64086]
Rango= [2963.17,3.27563]	Rango= [5029.98,4.16187]
Varianza= [169866,0.526067]	Varianza= [837199,0.852181]
Desviación Estándar= [412.148,0.725305]	Desviación Estándar= [914.986,0.923137]
Skewness= [1.93444,2.02616]	Skewness= [0.965483,1.05849]
Kurtosis= [10.5212,6.59205]	Kurtosis= [4.17386,3.53506]
-----	-----
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
-----	-----
CoVarianza= [[ 169866    79.2142 ]	CoVarianza= [[ 837199    -46.3028 ]
[ 79.2142    0.526067 ]]	[ -46.3028    0.852181 ]]
Desviación Estándar= [412.148,0.725305]	Desviación Estándar= [914.986,0.923137]
Matriz de Correlación de Pearson= [[ 1    0.26499 ]	Matriz de Correlación de Pearson= [[ 1    -0.0548185 ]
[ 0.26499    1 ]]	[ -0.0548185    1 ]]
Matriz de Correlación de Kendall= [[ 1    0.0901865 ]	Matriz de Correlación de Kendall= [[ 1    -0.0973576 ]
[ 0.0901865    1 ]]	[ -0.0973576    1 ]]
Matriz de Correlación de Spearman= [[ 1    0.137168 ]	Matriz de Correlación de Spearman= [[ 1    -0.139643 ]
[ 0.137168    1 ]]	[ -0.139643    1 ]]
-----	-----
pvalue=8.015394 x10^-7	pvalue=0.31569

Tabla 6.22| Estadísticas de Spatter y Domos con la cópula de Bernstein para el conjunto de DiamBasal vs Edad

Los Domos dejan de tener un p-value significativo y sus coeficientes de correlación promedian mayor a -0.1. Mientras que los Spatter mantienen un p-value bajo y la vecindad para abarcar todos sus coeficientes de correlación sería de 0.26 de radio, es decir, están lejos al cero. De modo que existe una dependencia positiva para los Spatter, y una negativa, pero menos marcada en los Domos.

Las cópulas mejor ajustadas a cada conjunto fueron Bernstein para Spatter (Figura 6.31) y Kernel para Domos, para estos últimos se pondrán las tres cópulas dado el resultado confuso que dieron sus estadísticas. (Figura 6.32 y 6.33):

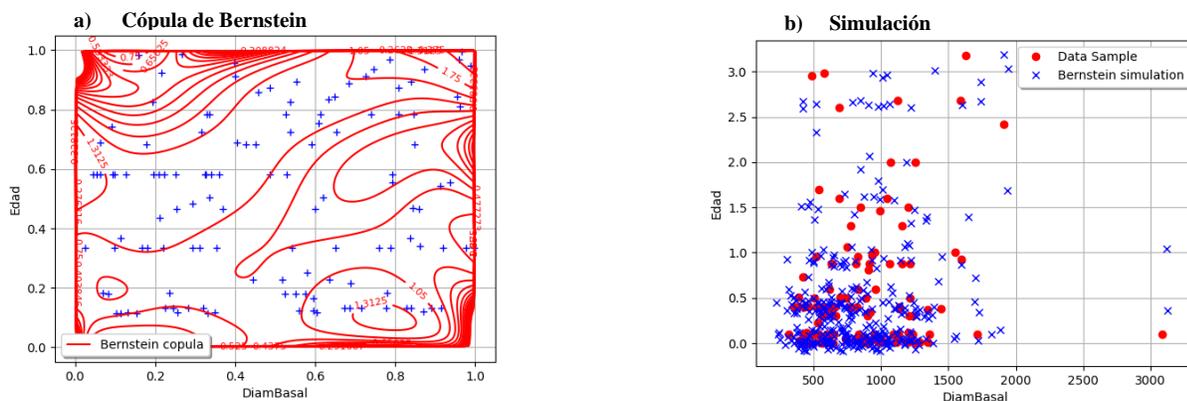
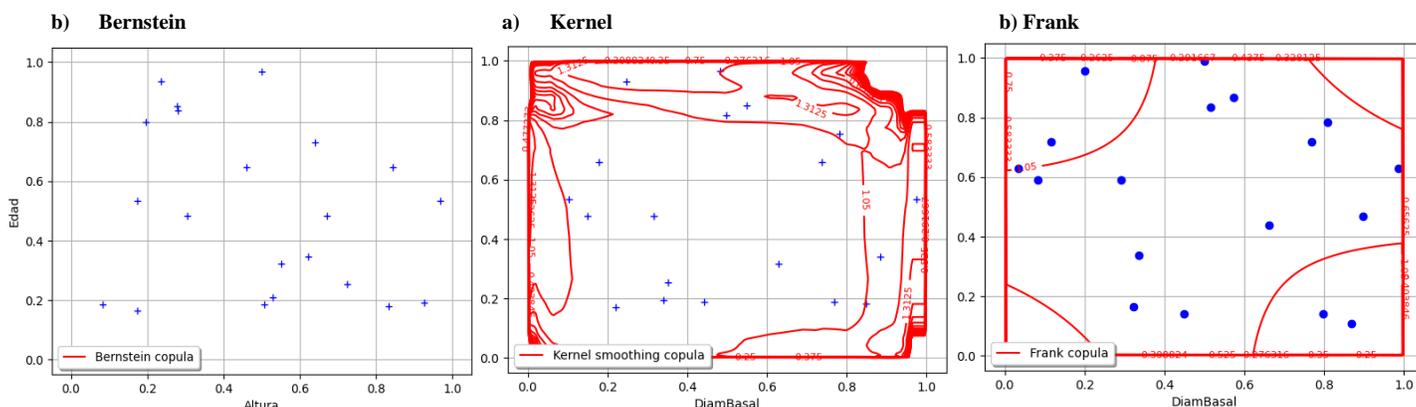


Figura 6.31| “Spatter Diámetro Basal vs Edad” con la Cópula de Bernstein: a) Gráfica de la cópula; b) Datos originales contra los simulados

El gráfico de dispersión muestra que la cópula está correctamente ajustada y su gráfico muestra la tendencia de pendiente positiva dominante en las curvas internas.

En cuanto a los Domos (Figura 6.32 y 6.33):

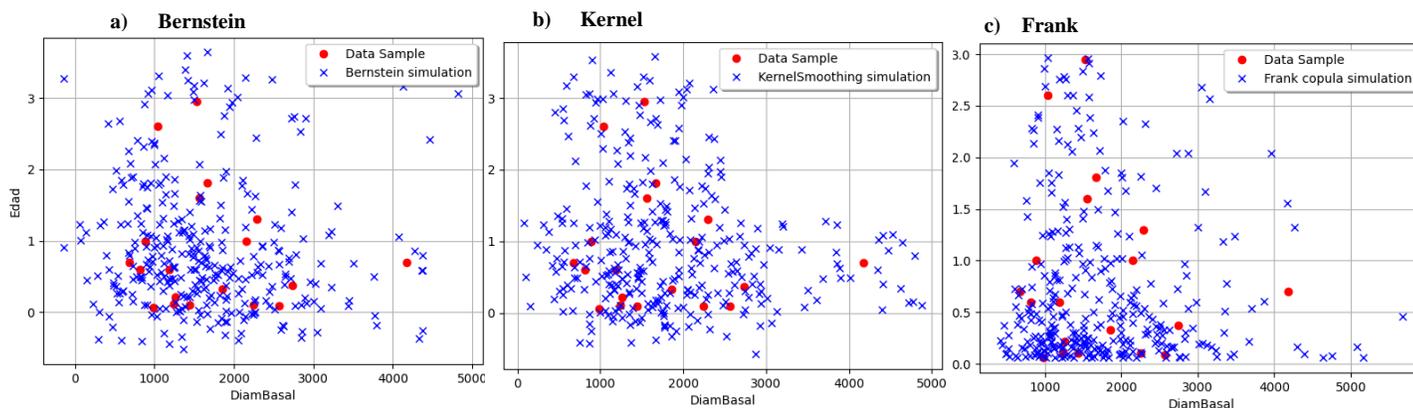
Al graficar la cópula de Bernstein se repite lo ocurrido con la figura 6.17, por lo que observamos todas las gráficas obtenidas (Figura 6.32)



**Figura 6.32** Grafica de la cópula de: a) Bernstein, b)Kernel, c)Frank; para el conjunto de Domos “Diámetro Basal vs Edad”

Se mantiene la diferencia notable entre las gráficas donde el comportamiento local de la cópula de Kernel alrededor del (1,1) aparenta ser lo que ocasiona los resultados estadísticos de dependencia, aunque, al ser local, es un comportamiento que se pierde al calcular el parámetro que tiene en cuenta el resto de datos y del cual proviene la cópula de Frank.

En las simulaciones: (Figura 6.33)



**Figura 6.33** Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing c) Frank en el conjunto de datos “Domos Diámetro Basal vs Edad”.

Se repite el comportamiento de la Figura 6.18 de poseer mayor aleatoriedad por haber menos datos originales a los que ajustarse.

En suma, hay correlación positiva entre Diámetro Basal y Edad para Spatter y negativa entre Diámetro Basal y Edad para los Domos, teniendo la reserva de que en los Domos la dependencia proviene de un comportamiento local en probabilidades altas.

En los Spatter la cópula mejor ajustada fue Bernstein, sin embargo, por lo mencionado al final de la sección 6.1.1 y dado que la de Kernel también estuvo bien ajustada y tuvo los mismos resultados será esta la que se usará como expresión que relaciona Diámetro Basal y Edad:

$$\text{Kernel}(Fx, Fy) = \left( \frac{3}{2.6730745} \right) \left( 1 - \left( \frac{Fx}{0.817477} \right)^2 - \left( \frac{Fy}{0.817477} \right)^2 \right)$$

Con Bandwith = 0.817477

Para los Domos la cópula mejor ajustada fue Kernel por lo que su expresión que relaciona Diámetro Basal y Edad es:

$$\text{Kernel}(Fx, Fy) = \left( \frac{3}{3.1357 \cdot 10^{12}} \right) \left( 1 - \left( \frac{Fx}{885.409} \right)^2 - \left( \frac{Fy}{885.409} \right)^2 \right)$$

Con Bandwith = 885.409

### 6.3 Conos y Spatter

Esta última comparación se hará en esos dos conjuntos de datos por ser las tipologías de mayor cantidad en nuestra área de estudio. Además, las conclusiones de esta comparación en una tipología se relacionan con la otra, es por esto que se le hace una sección aparte.

#### 6.3.1 Localización vs Elongación

##### 6.3.1.1 Conos

Por lo ya mencionado en *Localización vs Edad* se realizará la comparación de forma separada: latitud vs elongación y longitud vs elongación. Sólo esta vez se mostrarán los resultados a la par, pero no por ello descuidar su discusión.

#### Análisis Exploratorio de datos

Los resultados estadísticos son (Tabla 6.23):

Conos Elongación vs Latitud	Conos Elongación vs Longitud
Tamaño= 115	Tamaño= 115
Nombre= [Latitud,Elongacion]	Nombre= [Longitud,Elongacion]
Mínimo= [18.9409,2]	Mínimo= [-101.978,2]
Cuartil 1= [19.1442,42.25]	Cuartil 1= [-101.716,42.25]
Promedio= [19.2392,93.9391]	Promedio= [-101.635,93.9391]
Mediana= [19.2477,91]	Mediana= [-101.627,91]
Cuartil 3= [19.3459,142]	Cuartil 3= [-101.565,142]
Máximo= [19.4258,179]	Máximo= [-101.408,179]
Rango= [0.484909,177]	Rango= [0.570559,177]
Varianza= [0.0142216,2906.18]	Varianza= [0.0108154,2906.18]
Desviación Estándar= [0.119254,53.909]	Desviación Estándar= [0.103997,53.909]
Skewness= [-0.354874,0.0711138]	Skewness= [-0.297892,0.0711138]
Kurtosis= [2.33571,1.6357]	Kurtosis= [3.18967,1.6357]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
CoVarianza= [[ 0.0142216 0.14336 ] [ 0.14336 2906.18 ]]	CoVarianza= [[ 0.0108154 0.791614 ] [ 0.791614 2906.18 ]]
Desviación Estándar= [0.119254,53.909]	Desviación Estándar= [0.103997,53.909]
Matriz de Correlación de Pearson= [[ 1 0.0222994 ] [ 0.0222994 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.141199 ] [ 0.141199 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.00612515 ] [ -0.00612515 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.0894272 ] [ 0.0894272 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.0221127 ] [ -0.0221127 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.121016 ] [ 0.121016 1 ]]

**Tabla 6.23| Estadísticas de Conos Resultados de medidas estadísticas de los conjuntos de datos Elongación-Latitud y Elongación-Longitud**

### Marginales

Las funciones mejor acopladas de forma univariada son:

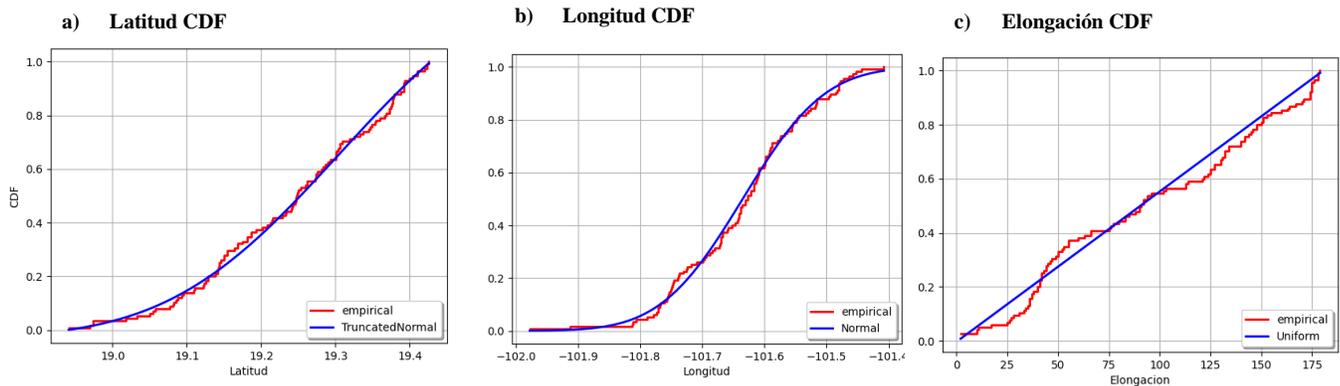
$$F_{\text{longitud}}(x) = \text{Normal}(\mu = -101.635, \sigma = 0.103997)$$

$$F_{\text{elongación}}(y) = \text{Uniform}(a = 0.487179, b = 180.513)$$

$$F_{\text{Elongación}} = \text{Uniform}(a = 0.487179, b = 180.513)$$

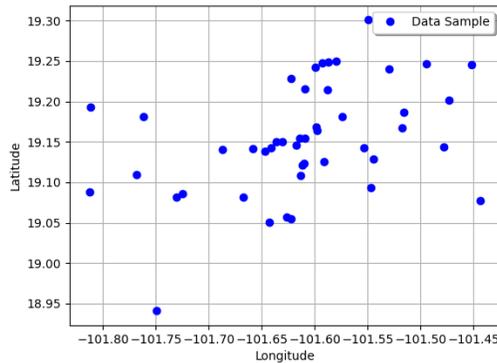
$$F_{\text{Latitud}} = \text{TruncatedNormal}(\mu = 19.3048, \text{sigma} = 0.180874, a = 18.9388, b = 19.4279)$$

Dada que la elongación en ambos conjuntos tiene los mismos datos, su función ajustada es la misma. De esa forma las gráficas de dichas funciones son (Figura 6.34):



**Figura 6.34** Funciones de Distribución Acumulativa (CDF) de las variables de "Conos Localización vs Elongación": a) Latitud | b) Longitud c) Elongación

De dichas gráficas se observa que los conos varían más en Longitud que Latitud, y eso resulta más claro al ver la gráfica de dispersión de latitud contra longitud de los 115 conos (Figura 6.35):



**Figura 6.35** Gráfico de dispersión de latitud-longitud de los conos.

### Cóputas

Los resultados estadísticos de las simulaciones de las cóputas no-paramétricas, empezando por Kernel: (Tabla 6.24)

----- Cópula de KernelSmoothing Conos Elongación vs Latitud -----	----- Cópula de KernelSmoothing Conos Elongación vs Longitud -----
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Elongacion]	Nombre= [Longitud,Elongacion]
Mínimo= [18.9079,-17.7376]	Mínimo= [-101.946,-17.7376]
Cuartil 1= [19.1604,42.8087]	Cuartil 1= [-101.697,42.8087]
Promedio= [19.2514,95.1084]	Promedio= [-101.624,95.1084]
Mediana= [19.2594,92.8427]	Mediana= [-101.62,92.8427]
Cuartil 3= [19.3468,142.956]	Cuartil 3= [-101.552,142.956]
Máximo= [19.4717,197.278]	Máximo= [-101.385,197.278]
Rango= [0.563804,215.015]	Rango= [0.560646,215.015]
Varianza= [0.0130547,3064.19]	Varianza= [0.0108748,3064.19]
Desviación Estándar= [0.114257,55.3552]	Desviación Estándar= [0.104282,55.3552]
Skewness= [-0.25632,0.0429316]	Skewness= [-0.153755,0.0429316]
Kurtosis= [2.28078,1.68505]	Kurtosis= [2.67879,1.68505]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 0.0130547 -0.117152 ] [ -0.117152 3064.19 ]]	CoVarianza= [[ 0.0108748 1.12792 ] [ 1.12792 3064.19 ]]
Desviación Estándar= [0.114257,55.3552]	Desviación Estándar= [0.104282,55.3552]
Matriz de Correlación de Pearson= [[ 1 -0.0185229 ] [ -0.0185229 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.195393 ] [ 0.195393 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0270595 ] [ -0.0270595 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.125971 ] [ 0.125971 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.0400126 ] [ -0.0400126 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.180072 ] [ 0.180072 1 ]]
----- pvalue=0.734760	----- pvalue=0.0003080349

**Tabla 6. 24| Estadísticas de Conos con la cópula de Kernel ajustada a datos de Latitud-Elongación y Longitud-Elongación**

De los respectivos p-values resulta significativo para Longitud vs Elongación en la cópula de Kernel. Con coeficientes de correlación positivos y lejanos a 0, indicando dependencia positiva. Por otra parte, para Latitud vs Elongación el p value no es significativo y sus coeficientes de correlación son cercanos a 0 por lo que no hay dependencia.

Los resultados de la cópula de Bernstein (Tabla 6.25):

----- Cópula de Bernstein Conos Elongación vs Latitud -----	----- Cópula de Bernstein Conos Elongación vs Longitud -----
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Elongacion]	Nombre= [Longitud,Elongacion]
Mínimo= [18.9168,-7.91598]	Mínimo= [-101.946,-8.45812]
Cuartil 1= [19.1472,47.6105]	Cuartil 1= [-101.705,44.9945]
Promedio= [19.2338,99.1014]	Promedio= [-101.638,96.4345]
Mediana= [19.2362,102.993]	Mediana= [-101.629,97.2455]
Cuartil 3= [19.3323,144.546]	Cuartil 3= [-101.576,144.916]
Máximo= [19.4839,194.405]	Máximo= [-101.407,195.44]
Rango= [0.567136,202.321]	Rango= [0.53888,203.898]
Varianza= [0.0144956,2924.03]	Varianza= [0.00947809,2840.7]
Desviación Estándar= [0.120398,54.0743]	Desviación Estándar= [0.0973555,53.2982]
Skewness= [-0.252886,-0.122584]	Skewness= [-0.174846,0.0058051]
Kurtosis= [2.50101,1.77501]	Kurtosis= [2.95411,1.69454]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 0.0144956 0.231177 ] [ 0.231177 2924.03 ]]	CoVarianza= [[ 0.00947809 0.481226 ] [ 0.481226 2840.7 ]]
Desviación Estándar= [0.120398,54.0743]	Desviación Estándar= [0.0973555,53.2982]
Matriz de Correlación de Pearson= [[ 1 0.0355086 ] [ 0.0355086 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.0927419 ] [ 0.0927419 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.0158966 ] [ 0.0158966 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.0645401 ] [ 0.0645401 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.025247 ] [ 0.025247 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.0943638 ] [ 0.0943638 1 ]]
----- pvalue=0.5159280 -----	----- pvalue=0.0891581622 -----

**Tabla 6.25| Estadísticas de Conos con la cópula de Bernstein ajustada a datos de Latitud-Elongación y Longitud-Elongación**

Estos resultados reafirman lo visto con las cópulas anteriores: para Latitud vs Elongación no hay dependencia, ya que el p-value vuelve a ser mayor a nuestro nivel de confianza.

En cuanto a Longitud vs Elongación el p-value es cercano pero mayor al nivel de confianza, y los coeficientes de correlación promedian algo positivo pero cercano al límite de 0.1 de que pasemos a los gráficos de las cópulas:

Para Latitud vs Elongación el resultado con ambas cópulas fue el mismo y para la Longitud el resultado cambió por poco, por lo que, por ser una comparación de particular relevancia por sus implicaciones geológicas (La elongación de una estructura volcánica es paralela a las fallas subyacentes y una relación con la localización nos diría mucho sobre el comportamiento estructural de la zona) por lo que se pondrán las gráficas de las tres cópulas (Bernstein, Kernel y Frank) y en ambas comparaciones (Elongación vs Latitud y Elongación vs Longitud).

Primero las de Frank por ser más suavizadas y sencillas de leer: (Figura 6.36)

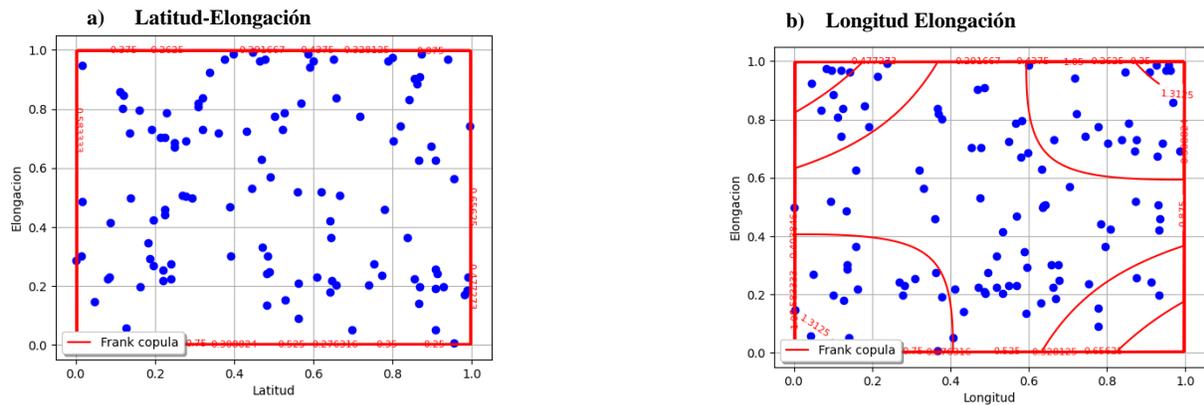


Figura 6.36 | Gráfica de las Cópulas de Frank: a) En el conjunto Latitud vs Elongación; b) En el conjunto Longitud vs Elongación.

Posteriormente las no paramétricas: (Figura 6.37)

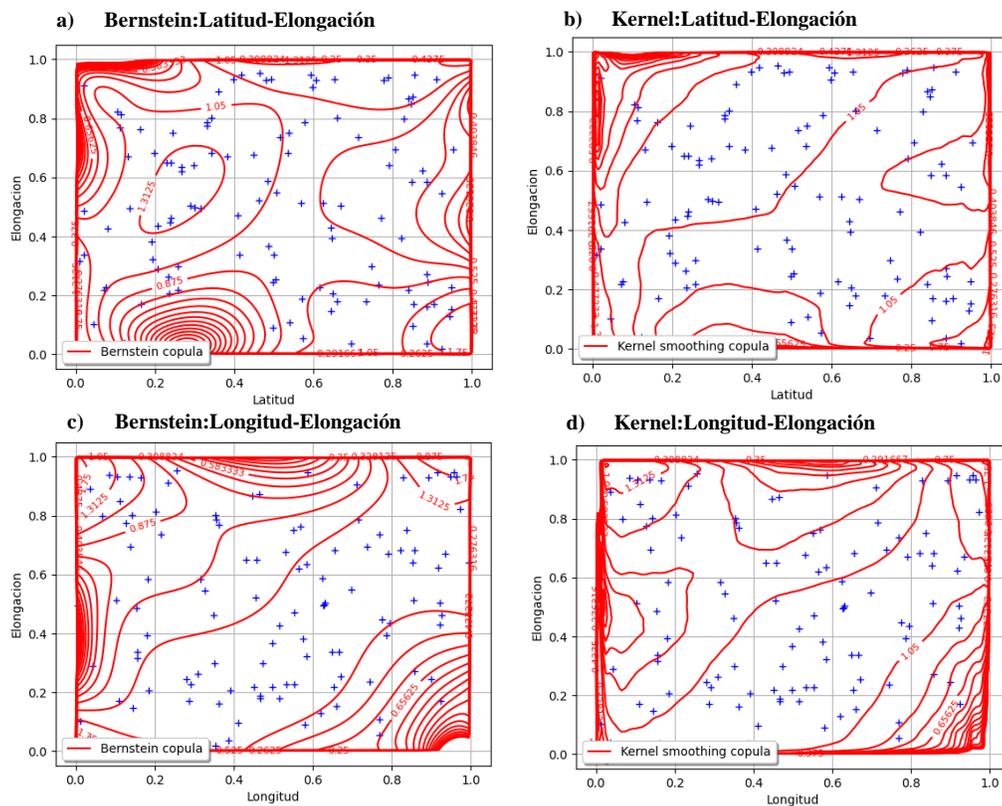
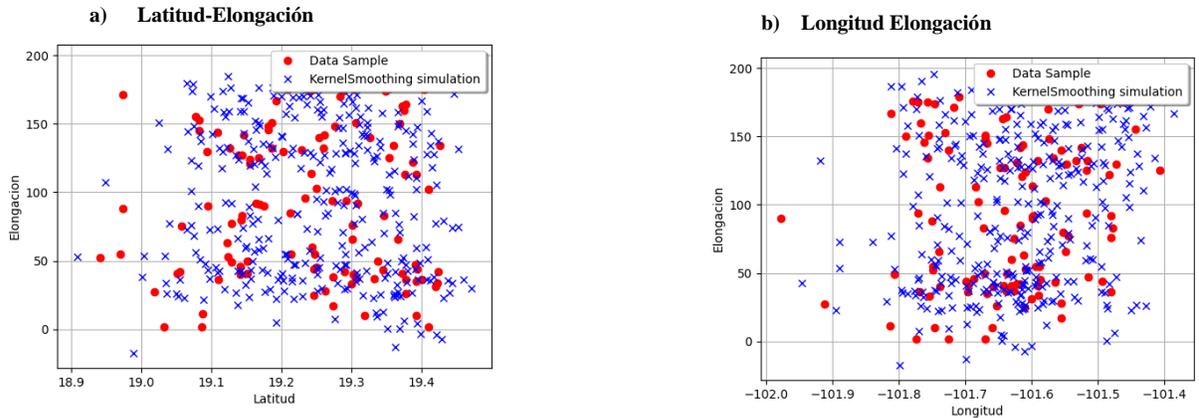


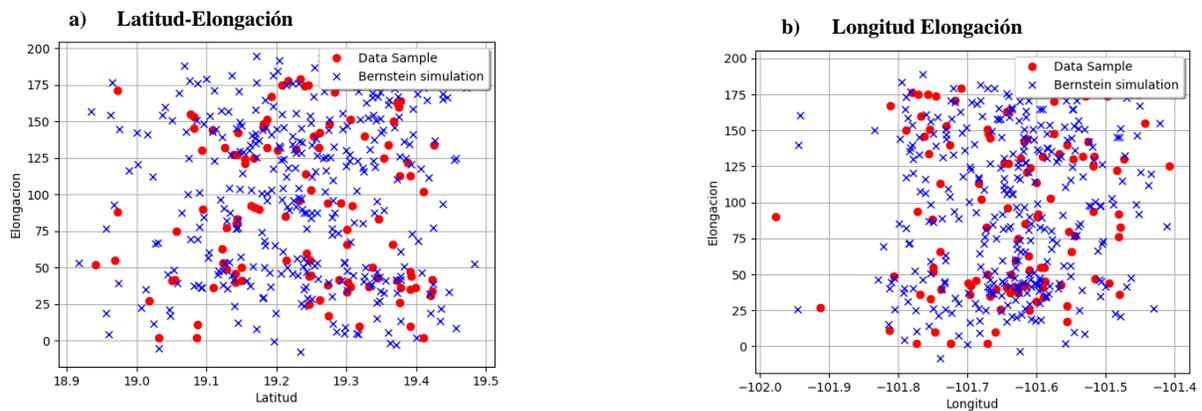
Figura 6.37 | Gráfica de las Cópulas de: a) Bernstein en el conjunto Latitud vs Elongación; b) Kernel en el conjunto Latitud vs Elongación; c) Bernstein en el conjunto Longitud vs Elongación; d) Kernel en el conjunto Longitud vs Elongación.

Con las respectivas simulaciones de las Cópulas no paramétricas, empezando por la Cópula de Kernel (Figura 6.38):



**Figura 6.38** Resultado de los datos originales contra los simulados con la Cópula de Kernel para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación

Y posteriormente la simulación proveniente de la cópula de Bernstein (Figura 6.39):



**Figura 6.39** Resultado de los datos originales contra los simulados con la Cópula de Bernstein para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación

De los resultados numéricos se obtuvo que *hay dependencia positiva entre la Longitud y la Elongación* en los Conos, mientras que la *Latitud y la Elongación* resultaron en el caso de “no hay suficiente información en contra de la hipótesis nula para descartarla”, es decir, *no hay dependencia*. Esto queda reflejado en la diferencia marcada de las gráficas de las cópulas de Frank (Figura 6.36) y en la diferencia de la tendencia de las curvas internas de las gráficas a y b contra la c y d en la Figura 6.37.

Si la mayoría de curvas van en la misma dirección, como en el caso b) de la Figura 6.36 donde 4 de las curvas internas tienen pendiente positiva y sólo 2 la tienen negativa, vemos que hay una tendencia en cierta dirección, y ello nos indica dependencia.

En síntesis, dadas las gráficas y coeficientes de correlación con p-valores confiables podemos afirmar que hay una dependencia positiva entre la Longitud y la Elongación, mientras que la Latitud y la Elongación resultaron independientes.

Para obtener la expresión precisa de dependencia se usa un procedimiento similar a los anteriores

En Longitud vs Elongación:

$$\text{Bandwidth} = 0.062747$$

$$\text{Bandwidth}^2 = 0.003937186$$

$$\frac{3}{4} * \text{Bandwidth}^2 = \frac{3}{0.015748}$$

$$\text{Kernel}(Fx, Fy) = \left( \frac{3}{0.015748} \right) \left( 1 - \left( \frac{Fx}{0.062747} \right)^2 - \left( \frac{Fy}{0.062747} \right)^2 \right)$$

Con las funciones de distribución asociadas a cada variable que fueron:

$$F_{\text{longitud}}(x) = \text{Normal}(\mu = -101.635, \sigma = 0.103997)$$

$$F_{\text{elongación}}(y) = \text{Uniform}(a = 0.487179, b = 180.513)$$

Aunque dichas expresiones son correctas, el desarrollo del porqué no hay un despeje para poner directamente a "x" en función de "y" se encuentra en el Capítulo 7.

### 6.3.1.2 Spatter

#### Análisis Exploratorio de datos

Los resultados estadísticos para los Spatter fueron (Tabla 6.26):

----- Spatter Elongación vs Latitud -----	----- Spatter Elongación vs Longitud -----
Tamaño= 251	Tamaño= 251
Nombre= [Latitud,Elongacion]	Nombre= [Longitud,Elongacion]
Mínimo= [18.8798,2]	Mínimo= [-101.998,2]
Cuartil 1= [19.0712,45]	Cuartil 1= [-101.741,45]
Promedio= [19.2003,95.0717]	Promedio= [-101.622,95.0717]
Mediana= [19.1916,91]	Mediana= [-101.618,91]
Cuartil 3= [19.3169,145]	Cuartil 3= [-101.505,145]
Máximo= [19.548,180]	Máximo= [-101.362,180]
Rango= [0.668256,178]	Rango= [0.636158,178]
Varianza= [0.0221314,3041.34]	Varianza= [0.0190677,3041.34]
Desviación Estándar= [0.148766,55.1483]	Desviación Estándar= [0.138086,55.1483]
Skewness= [0.160653,-0.0455539]	Skewness= [-0.110282,-0.0455539]
Kurtosis= [2.16334,1.68804]	Kurtosis= [2.14324,1.68804]
Resumen de Estadísticas Bivariadas	Resumen de Estadísticas Bivariadas
-----	-----
CoVarianza= [[ 0.0221314 -0.406872 ] [ -0.406872 3041.34 ]]	CoVarianza= [[ 0.0190677 -0.760373 ] [ -0.760373 3041.34 ]]
Desviación Estándar= [0.148766,55.1483]	Desviación Estándar= [0.138086,55.1483]
Matriz de Correlación de Pearson= [[ 1 -0.0495931 ] [ -0.0495931 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.0998493 ] [ -0.0998493 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0295239 ] [ -0.0295239 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.0647095 ] [ -0.0647095 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.047324 ] [ -0.047324 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.0908356 ] [ -0.0908356 1 ]]
-----	-----

**Tabla 6.26| Estadísticas de Spatter de los datos Latitud-Elongación y Longitud-Elongación**

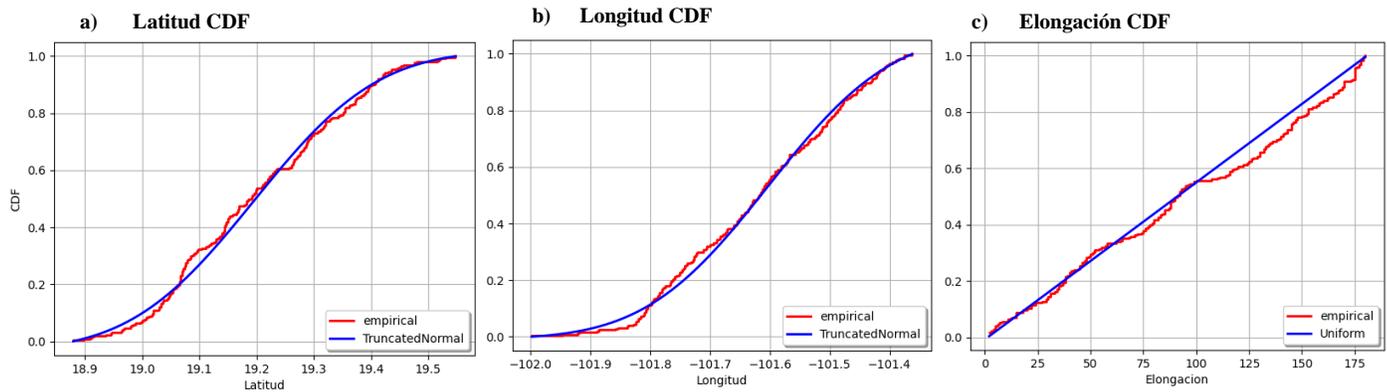
#### Marginales

Las marginales ajustadas son (Figura 6.40):

$$F_{\text{Elongación}} = \text{Uniform}(a = 1.29644, b = 180.704)$$

$$F_{\text{Latitud}} = \text{TruncatedNormal}(\mu = 19.196, \sigma = 0.170209, \\ a = 18.8784, b = 19.5494)$$

$$F_{\text{Longitud}} = \text{TruncatedNormal}(\mu = -101.603, \sigma = 0.162401, \\ a = -101.999, b = -101.361)$$



**Figura 6.40** | Funciones de Distribución Acumulativa (CDF) de las variables de "Spatter Localización vs Elongación": a) Latitud  
| b) Longitud c) Elongación

Lo primero a destacar es que tanto en Conos como en Spatter la mejor función marginal ajustada a la Elongación fue una Uniforme, esto tiene sentido ya que la elongación fue considerada en un rango de 0 a 179 grados y sólo en números enteros, y dado que los datos tuvieron 115 y 251 entradas respectivamente, resulta comprensible que, al tomar valores aleatorios en una cantidad de experimentos similar o mayor a nuestra cantidad de posibles resultados, estos se distribuyan de forma uniforme.

## Cóputas

Al ajustar la cópula de Bernstein, simular los datos y obtener sus estadísticas resulta (Tabla 6.27):

-----Cópula de Bernstein Spatter Elongación vs Latitud -----	-----Cópula de Bernstein Spatter Elongación vs Longitud -----
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Elongacion]	Nombre= [Longitud,Elongacion]
Mínimo= [18.8102,-11.3706]	Mínimo= [-102.054,-14.4709]
Cuartil 1= [19.0745,45.6622]	Cuartil 1= [-101.734,45.3246]
Promedio= [19.1933,96.428]	Promedio= [-101.635,92.6408]
Mediana= [19.1797,94.7648]	Mediana= [-101.62,91.7141]
Cuartil 3= [19.2941,146.11]	Cuartil 3= [-101.537,139.023]
Máximo= [19.6046,194.542]	Máximo= [-101.332,190.931]
Rango= [0.794405,205.913]	Rango= [0.722812,205.402]
Varianza= [0.0243196,3103.86]	Varianza= [0.0171086,3002.62]
Desviación Estándar= [0.155947,55.7123]	Desviación Estándar= [0.1308,54.7961]
Skewness= [0.333101,-0.0518896]	Skewness= [-0.202662,-0.103808]
Kurtosis= [2.55753,1.80362]	Kurtosis= [2.61854,1.8711]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 0.0243196 -0.504957 ] [ -0.504957 3103.86 ]]	CoVarianza= [[ 0.0171086 1.16505 ] [ 1.16505 3002.62 ]]
Desviación Estándar= [0.155947,55.7123]	Desviación Estándar= [0.1308,54.7961]
Matriz de Correlación de Pearson= [[ 1 -0.05812 ] [ -0.05812 1 ]]	Matriz de Correlación de Pearson= [[ 1 0.16255 ] [ 0.16255 1 ]]
Matriz de Correlación de Kendall= [[ 1 -0.0323583 ] [ -0.0323583 1 ]]	Matriz de Correlación de Kendall= [[ 1 0.0971104 ] [ 0.0971104 1 ]]
Matriz de Correlación de Spearman= [[ 1 -0.0478427 ] [ -0.0478427 1 ]]	Matriz de Correlación de Spearman= [[ 1 0.139538 ] [ 0.139538 1 ]]
----- pvalue=0.2873849	----- pvalue=0.002763801

**Tabla 6. 27| Estadísticas de Spatter con la cópula de Bernstein ajustada a datos de Latitud-Elongación y Longitud-Elongación**

Se repite algo similar a lo obtenido con los Conos: el p-value menor al nivel de confianza con coeficientes de correlación positivos y mayores a 0.1 corresponden a la comparación Elongación vs Longitud, mientras que el p-value mayor al nivel de confianza y con coeficientes de correlación cercanos al 0 son de Latitud vs Elongación.

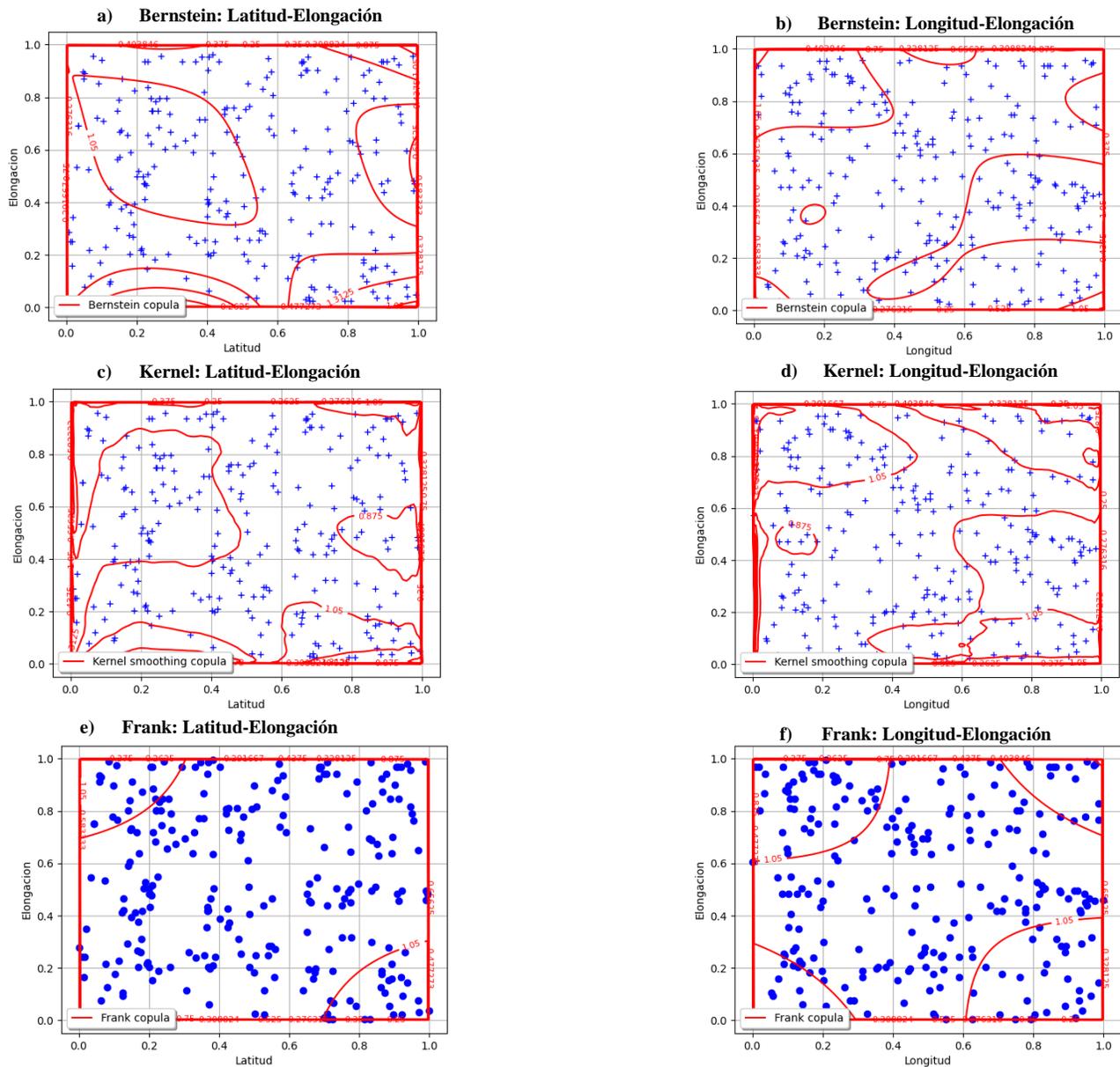
Por otra parte, al realizar el análogo con la Cópula de Kernel: (Tabla 6.28)

----- Cópula de KernelSmoothing Spatter Elongación vs Latitud -----	----- Cópula de KernelSmoothing Spatter Elongación vs Longitud -----
Tamaño= 337	Tamaño= 337
Nombre= [Latitud,Elongacion]	Nombre= [Longitud,Elongacion]
Mínimo= [18.8542,-4.57753]	Mínimo= [-101.932,-4.57753]
Cuartil 1= [19.0762,47.5441]	Cuartil 1= [-101.743,47.5441]
Promedio= [19.1972,100.449]	Promedio= [-101.634,100.449]
Mediana= [19.1826,109.127]	Mediana= [-101.645,109.127]
Cuartil 3= [19.3267,150.08]	Cuartil 3= [-101.529,150.08]
Máximo= [19.5835,193.862]	Máximo= [-101.322,193.862]
Rango= [0.729278,198.439]	Rango= [0.609233,198.439]
Varianza= [0.024575,3195.98]	Varianza= [0.0198918,3195.98]
Desviación Estándar= [0.156764,56.533]	Desviación Estándar= [0.141038,56.533]
Skewness= [0.193867,-0.199725]	Skewness= [0.150691,-0.199725]
Kurtosis= [2.31488,1.69862]	Kurtosis= [2.18925,1.69862]
----- Resumen de Estadísticas Bivariadas -----	----- Resumen de Estadísticas Bivariadas -----
CoVarianza= [[ 0.024575 0.12856 ] [ 0.12856 3195.98 ]]	CoVarianza= [[ 0.0198918 -0.564047 ] [ -0.564047 3195.98 ]]
Desviación Estándar= [0.156764,56.533]	Desviación Estándar= [0.141038,56.533]
Matriz de Correlación de Pearson= [[ 1 0.0145063 ] [ 0.0145063 1 ]]	Matriz de Correlación de Pearson= [[ 1 -0.0707418 ] [ -0.0707418 1 ]]
Matriz de Correlación de Kendall= [[ 1 0.0133531 ] [ 0.0133531 1 ]]	Matriz de Correlación de Kendall= [[ 1 -0.0572983 ] [ -0.0572983 1 ]]
Matriz de Correlación de Spearman= [[ 1 0.0188815 ] [ 0.0188815 1 ]]	Matriz de Correlación de Spearman= [[ 1 -0.0818372 ] [ -0.0818372 1 ]]
----- pvalue=0.790759	----- pvalue=0.195166

**Tabla 6. 28| Estadísticas de Spatter con la cópula de Kernel ajustada a datos de Latitud-Elongación y Longitud-Elongación**

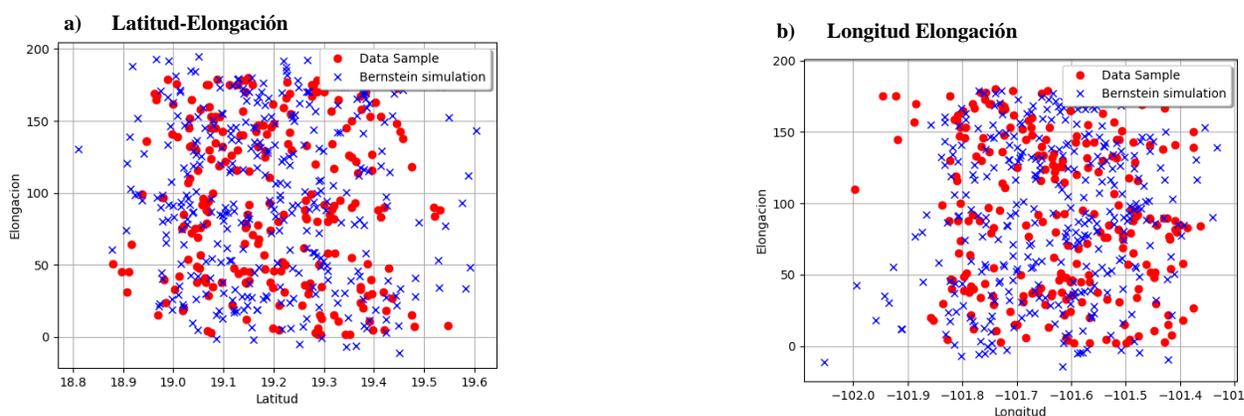
Sus p-values y coeficientes de correlación resultan en el caso de no haber suficiente evidencia en contra de la hipótesis nula para ambas comparaciones. Por lo que al analizar las gráficas volveremos a usar todas.

Queda observar las gráficas de las cópulas (Figura 6.41):

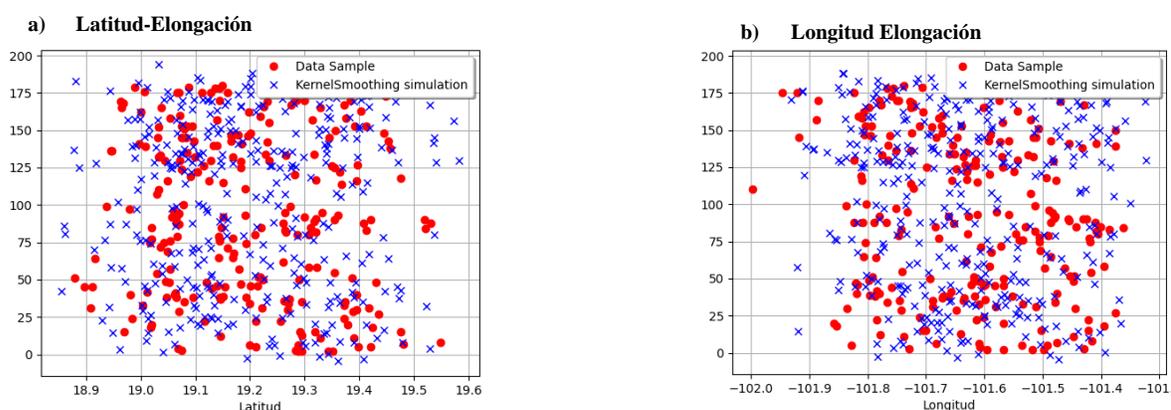


**Figura 6.41** Gráficos de las cópulas: a) De Bernstein para Latitud-Elongación, b) De Bernstein para Longitud-Elongación, c) De Frank para Latitud- Elongación, d) De Kernel para Longitud-Elongación, e) De Frank para Latitud-Elongación, f) De Frank para Longitud- Elongación.

Con la simulaciones de las Cópulas no paramétricas: (Figura 6.42 y 6.43)



**Figura 6.42** Resultado de los datos originales contra los simulados con la Cópula de Bernstein para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación, en Spatter



**Figura 6.43** Resultado de los datos originales contra los simulados con la Cópula de Kernel para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación, en Spatter

De los gráficos de dispersión de las simulaciones se observa la gran cantidad de datos además de confirmar el ajuste de las cópulas. Por otra parte, de las gráficas de las cópulas la acumulación de probabilidad en los márgenes está presente en todas, por lo que conviene centrarse en las curvas internas. Ya que en el caso e) de la Figura 6.42 todas las curvas internas tienen la misma tendencia, siendo que es una comparación que resultó independiente, no obstante hay que considerar que son pocas y cercanas a los márgenes, a comparación de aquellas del caso f), más grandes y alejadas de los márgenes.

Además al observar los casos b) y d) existe esta tendencia de pendiente positiva de las curvas internas, la cual no está presente para la comparación independiente en los casos a) y c). Esto siendo menos evidente que en el caso de los Conos, posiblemente debido a que se trata del doble de datos lo que ocasiona una gráfica con curvas en general más variantes, más cercanas en cada punto a algún valor original, que las obtenidas anteriormente.

En resumen, *hay dependencia positiva entre la Longitud y la Elongación de los Spatter, mientras que la Latitud y la Elongación resultó independiente.*

En este caso, la expresión de la cópula que relaciona Elongación y Longitud para Spatter es únicamente con la cópula de Bernstein por los resultados obtenidos, sin embargo, dado por lo explicado en la sección 6.1.1, no es posible obtener una expresión resumida para dicho polinomio por lo que lo dejaremos expresado con los parámetros estimados cómo:

Bernstein ( $F_x, F_y; 0.0783306, 18.7283$  )

## Capítulo 7- Relaciones de Dependencia

A continuación, examinaremos con mayor detenimiento los resultados obtenidos de dependencia: primero, viendo los problemas que se tienen al intentar despejar las variables originales de la cópula obtenida, y posteriormente realizando la comparación de los resultados obtenidos con cópulas en términos de dependencia contra el método de regresión lineal, por ser de los más utilizados al modelar variables.

### Despeje

Primeramente, mostraremos la imposibilidad de obtener una expresión exacta para las dos relaciones de dependencia que hemos obtenido. Veremos la prueba usando el caso concreto de la relación Edad e Índice de Aspecto.

Comenzamos recordando que la cópula de Kernel resultante fue:

$$\text{Bandwidth} = 0.120195$$

$$\text{Bandwidth}^2 = 0.014446839$$

$$\frac{3}{4} * \text{Bandwidth}^2 = \frac{3}{0.0577873521}$$

$$\text{Kernel}(x,y) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{x}{0.120195} \right)^2 - \left( \frac{y}{0.120195} \right)^2 \right)$$

De aquí hay que notar el abuso de notación que se está haciendo ya que los parámetros de la Cópula no son los valores numéricos del espacio muestral original sino los resultados (valores en el [0,1]) de las marginales de cada variable, en otras palabras, los parámetros de kernel son las funciones marginales, de forma que la notación correcta es:

$$\text{Kernel}(F_{\text{Edad}}, F_{H/W}) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{F_{\text{Edad}}(x)}{0.120195} \right)^2 - \left( \frac{F_{H/W}(y)}{0.120195} \right)^2 \right)$$

Que es lo mismo a:

$$\text{Kernel}(F_{H/W}, F_{\text{Edad}}) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - \left( \frac{F_{\text{Edad}}(y)}{0.120195} \right)^2 \right)$$

Por ser de la forma:

$$\text{Kernel}(F(x), F(y)) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{F(x)}{a} \right)^2 - \left( \frac{F(y)}{a} \right)^2 \right)$$

Con las funciones marginales:

$F_{Edad} = \text{Laplace } (\mu= 780000, \lambda= 4.54545 \times 10^{-6})$   
 $F_{H/W} = \text{Exponential } (\lambda = 9.44946, \text{gamma} = 0.113431)$

Con las expresiones:

$$F_{Edad}(y) = \int_0^y \frac{1}{2\lambda} e^{-\frac{t-\mu}{\lambda}} dt$$

$$F_{H/W}(x) = \int_0^x \lambda e^{-\lambda(t-\gamma)} dt$$

Donde "x" sí es el valor numérico del índice de aspecto y "y" la Edad.

De ahí resulta evidente que no es trivial el despejar "y" en función de "x" de la cópula de Kernel.

Para empezar, quisiéramos eliminar las integrales donde están "x" e "y", pero notemos que al derivar con respecto a cualquiera de las variables la otra se elimina:

Si derivamos con respecto a "x" y usamos la linealidad de la derivada que nos dice que:

$$\frac{d}{dx}(af(x) + b) = a \frac{d}{dx}f(x) + \frac{d}{dx}b \text{ con } a, b \text{ en } \mathbb{R}$$

entonces obtenemos:

$$\frac{d}{dx} \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{3}{0.0577873521} \right) \left( \frac{d}{dx} 1 - \frac{d}{dx} \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - \frac{d}{dx} \left( \frac{F_{Edad}(y)}{0.120195} \right)^2 \right)$$

$$\begin{matrix} \rightarrow \\ \leftarrow \end{matrix} \frac{d}{dx} \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{3}{0.0577873521} \right) \left( 0 - \frac{d}{dx} \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - 0 \right)$$

$$\begin{matrix} \rightarrow \\ \leftarrow \end{matrix} \frac{d}{dx} \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{-3}{0.0577873521} \right) \frac{d}{dx} \left( \frac{F_{H/W}(x)}{0.120195} \right)^2$$

Y algo análogo pasa con la derivada con respecto a "y":

$$\frac{d}{dy} \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{3}{0.0577873521} \right) \left( \frac{d}{dy} 1 - \frac{d}{dy} \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - \frac{d}{dy} \left( \frac{F_{Edad}(y)}{0.120195} \right)^2 \right)$$

$$\begin{matrix} \rightarrow \\ \leftarrow \end{matrix} \frac{d}{dy} \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{-3}{0.0577873521} \right) \frac{d}{dy} \left( \frac{F_{Edad}(y)}{0.120195} \right)^2$$

Por tanto, para evitar que se elimine el otro término, basta con usar lo que ya sabemos: hay una relación entre "x" y "y". Por tanto, sabemos que existe  $g(x): \mathbb{R} \rightarrow \mathbb{R}$  tal que  $y=g(x)$ . Justamente queremos saber quién es esta

$g(x)$  que relaciona “ $x$ ” con “ $y$ ”. De poder despejarse, convertiremos la cópula en una función en términos de  $x$ ,  $g(x)$  y  $g'(x)$ , que se puede resolver como ecuación diferencial.

De esta forma:

$$\text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - \left( \frac{F_{Edad}(y)}{0.120195} \right)^2 \right)$$

$$\leftarrow \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - \left( \frac{F_{Edad}(g(x))}{0.120195} \right)^2 \right)$$

Derivamos con respecto a “ $x$ ”

$$\leftarrow \frac{d}{dx} \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{3}{0.0577873521} \right) \left( \frac{d}{dx} 1 - \frac{d}{dx} \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - \frac{d}{dx} \left( \frac{F_{Edad}(g(x))}{0.120195} \right)^2 \right)$$

$$\leftarrow \frac{d}{dx} \text{Kernel}(F_{H/W}, F_{Edad}) = \left( \frac{-3}{0.0577873521} \right) \left( \frac{d}{dx} \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 + \frac{d}{dx} \left( \frac{F_{Edad}(g(x))}{0.120195} \right)^2 \right)$$

Aquí conviene recordar dos cosas:

- 1- La regla de la cadena nos dice que:

$$\frac{d}{dx} f(h(x)) = f'(h(x)) h'(x)$$

- 2- El Teorema Fundamental del Cálculo (TFC) nos dice que:

$$\frac{d}{dx} \int_a^x f(t) dt = f(x)$$

Por el TFC se observa que  $F'_{H/W}(x)$  ya no sería una Integral, que es un primer paso para poder despejar “ $y$ ” en función de “ $x$ ”. Lo mismo con  $F'_{Edad}(y)$ .

Regresando a nuestra última igualdad, intentemos desarrollar el último término:

$$\frac{d}{dx} \left( \frac{F_{Edad}(g(x))}{0.120195} \right)^2$$

Por facilitar la notación, haremos  $0.120195 = a$ .

$$\leftarrow \frac{d}{dx} \left( \frac{F_{Edad}(g(x))}{a} \right)^2$$

$$\leftarrow \frac{d}{dx} \left( \frac{F_{Edad}(g(x))^2}{a^2} \right)$$

$$\leftarrow \frac{1}{a^2} \frac{d}{dx} \left( F_{Edad}(g(x))^2 \right)$$

Hacemos una función auxiliar para reescribir lo que tenemos. Sea  $h(k) = k^2$

$$\begin{aligned} &\rightarrow \frac{1}{a^2} \frac{d}{dx} (h(F_{\text{Edad}}(g(x)))) \\ &\leftarrow \end{aligned}$$

Aplicamos regla de la cadena a  $h(R(x))$  con  $R(x)=F_{\text{Edad}}(g(x))$

$$\begin{aligned} &\rightarrow \frac{1}{a^2} \frac{d}{dx} (h(R(x))) \\ &\leftarrow \end{aligned}$$

$$\begin{aligned} &\rightarrow \frac{1}{a^2} (h'(R(x))) \left( \frac{d}{dx} R(x) \right) \\ &\leftarrow \end{aligned}$$

Dado, regresamos  $R(x)$  a su expresión original y al desarrollar volvemos a usar regla de la cadena:

$$\begin{aligned} &\rightarrow \frac{1}{a^2} (h'(F_{\text{Edad}}(g(x)))) \left( \frac{d}{dx} F_{\text{Edad}}(g(x)) \right) \\ &\leftarrow \end{aligned}$$

$$\begin{aligned} &\rightarrow \frac{1}{a^2} (h'(F_{\text{Edad}}(g(x)))) (F'_{\text{Edad}}(g(x))) (g'(x)) \\ &\leftarrow \end{aligned}$$

Dado que  $h(x) = x^2$  entonces  $h'(x) = 2x$

$$\begin{aligned} &\rightarrow \frac{1}{a^2} (2(F_{\text{Edad}}(g(x)))) (F'_{\text{Edad}}(g(x))) (g'(x)) \\ &\leftarrow \end{aligned}$$

Entonces, la integral que queríamos eliminar para poder hacer el despeje, sigue presente ya que tenemos aún  $F_{\text{Edad}}(g(x))$ .

Además ahora tenemos un producto de funciones de  $x$ , por lo que si intentamos volver a derivar, la integral no se va a eliminar dado que:

$$\frac{d}{dx} f(x) h(x) = f'(x)h(x) + f(x)h'(x)$$

Y la igualdad anterior vuelve a usarse sin importar cuántas veces derivemos. Es por esto que no podemos quitarnos a  $F_{\text{Edad}}(g(x))$ .

Dado que eso fue ocasionado por estar dentro de otra función ( $h(x)=x^2$ ), algo análogo ocurre con el otro término:

$$\frac{d}{dx} \left( \frac{F_{H/W}(x)}{0.120195} \right)^2$$

$$\begin{aligned} &\rightarrow \frac{1}{a^2} (2(F_{H/W}(x))) (F'_{H/W}(x)) \\ &\leftarrow \end{aligned}$$

Donde se repite la presencia de un producto de funciones. Por todo lo anterior, concluimos que no es posible dar la función precisa de "y" en función de "x" más allá de la Cópula:

$$\text{Kernel } (F_{H/W}, F_{Edad}) = \left( \frac{3}{0.0577873521} \right) \left( 1 - \left( \frac{F_{H/W}(x)}{0.120195} \right)^2 - \left( \frac{F_{Edad}(y)}{0.120195} \right)^2 \right)$$

Si bien eso ocasiona que no sea factible un despeje de la edad en términos del índice de aspecto de forma directa, sí tenemos la función que relaciona sus funciones de distribución y además las funciones de distribución marginales estimadas junto con sus valores numéricos. Es decir, no hay más incógnitas que la misma  $x$  de Índice de aspecto y “ $y$ ” de edad (que ya argumentamos por qué se puede ver como  $y=g(x)$ ) de forma que la aproximación de una variable en términos de la otra es obtenible por métodos numéricos. Sin olvidar que este modelo es estocástico, es decir, probabilista, no determinista, lo cual hace más confiables nuestros resultados a pesar de que cambien cada vez que se calculen. Dicho margen de variabilidad es pequeño por el nivel de confianza contra el que comparamos los  $p$ -values.

Queda únicamente ver una comparación contra un método clásico y no-probabilista de estimación como lo es la regresión lineal. Lo usaremos únicamente en los resultados de dependencia que obtuvimos, ya que bastarán para notar las principales diferencias entre ambos modelos y justificar la introducción del uso de cópulas a este tipo de estudios.

## Regresión

Entonces usaremos la regresión lineal en las comparaciones: Edad contra Índice de Aspecto ( en los conjuntos “Conos edad ajustada” y “Conos B”) y Localización contra Elongación de los conos (en los conjuntos “latitud-elongación” y “longitud-elongación”).

### 1. Índice de Aspecto vs Edad

De una regresión lineal usando la Edad en función del índice (es decir, la Edad como la variable dependiente), en el conjunto de “Conos B” obtenemos la ecuación  $y=755,556 -404,569X$ . De ver la ecuación y teniendo en cuenta que  $x$  ( el Índice de aspecto) es una variable positiva continua en el  $[0,1]$ , se obtiene que la edad máxima considerada o estimable con el modelo es de 755 mil años (a pesar de haber sido calibrada con una máxima de 1.5Ma), así como la mínima resulta de 350,987 mil años (en el caso no-factible que  $H/W$  fuese 1). Si consideramos los casos reales (con  $\max\{H/W\}=0.6$ ) resulta una edad mínima estimable de 512,814 años. Nuevamente, alejada de los datos con los que fue calibrado ya que el mínimo en el conjunto Conos B son 80,000 años.

Aun dejando el plano práctico y siendo puramente teóricos, los datos sugieren que estadísticamente la prueba no resulta significativa debido a que, una vez hecho el ajuste, si bien el primer paso usual es observar el valor de la  $R^2$  estandarizada y buscar si es cercano a 1 para ver el buen o mal ajuste del modelo, lo correcto es hacer una prueba de Bondad de Ajuste. Además, dados los análisis previos en la sección de resultados, resulta trivial ver que los datos no son lineales, por lo cual la  $R^2$  no es confiable.

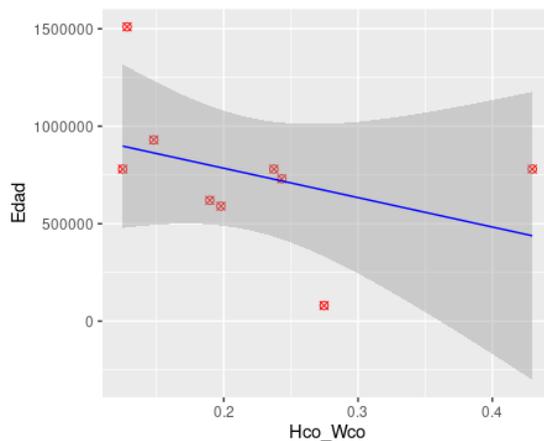
Entonces, para una aceptación o rechazo significativo del modelo es necesario una prueba de bondad de ajuste, utilizando el valor crítico (obtenido a partir del nivel de confianza, los grados de libertad del numerador que resultan los mismos a los grados de libertad usados en la regresión, y los grados de libertad del denominador que son la cantidad de observaciones restando los grados de libertad del numerador). Este valor crítico se compara con el valor obtenido de la estadística  $F$  que como sólo hay un regresor equivale al estadístico  $t$ , y si

resulta el valor crítico menor que la estadística, entonces se rechaza la Hipótesis nula. En caso de ser mayor, nuevamente, lo único que es posible concluir es que no hay suficiente Evidencia en contra de la hipótesis nula, lo cual en nuestro caso se traduce en que el modelo de regresión Lineal no es significativo.

Cabe recordar que en este caso  $H_0: B_1=0$  y  $H_a: B_1 \neq 0$ .

Dado que  $B_1$  es el coeficiente que hace que X aparezca en el modelo, ésta prueba es la que nos dice si el modelo ajustado tiene sentido, en otras palabras, si es correcto buscar una relación entre X y Y. Todo ello utilizando los mejores estimadores lineales insesgados de mínima varianza para  $B_0$  y  $B_1$

En nuestro caso particular los resultados para la regresión lineal fueron (Figura 7.1):



**Figura 7.1** Gráfica de la regresión lineal del conjunto "Conos B" al comparar Índice de Aspecto con Edad

De la gráfica es apreciable como los intervalos de predicción aumentan conforme nos alejamos de la media, esto es porque estos modelos están centrados en la media de las observaciones. En cuanto a los resultados numéricos obtenidos:

```

Estimate Std. Error    t value  Pr(>|t|)
(Intercept)  755556    122186   6.184 0.000452 ***
poly(x, 1)  -404569   366556  -1.104 0.306213
Residual standard error: 366600 on 7 degrees of freedom
Multiple R-squared:  0.1482,    Adjusted R-squared:  0.02655
F-statistic: 1.218 on 1 and 7 DF
Valor crítico:5.317655
p-value: 0.3062
    
```

De forma que, dado que el valor crítico fue mayor que la estadística F, es decir un equivalente a observar que el p-value es mayor a 0.05 (nuestro nivel de confianza utilizado) decimos que nos falta evidencia para rechazar  $H_0$  de forma que el modelo ajustado no es representativo para nuestros datos.

Una razón más para rechazar el modelo es aquello que suele saltarse en sus ocasiones de uso: la corroboración de supuestos, como lo son la Homocedasticidad, el uso de errores estudiantizados en vez de estandarizados si la muestra es pequeña (menor a 50 datos), la Linealidad y la Normalidad de los errores, además de que estos sean independientes e idénticamente distribuidos con distribución Normal de media 0 y varianza sigma cuadrada.

De estos, nos centraremos en la Homocedasticidad que es verificar que la varianza es constante. Y en este caso particular, en sus resultados estadísticos previos a la aplicación de cópulas (Tabla 6.1) se observó justamente que la varianza incrementa al ir incluyendo los datos de menores edades, lo cual ya nos dice que no es constante y por tanto no es ajustable un modelo de regresión lineal.

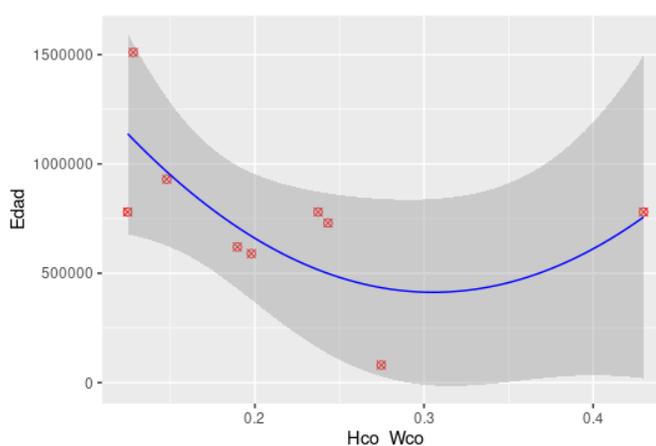
Para los estadísticos esto puede resultar evidente y repetitivo, ya que obtuvimos el coeficiente de correlación de Pearson para los datos originales en cada comparación, y dicho coeficiente al cuadrado te da la  $R^2$ , además existen actualmente pruebas de bondad de ajuste usando directamente la  $R^2$ , de forma que para que este análisis extra usando métodos de regresión para argumentar de forma más certera la utilidad que presenta el uso de cópulas en lugar de estos métodos, será necesario ir más allá de la regresión lineal e intentar suponer que los datos se distribuyen de otras formas como cuadrática o cúbica.

Por ello, la siguiente posible hipótesis sería tratar de acoplar modelos de regresión cuadráticos (Figura 7.2) o cúbicos (Figura 7.3).

Para la regresión cuadrática la ecuación obtenida es:

$$y=2,485,345 -13,570,606x+22,216,645x^2.$$

De donde:



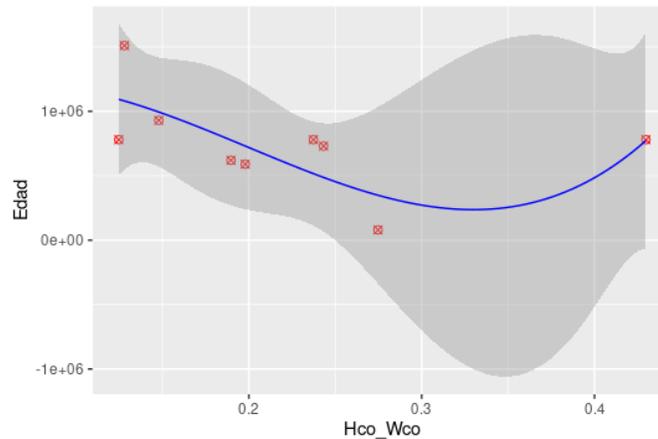
**Figura 7.2|** Grafica de la regresión cuadrática del conjunto "Conos B" al comparar Índice de Aspecto con Edad

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2485345	730550	3.402	0.0145 *
x	-13570606	5968707	-2.274	0.0634 .
I(x^2)	22216645	10794394	2.058	0.0853 .

Residual standard error: 303100 on 6 degrees of freedom  
 Multiple R-squared: 0.5007, Adjusted R-squared: 0.3343  
 F-statistic: 3.009 on 2 and 6 DF  
 Valor crítico: 4.737414  
 p-value: 0.1245

Con un valor crítico resultante de 4.73, mientras que la estadística F fue de 3.0, por lo que se no hay evidencia suficiente para rechazar  $H_0$ , entonces el modelo predictivo no es asegurable que tenga sentido. Entonces pasamos a intentarlo con una regresión cúbica de donde obtenemos:

$$Y = 965718 + 7,486,080x - 65,472,035 x^2 + 109,375,693 x^3$$



**Figura 7.3|** Grafica de la regresión cúbica del conjunto "Conos B" al comparar Índice de Aspecto con Edad

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	965718	3621741	0.267	0.800
x	7486080	49407853	0.152	0.885
I(x^2)	-65472035	204339940	-0.320	0.762
I(x^3)	109375693	254465070	0.430	0.685

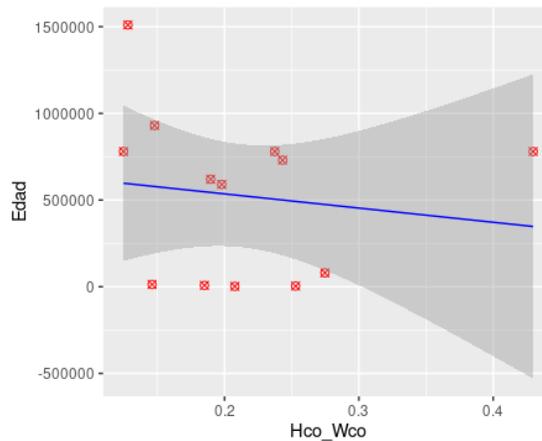
Residual standard error: 326100 on 5 degrees of freedom  
 Multiple R-squared: 0.5185, Adjusted R-squared: 0.2296  
 F-statistic: 1.795 on 3 and 5 DF  
 Valor crítico: 4.757063  
 p-value: 0.2647

Nuevamente sin poder rechazar  $H_0$ , concluimos que no es factible ajustar una regresión a nuestros datos de manera significativa.

Los resultados para la comparación del conjunto "Conos Edad Ajustada" no cambia mucho, seguramente debido a esta varianza no-constante que se manifiesta en mayor medida al agregar los pocos datos que diferencian ambos conjuntos.

De la regresión lineal (Figura 7.4) se obtuvo:

$$Y=525,165-230,239x$$



**Figura 7.4|** Grafica de la regresión lineal del conjunto “Conos Edad Ajustada” al comparar Índice de Aspecto con Edad

```

Estimate Std. Error t value Pr(>|t|)
(Intercept)  525165    134940   3.892  0.00251 **
poly(x, 1)   -230239    486533  -0.473  0.64531
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

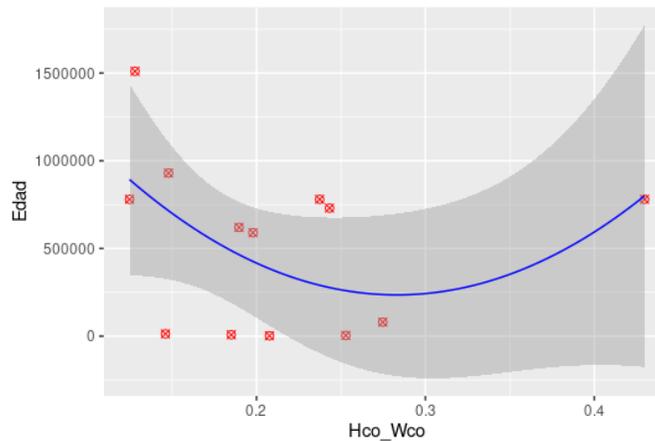
```

Residual standard error: 486500 on 11 degrees of freedom  
Multiple R-squared: 0.01995, Adjusted R-squared: -0.06914  
F-statistic: 0.2239 on 1 and 11 DF  
Valor crítico: 4.747225  
p-value: 0.6453

Como se esperaba, el valor crítico queda lejos de ser menor que la estadística F, de donde no podemos concluir que el modelo tenga sentido.

En caso de una regresión cuadrática (Figura 7.5):

$$Y = 2,337,894 - 14,854,114x + 26,234,464x^2$$



**Figura 7.5|** Grafica de la regresión cuadrática del conjunto “Conos Edad Ajustada” al comparar Índice de Aspecto con Edad

```

Estimate Std. Error t value Pr(>|t|)

```

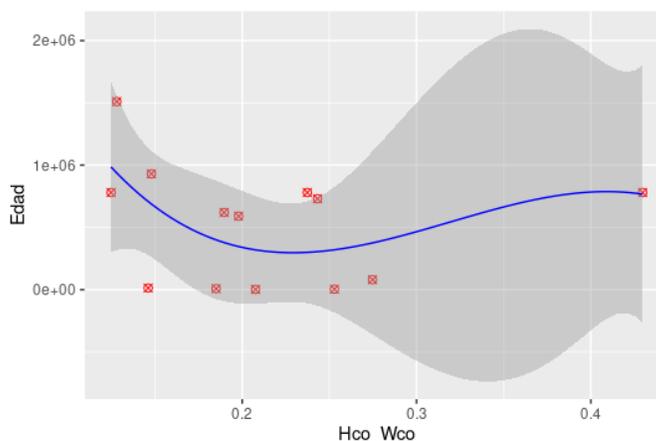
```

(Intercept) 2337894 961012 2.433 0.0353 *
x -14854114 7807158 -1.903 0.0862 .
I(x^2) 26234464 14294827 1.835 0.0963 .
Residual standard error: 441300 on 10 degrees of freedom
Multiple R-squared: 0.2669, Adjusted R-squared: 0.1203
F-statistic: 1.82 on 2 and 10 DF
Valor crítico: 3.982298
p-value: 0.2118

```

No hay cambios en nuestras conclusiones, así que pasemos a ver el último de estos casos, la regresión cúbica (Figura 7.6):

$$Y = 965718 + 7,486,080x - 65,472,035 x^2 + 109,375,693 x^3$$



**Figura 7.6** Gráfica de la regresión cúbica del conjunto "Conos Edad Ajustada" al comparar Índice de Aspecto con Edad

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4719605	4385394	1.076	0.310
x	-47471453	59045312	-0.804	0.442
I(x^2)	161624644	243228812	0.664	0.523
I(x^3)	-168854755	302784831	-0.558	0.591

```

Residual standard error: 457400 on 9 degrees of freedom
Multiple R-squared: 0.2914, Adjusted R-squared: 0.05515
F-statistic: 1.233 on 3 and 9 DF
Valor crítico: 3.708265
p-value: 0.3533

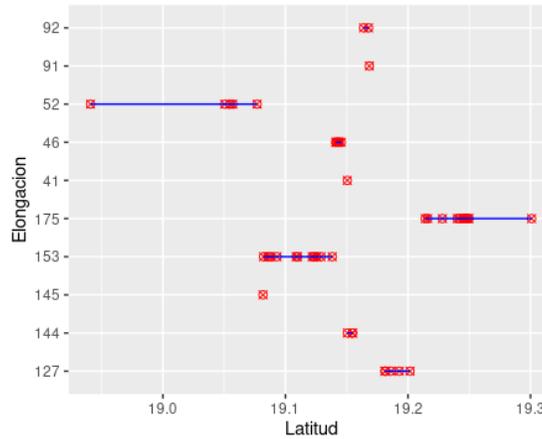
```

De forma que por el comportamiento esencial de nuestros datos no es posible ajustar algún modelo de regresión clásico de manera significativa. En cambio, con las cópulas si obtuvimos ajustes significativos debido a que las cópulas es un ajuste probabilista mientras que la regresión clásica es más determinista. Sin embargo, hay modelos polinómicos a pedazos (splines), que son regresiones que toman en cuenta la vecindad de ciertos puntos, aunque su ajuste es más tardado que el de una cópula.

Para poder generalizar y observar que eso se cumple en el resto de nuestras comparaciones, veremos el otro caso en que las cópulas nos regresaron resultados bien ajustados y con dependencia: la comparación de Localización contra elongación en los conjuntos "Latitud-Elongación" (Figura 7.7) y "Longitud-Elongación" por separado

## 2. Localización vs Elongación

Al momento de hacer el ajuste resulta un fallo:



**Figura 7.7|** Grafica de la regresión lineal del conjunto "Latitud-Elongación" al comparar Localización contra Elongación

La elevada densidad de estructuras en el área ocasiona que, al considerar por separado cada una de las coordenadas, inevitablemente algunas de ellas presenten valores coincidentes. Esto conlleva a la presencia de datos colineales, lo cual impide la aplicación de un modelo de regresión.

Este problema se sustenta en la propia naturaleza densa de la distribución espacial de las estructuras en la zona de estudio (y en el CVMG en general), de donde se deriva que ciertos datos adquieran valores iguales al enfocarse el análisis solo sobre un eje de coordenadas. Pretender eliminar datos para evitar la colinealidad conduciría a sesgar la muestra y, por ende, a tergiversar los resultados.

Tampoco es viable soslayar el inconveniente intercambiando las variables dependiente e independiente. Aun considerando la elongación como variable independiente, su rango discreto en grados implica que igualmente se den repeticiones de valores entre los datos. Esto nuevamente invalida la aplicación de un modelo de regresión.

En resumen, el método de regresión lineal es utilizable para inferencia, no para predicción. Ello no quita su utilidad ya que sigue siendo utilizable en estudios de enfoque explicativo, no predictivos. Siempre y cuando los datos a los que es aplicado cumplan con las hipótesis del modelo, hipótesis que nuestra muestra no cumple.

## Capítulo 8- Conclusión y Discusión

Este trabajo buscó modelar variables morfométricas de tres tipologías distintas de estructura volcánica por medio de métodos probabilísticos como son las cópulas. Generando modelos en los que existe una secuencia cambiante de eventos analizables probabilísticamente, es decir, estocásticos. A fin de buscar posibles relaciones de dependencia entre dichas variables y proponer el uso de estos modelos probabilistas para futuros estudios.

Se encontraron dichas relaciones de dependencia por medio de pruebas de hipótesis aplicadas a los coeficientes de correlación obtenidos de las simulaciones de las cópulas ajustadas. Se comprobaron las ventajas de estas para este tipo de datos en lugar de métodos de modelación no-probabilistas como las regresiones, y futuras líneas de investigación podrán corroborar lo eficaz de este método con la cantidad suficiente de datos.

En detalle, los resultados fueron (Tabla 8.1):

tipología	X	Y	Marginal X	Marginal Y	pvalue		parametro	promedio coef de correlacion
					bernstein	kernel		
Conos	Latitud	Elongación	TruncatedNormal	Uniforme	5.16E-01	7.35E-01	-0.06	0.03
Conos	Longitud	Elongación	Normal	Uniforme	8.90E-02	3.08E-04	0.81	0.16
Spatter	Latitud	Elongación	TruncatedNormal	Uniforme	2.87E-01	7.91E-01	-0.27	-0.04
Spatter	Longitud	Elongación	TruncatedNormal	Uniforme	2.70E-03	1.95E-01	-0.58	0.13
Conos Edad Ajustada	Índice de Aspecto	Edad	Exponencial	Beta	5.00E-05	2.80E-01	-	-0.21
Conos B	Índice de Aspecto	Edad	Exponencial	Laplace	5.59E-01	2.60E-08	-	-0.26
Conos	Latitud	Edad	Logistic	Weibul Min	4.07E-02	6.60E-01	-	-0.08
Conos	Longitud	Edad	Laplace	Weibul Min	3.26E-01	1.77E-01	0.01	0.04
Spatter	Latitud	Edad	Gumbel	Gamma	6.70E-01	1.40E-01	-0.17	0.03
Spatter	Longitud	Edad	NormalTruncada	Gamma	4.60E-01	1.30E-01	0.69	0.15
Domos	Latitud	Edad	Uniforme	LogUniforme	8.40E-02	8.20E-01	1.32	-0.01
Domos	Longitud	Edad	WeibullMax	LogUniforme	1.30E-01	7.40E-01	0.67	-0.10
Spatter	Altura	Edad	LogNormal	Gamma	6.50E-01	5.00E-01	-0.45	-0.02
Spatter	Diámetro Basal	Edad	NormalInversa	Gamma	7.00E-08	4.04E-05	-	0.16
Domos	Altura	Edad	NormalInversa	LogUniforme	1.89E-01	4.00E-04	1.09	-0.18
Domos	Diámetro Basal	Edad	NormalInversa	LogUniforme	3.15E-01	2.70E-03	-	-0.12

**Tabla 8.1** | Tabla de los la tipología usada, cada pareja de variables comparada, las marginales estimadas a cada variable, los p-values de las copulas no paramétricas, el parámetro de la cópula paramétrica y el promedio de los coeficientes de correlación de alguna de las cópulas no paramétricas.

Se repasarán y discutirán los resultados de cada comparación:

Los conos (de escoria) si tienen relación de dependencia entre el Índice de Aspecto y la Edad, dicha dependencia es negativa y no-lineal (en general, todas las dependencias obtenidas no son lineales) y, por lo tanto, el Índice de Aspecto o Juventud es una buena aproximación para inducir edades de Conos. Ello confirma la hipótesis de Wood, pudiendo obtener aproximaciones certeras de este dato sin necesidad de más información que su Altura y Diámetro Basal, obtenible hoy día por diversos métodos como el satelital.

En cuanto a esta comparación cabe observar que la escasez de datos intermedios (edades entre 80,000 y 20,000 años) que ocasiona la variabilidad discutida en la sección 6.1.1, incide negativamente en los modelos. Al pretender estimar y ajustar las funciones de distribución marginales (y la consiguiente cópula), las funciones carecen de datos reales contra los cuales calibrarse en un amplio rango de valores. Más allá de ello, muchos métodos de estimación parten de estadísticas generales de los datos, incluyendo media y varianza (o funciones generadoras), de modo que alterar en tal grado dichos parámetros con observaciones tan escasas (ocasionando

que al agregar un par de datos la diferencia entre el mayor y el menor incremento 60 veces) redundante en mayor incertidumbre a la hora de generar un modelo. Lo cual dificulta alcanzar un ajuste sólidamente confiable.

Los Conos, Domos y Spatter no tienen dependencia de Edad con la localización dentro de nuestra área de estudio. Sin embargo, dicho resultado puede deberse no sólo a la poca cantidad de datos disponibles (teniendo dos datos mayores a un 1 Ma y el resto menores) sino, y probablemente en mayor medida, a la región en particular del campo que se está estudiando: que es la zona sureste del CVMG en un cuadrado de  $0.6^\circ$  por lado, lo cual puede ser un área pequeña (a comparación del tamaño del campo) para mostrar o rechazar una posible dependencia más general del campo.

Los Conos y Spatter tienen relación de dependencia entre su Longitud y su Elongación, y es positiva, en otras palabras, entre más cercano al este está la estructura, su elongación tiende a ser en dirección Norte-Sur.

Los Spatter no tienen relación de dependencia entre su Altura y su Edad. Mientras que los Domos sí la tienen y es negativa, es decir, a mayor edad, menor altura.

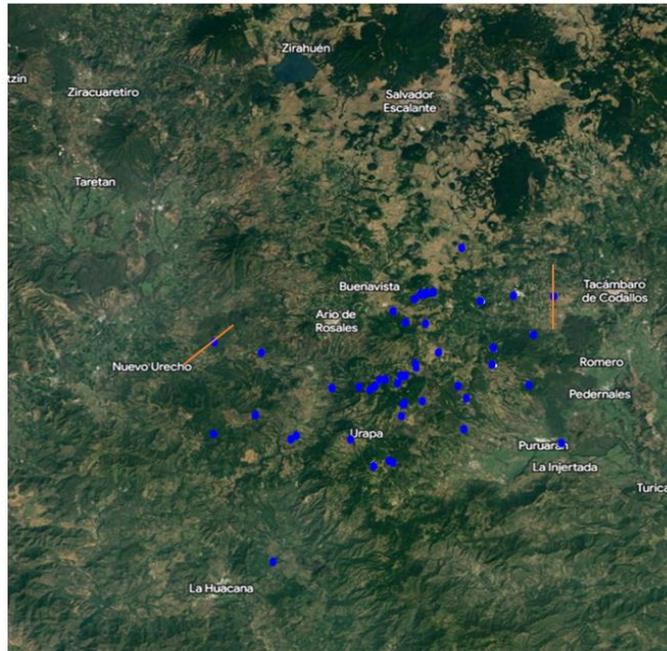
Los Domos y Spatter tienen dependencia entre su Diámetro Basal y su Edad, los Domos de forma negativa y los Spatter de forma positiva. Es decir, en Domos resulta que, a menor edad, mayor Diámetro Basal y en Spatter a mayor edad, mayor Diámetro Basal.

Sin embargo, estos últimos dos resultados (Diámetro Basal vs Edad y Altura vs Edad) para los Domos resultó ser un comportamiento local de las cópulas en probabilidades altas (Figura 6.26). Esto quiere decir que si en una de las variables se obtiene un valor cuya probabilidad en la función de distribución marginal asociada es alta, entonces con una probabilidad igual de alta la otra variable decrece, sin embargo, en tanto nos alejamos de esas probabilidades altas (tomando valores más pequeños) deja de haber asociación entre las variables. Y esto en general tiene sentido porque la mayoría de los datos utilizados se concentran en Edades menores a 1 Ma, por lo que son pocos datos aquellos con Edades Altas y eso explica que en probabilidades altas se haya encontrado dependencia. Por todo lo anterior, este resultado puede estar sesgado por la baja cantidad de datos en Edades altas.

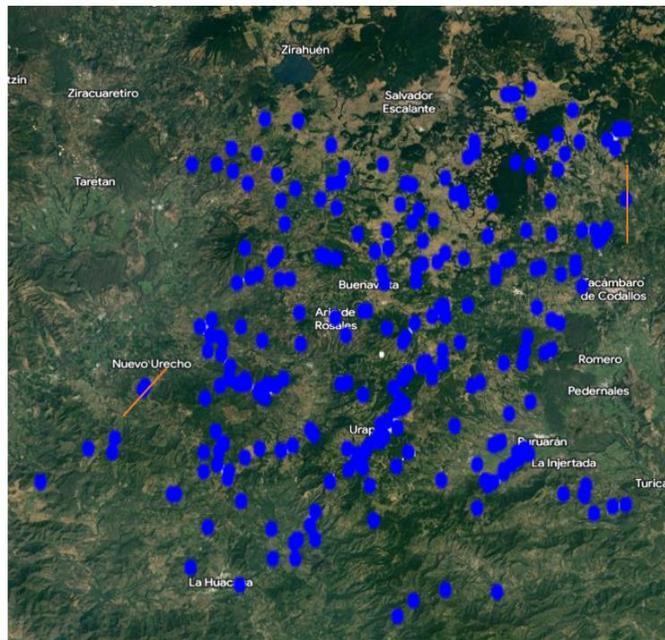
En tanto a la comparación de Localización vs Elongación resultó una dependencia positiva tanto en Conos como en Spatter, ello significa que, a mayor elongación, mayor será la longitud. En otras palabras, si la elongación es en dirección Norte-Sur (cero grados con respecto al norte), la estructura es más probable que se encuentre más cercana al Este de nuestra zona de estudio. Y entre más nos acerquemos a la zona oeste, la elongación tenderá a ser en dirección Suroeste-Noreste. Este resultado infiere, dado que la elongación es paralela a las fallas subyacentes del área, que las mismas fallas cambian de dirección conforme a la longitud.

Una forma más visual de ver estos resultados obtenidos para Conos y Spatter es la siguiente:

**a) Conos: Localización-Elongación**



**b) Spatter: Localización-Elongación**



**Figura 8.1** | Foto satelital del área de estudio con las localizaciones de las estructuras sobrepuestas (círculos azules) y el cambio de la elongación (barras naranjas) conforme a la Longitud yendo de dirección Norte-Sur en el este del campo a dirección Suroeste - Noreste en el Oeste del campo para: a) Conos, b) Spatter

En general, forma más ilustrativa de ver dichos resultados es (Tabla 8.2):

Datos	Tipología	X	Y	Marginal X	Marginal Y	P-value		θ Frank	Promedio coef corr.
						Bernstein	Kernel		
115	Conos	Latitud	Elongación	TruncatedNor	Uniforme	5.16E-01	7.35E-01	-0.06	0.03
115	Conos	Longitud	Elongación	Normal	Uniforme	8.90E-02	3.08E-04	0.81	0.16
251	Spatter	Latitud	Elongación	TruncatedNor	Uniforme	2.87E-01	7.91E-01	-0.27	-0.04
251	Spatter	Longitud	Elongación	TruncatedNor	Uniforme	2.70E-03	1.95E-01	-0.58	0.13
13	Conos Edad A	Indice de Aspe	Edad	Exponencial	Beta	5.00E-05	2.80E-01	-	-0.21
9	Conos B	Indice de Aspe	Edad	Exponencial	Laplace	5.59E-01	2.60E-08	-	-0.26
46	Conos	Latitud	Edad	Logistic	Weibul Min	4.07E-02	6.60E-01	-	-0.08
46	Conos	Longitud	Edad	Laplace	Weibul Min	3.26E-01	1.77E-01	0.01	0.04
138	Spatter	Latitud	Edad	Gumbel	Gamma	6.70E-01	1.40E-01	-0.17	0.03
138	Spatter	Longitud	Edad	NormalTrunca	Gamma	4.60E-01	1.30E-01	0.69	0.15
22	Domos	Latitud	Edad	Uniforme	LogUniforme	8.40E-02	8.20E-01	1.32	-0.01
22	Domos	Longitud	Edad	WeibullMax	LogUniforme	1.30E-01	7.40E-01	0.67	-0.10
133	Spatter	Altura	Edad	LogNormal	Gamma	6.50E-01	5.00E-01	-0.45	-0.02
126	Spatter	Diámetro Bas	Edad	NormalInvers	Gamma	7.00E-08	4.04E-05	-	0.16
21	Domos	Altura	Edad	NormalInvers	LogUniforme	1.89E-01	4.00E-04	1.09	-0.18
19	Domos	Diámetro Bas	Edad	NormalInvers	LogUniforme	3.15E-01	2.70E-03	-	-0.12

**Tabla 8.2]** Tabla de los la tipología usada, cada pareja de variables comparada, las marginales estimadas a cada variable, los p-values de las copulas no paramétricas, el parámetro de la cópula paramétrica y el promedio de los coeficientes de correlación de alguna de las cópulas no paramétricas. En verde, las comparaciones donde hubo dependencia, el coeficiente de correlación respectivo es azul si la dependencia es positiva y amarillo si es negativa. En blanco, las comparaciones donde no hay dependencia evidente, sus p-values en rojo. En naranja las comparaciones donde hubo p-values menores al nivel de significancia, sin embargo, el coeficiente de correlación promedió cercano a 0.

Cabe recordar que los resultados que llamamos independencia no implicarían que  $F_{XY} = F_X * F_Y$ . Es decir, sí existe una distribución conjunta que los simula, pero esta es el mero producto de las marginales, de forma que el comportamiento de una de las variables no afecta a la otra. Ya que las Independencias no son seguras, puesto que con las pruebas de hipótesis o se rechaza la Hipótesis nula (que en nuestro caso era la independencia) o se dice que no hay evidencia suficiente para rechazarla, lo cual no significa afirmar que esa hipótesis nula sea necesariamente verdadera.

En cuanto a las funciones marginales estimadas y ajustadas a cada variable en cada comparación, podemos ver que, dado que cuando comparamos latitud con edad en Domos nos ajusta una uniforme en la edad que puede estar reflejando la falta de datos (a pocos datos, mucha es la variación entre unos y otros sin datos intermedios, ocasionando facilidad de ajuste a una recta) y cuando comparamos altura con edad para la misma tipología nos ajusta una loguniforme (reflejando lo alejado de los datos), y además que dichas funciones no aparecieron en la edad de los Spatter para ninguna comparación. Esto nos dice lo mucho que una baja cantidad de datos puede afectar a un estudio, incluyendo el nuestro. De tener más datos en los casos en los que se tiene menos que 100 (Conos y Domos), el agregar pocos datos (como vimos en la comparación de “Conos edad ajustada” y “Conos B”) nos cambia las funciones marginales estimadas, lo cual cambia la cópula y por ende altera los resultados en los coeficientes de correlación, pudiendo cambiar los resultados de dependencia.

Por otro lado, la falta de información constituye precisamente una de las motivaciones de este estudio: el intento de llenar los vacíos de información mediante estimaciones. Entonces lo importante sería usar más datos disponibles. Así, una posible vía de investigación que este trabajo ha abierto, sería aplicar las mismas

comparaciones a dos áreas: una ligeramente mayor que la actual, lo suficiente para incorporar datos de tipologías a aquellas que ahora tuvimos poca información y una segunda prueba aplicada a una mayor escala del campo, a fin de encontrar una cantidad de datos con la cual los resultados no cambien. Con la información actual, podemos afirmar que ello sucede con cantidades mayores a 100 datos.

Esto permitiría corroborar no sólo sobre la utilidad de las cópulas para este tipo de estimaciones, cosa que este trabajo ya probó, sino también evaluar si los patrones de dependencia obtenidos con ellas son extrapolables al CVMG en general. De repetirse los resultados de forma consistente, podríamos afirmar que la muestra original fue representativa. Ello facilitaría la identificación de los factores clave en la selección de datos que permitan reproducir el estudio y generar conclusiones aplicables a otros campos volcánicos monogenéticos.

Ya se dispone de un método que permitirá trabajar con información limitada (Los ajustes para la edad de Spatter variaron menos entre comparaciones, lo cual no sucedió con otras tipologías y ello es lo que sugiere que 100 registros podrían ser suficientes). Queda por mejorar la obtención sistemática de datos para nutrir los modelos, a fin de evitar tener una base de n entradas con la mitad de entradas con la información necesaria para considerarse útil, como sucedió en este estudio al momento de, por ejemplo, no tener la misma cantidad de Domos con alturas registradas que Domos con Diámetro basal registrado. O tener 115 Conos con elongación registrada pero sólo 13 con Edad radiométrica asignada.

En resumen:

- > Cien datos parecen resultar suficientes para que las estimaciones marginales preserven propiedades en diferentes modelos.
- > Es posible deducir características de las estructuras no estudiadas a partir del modelado estocástico de una muestra relativamente pequeña ( 100 datos ) a comparación de la cantidad de estructuras existentes en el CVMG.
- > El uso de modelos probabilísticos, como los de cópulas, resulta más confiable que aquellos como los de regresión clásica para estudios de este tipo.

## Capítulo 9- Anexo

### ANEXO: GLOSARIO PROBABILIDAD Y ESTADÍSTICA

#### 1.1 Conceptos de probabilidad

##### a. Variable Aleatoria

Es una función que asigna un valor numérico a un experimento aleatorio. De esta forma, si al evento A le asignamos el número 3 en una variable aleatoria X. Al preguntarnos por si ocurrió un evento "A", podemos reducirlo a preguntar si nuestra variable aleatoria valió 3 ( $X=3$ ).

Suelen representarse con letras mayúsculas.

##### b. Espacio de probabilidad

Dado que las funciones son relaciones entre dos conjuntos (donde a todo elemento del dominio se le asocia un único elemento del contra dominio), las funciones de probabilidad están definidas entre dos conjuntos: la sigma-álgebra y el  $[0,1]$ .

##### i. Espacio

##### Muestral

Iniciamos por un conjunto Omega llamado el espacio muestral, que es el conjunto con los posibles resultados.

##### ii. Sigma-álgebra

De Omega obtenemos a nuestra sigma-álgebra, que es el conjunto de posibles eventos (incluyendo aquellos en los que hay varios resultados). Al que se denota con una *f* cursiva.

##### iii. Función

##### de

##### probabilidad

La medida o función de probabilidad "P", la cual asigna una probabilidad (un valor entre el 0 y el 1) a cada evento de nuestra sigma-álgebra.

A la unión de estos tres elementos ( $\Omega, f, P$ ) le llamamos el espacio de probabilidad.

##### c. Función de densidad

Es nuestra función de probabilidad "P", es decir:

$$f(x)=P[X=x]$$

Por tanto, la función toma valores en el  $[0,1]$ .

##### d. Función de distribución

Describe la probabilidad acumulada, es decir:

$$F(x)=P[X\leq x]$$

Aquí la función es creciente y si X toma valores  $\{x_1, \dots, x_n\}$  resulta que:

$$F(x_n)=P[X\leq x_n]=1$$

Cuando dichas funciones siguen patrones específicos, se agrupan de acuerdo a estos, es decir, se le asigna un nombre a la familia de la distribución, una expresión algebraica

a la función, y se le asigna al menos un parámetro a variar (ej: Distribución Normal, Distribución Poisson...).

En dicho caso diremos que X, en este espacio de probabilidad, “se distribuye” de tal forma.

**e. Marginales**

Cuando tenemos más de una variable aleatoria, se le llama conjunta a la función de densidad o distribución que describa ambas variables a la vez.

Y le llamaremos marginal de X, a la función de distribución o densidad que describe únicamente a X.

Es decir, nótese que, para una función de densidad conjunta de cuatro variables aleatorias, existirían cuatro funciones de densidad marginales, una por cada variable.

**f. Algunas funciones de probabilidad:**

**i. Normal  $N(\mu, \sigma)$**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Con

$\mu$ : la media

$\sigma$ : la desviación estándar.

Describe la probabilidad relativa de que una variable aleatoria siga esta distribución, su gráfica tiene forma de campana y es ampliamente utilizada en estadística, análisis de riesgo y análisis de datos.

**ii. Normal truncada<sup>55</sup>**

$$f(x) = \frac{1}{\sigma (\phi(b) - \phi(a))\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Con

$\mu$ : el parámetro de media de la distribución normal truncada

$\sigma$ : el parámetro de desviación estándar de la distribución normal truncada

a: el límite inferior del truncamiento

b: el límite superior del truncamiento

$\phi$ : la función de distribución acumulada de la distribución normal estándar

Es una distribución normal ajustada a un intervalo específico. Es un utilizada en análisis e datos censurados, modelos de elección discreta, inferencia estadística en presencia de límites y modelado de variables con valores restringidos.

**iii. Normal Inversa (GIG)<sup>56</sup>**

$$f(x) = \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \left(\frac{x}{\mu}\right)^{\lambda-\frac{1}{2}} e^{-\frac{1}{2}(\lambda x^2 + 2x\mu)}$$

Con

$\rho$ : la probabilidad

$\mu$ : el parámetro de localización de la distribución GIG

$\lambda$ : el parámetro de escala de la distribución GIG

También es llamada Gussiana Inversa Generalizada (GIG por sus siglas en inglés) y generaliza tanto la distribución normal como la gamma inversa, describiendo la probabilidad de que una variable aleatoria siga esta distribución. Es usada en teoría de colas, econometría y análisis de supervivencia.

#### iv. Exponencial Exp ( $\lambda$ )

Con

$\lambda$ : el número de veces que se espera que ocurra el evento estudiado durante un periodo de tiempo determinado.

Y función de Densidad:

$$f(x) = \lambda e^{-\lambda x}$$

Aunque el ajuste que usaremos es:

$$f(x) = \lambda e^{-\lambda(x-\gamma)}$$

Con

$\lambda$ : el parámetro de tasa o inverso de la media de la distribución exponencial

gamma: el parámetro de desplazamiento o ubicación de la distribución exponencial.

Es usada para modelar el tiempo de espera hasta que ocurra un evento en un proceso de Poisson.

#### v. Uniforme

$$f(x) = \frac{1}{b-a}$$

$$F(x) = \frac{x-a}{b-a}$$

Con

a: el límite inferior del intervalo

b: el límite superior del intervalo

Establece una distribución constante en el intervalo [a,b], es decir, les da la misma probabilidad de ocurrir a cualquier valor dentro del intervalo.

**vi. LogUniforme**

$$F(x) = \frac{\log(x) - a \log}{b \log - a \log}$$

Con

$a \log$ ,  $b \log$ : los parámetros que definen el rango de la distribución loguniforme.

En esta distribución la probabilidad es uniforme en una escala logarítmica, y es usada para modelar variables con un rango amplio como los ingresos, precios de activos financieros o tamaños de población.

**vii. Laplace**

$$f(x) = \frac{1}{2\lambda} e^{-\frac{|t-\mu|}{\lambda}}$$

Con

$\mu$  : el parámetro de ubicación o media de la distribución Laplace

$\lambda$ : el parámetro o factor de escala de la distribución Laplace

Es usada para modelar datos con valores atípicos o picos pronunciados, ya que la cola exponencial de la distribución permite capturar eventos raros o extremos.

**viii. Weibull mínima (WeibullMin)**

$$F(x) = e^{-\left(\frac{x-\text{gamma}}{\beta}\right)^\alpha}$$

Con

$\beta$  : el parámetro de escala de la distribución Weibull mínima

$\alpha$  : el parámetro de forma de la distribución Weibull mínima

gamma: el parámetro de desplazamiento de la distribución Weibull mínima

Se utiliza para modelar datos con tasa de falla decreciente con el tiempo, siendo adecuada para describir la vida útil de sistemas o componentes.

**ix. Weibull máxima (WeibullMax)**

$$F(x) = 1 - e^{-\left(\frac{x-\text{gamma}}{\beta}\right)^\alpha}$$

Con

$\beta$  : el parámetro de escala de la distribución Weibull máxima

$\alpha$  : el parámetro de forma de la distribución Weibull máxima

gamma: el parámetro de desplazamiento de la distribución Weibull máxima

Se utiliza para modelar datos positivos y asimétricos donde los valores más grandes tienen mayor probabilidad de ocurrencia. Si  $\beta = 1$  es una distribución exponencial, y si  $\beta > 1$  es una distribución de cola pesada.

#### x. Logistic

$$F(x) = \frac{1}{1 + e^{\frac{-(x-\mu)}{\beta}}}$$

Con

$\beta$  : el parámetro de escala de la distribución logística

$\mu$  : el parámetro de ubicación de la distribución logística

Su gráfica tiene forma de "S" característica, y la función es utilizada para modelar variables continuas que pueden tomar valores en un rango amplio y tienen una cola más pesada que la distribución normal. Se usa en análisis de supervivencia, modelización de variables socioeconómicas, análisis de respuesta en psicología y en teoría de la información.

#### xi. Beta

$$F(x) = \frac{B(x; \alpha, \beta) - B(a; \alpha, \beta)}{B(b; \alpha, \beta) - B(a; \alpha, \beta)}$$

Con

$B(x; \alpha, \beta)$  : la función beta incompleta que es la función de densidad beta ( $f(x)$ ) desde 0 hasta x con parámetros alfa y beta.

$a$  y  $b$  : los límites inferior y superior de la variable aleatoria x.

$$f(x) = t^{(\alpha-1)}(1-t)^{(\beta-1)}$$

Esta función proporciona la probabilidad de que una variable aleatoria siga una distribución beta. Es utilizada en modelado de proporciones, análisis de datos de proporción, pruebas de hipótesis y en el campo de la teoría de la confiabilidad.

#### xii. Gamma

$$F(x) = 1 - \Gamma(k, \lambda(x - \text{gamma}))$$

Con

$\Gamma(k, z)$  : la función gamma incompleta que es la integral de la función de densidad gamma ( $g(k, z)$ ) desde  $z$  hasta el infinito.

$k$  : el parámetro de forma de la distribución gamma

$\lambda$  : el parámetro de tasa o inverso de la media de la distribución gamma

$gamma$  : el parámetro de desplazamiento de la distribución gamma

$$g(k, z) = t^{(k-1)} e^{-z}$$

Se utiliza para calcular probabilidades y cuantiles asociados con la distribución gamma en un rango específico. Es útil en el análisis de supervivencia, modelado de tiempos de vida y teoría de colas.

#### xiii. Gumbel

$$f(x) = \frac{1}{\beta} e^{-\frac{(gamma-x)}{\beta}} e^{-e^{-\frac{(gamma-x)}{\beta}}}$$

Con

$\beta$  : el parámetro de escalado de la distribución Gumbel.

$gamma$  : el parámetro de ubicación de la distribución Gumbel.

Tiene una cola larga en el lado derecho y una caída mayor a la de la distribución normal. Se utiliza para modelar eventos o valores extremos en campos como hidrología, ingeniería, economía y ciencias sociales; siendo útil para analizar inundaciones, temperaturas máximas o mínimas y en estudios de riesgo y fiabilidad.

#### xiv. Función de densidad Empírica

Se consigue asignando a cada observación el mismo valor probabilístico, es decir, si tenemos 10 observaciones,  $f(X_i)=1/10$  para toda  $i= 1, \dots, 10$ . Y la de densidad

## 1.2 Estadística descriptiva

### a. Cuantil

El Cuantil  $p$  es el valor  $x$  tal que  $F(x)=p$  con  $F(x)$  función de distribución.

**b. Cuartil**

Los cuartiles 1,2,3 y 4 son los cuantiles 25,50,75 y 100 respectivamente.

**c. Media**

La media es la esperanza ponderada de los datos, es decir, la suma de todos entre el número de datos.

$$\text{Media} = \sum_{i=1}^n \frac{X_i}{n} \text{ con } X_1, X_2, \dots, X_n \text{ observaciones}$$

**d. Mediana**

La mediana es un estadístico de posición central que parte la distribución en dos, deja la misma cantidad de valores a un lado que a otro. Se ordenan los valores de n datos. Si es impar se toma el de en medio, si es par se toma el  $\frac{n}{2}$  - décimo, el  $\frac{n}{2} + 1$ - décimo la mediana resulta el promedio de estos dos datos.

$$\text{Mediana} = \frac{\frac{X_n + X_{\frac{n}{2}+1}}{2}}{2}$$

**e. Varianza**

Es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

$$\text{Var} = \sigma^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n}$$

o bien

$$\text{Var} = E(X^2) - E(X)^2$$

**f. Desviación Estándar**

Es una medida que se utiliza para cuantificar la variación o la dispersión de un conjunto de datos numéricos. Nos dice qué tanto se alejan los datos de la media.

$$\sigma = \sqrt{\text{Var}}$$

**ANEXO: Implementación Computacional**

Para las estimaciones, cálculos, simulaciones y gráficos se empleará un programa en Python y el editor de textos PyCharm ya que es un entorno de desarrollo integrado (IDE) específico para este lenguaje, lo que facilita la lectura y corrección del programa. Este software fue creado por Díaz-Viera<sup>57</sup> y modificado para el trabajo actual, y funciona de la siguiente manera:

1. Importar una base de datos de un csv, guardándolo en una variable con el nombre de sample
2. Definir la función de grafica de dispersión que pida coordenadas (x,y), y los coeficientes de Pearson, spearman, kendall, titulo, y rangos de la gráfica.
3. Crear las funciones (def) graficadoras para las marginales y una función para guardar las imágenes en pdf
4. Construir un árbol (con tkinter as tk)
5. Guardar una variable para saber la dimensión (y una función auxiliar en caso de 3 dimensiones)
6. Extraer los nombres de las variables del primer recuadro del Excel
7. Imprimir el primer resumen :
  - a. Tamaño de la muestra

- b. Matriz de títulos de las variables
- c. Mínimo
- d. Primer cuartil
- e. Promedio
- f. Mediana
- g. 3er cuartil
- h. Máximo
- i. Rango
- j. Varianza
- k. Desviación estándar
- l. Oblicuidad (skewness)
- m. Kurtosis

Todo lo anterior en matrices de la misma dimensión que los datos

8. Imprimir el segundo resumen
  - a. Covariancia
  - b. Desviación estándar
  - c. Matriz de correlación de Pearson
  - d. Matriz de correlación de kendall
  - e. Matriz de correlación de spearman
9. Guardar la dispersión de los datos
10. Obtener los histogramas de la distribución de cada variable además de la matriz de los mejores modelos BIC para cada marginal y la imprimimos como “distribución P de la variable i”.
11. Graficar la marginal teórica o estimada contra la empírica
12. Guardar en la variable “best estimated P” al BIC de cada marginal
13. Crear el histograma de la primera variable contra la misma P estimada
14. Repetir lo mismo con la segunda variable y su P correspondiente, creando así “Figures parametric vs empirical pdf”, uno para cada variable.
15. Definir una función que nos regresa la lista de índices ordenados cuya covarianza entre el anterior y el actual sea estrictamente positiva, la llamamos “find\_neighbours”
16. Crear otra función que, dada una covarianza de n variables, crea una lista, la ingresa a la función anterior, y nos ordena el resultado le llamaremos “connected-components”
17. Calcular la correlación de spearman de la base de datos denominada “sample” (en punto 1)
18. Por cada dimensión (cantidad de variables), iteramos sobre ambas entradas de nuestra correlación de spearman y si su valor absoluto es mayor a uno entre la raíz cuadrada del tamaño de la muestra, entonces hacemos esa entrada de la correlación 1, y 0 en otro caso
19. Ingresar la nueva correlación obtenida en el paso anterior como la covarianza en la función “connected-components” (paso 16) y esos serán los *bloques*  
*NOTA: Dichos bloques casis siempre saldrán el primero 0 y el segundo 1*
20. Dentro de la librería openturns, llamamos a una lista de distribuciones continuas, iteramos sobre cada una y sobre las que sean copulas checamos que su nombre no sea la de Bernstein, y entonces la añadimos a una lista llamada “copulaDactories”.
21. Crear una lista de copulas llamada “estimated P copulas” y le añadimos una lista por cada bloque con:
  - a. El BIC de:

- i. Las marginales conforme al parámetro bloque
- ii. Una copula de Frank generada por la librería openturns

NOTA y ejemplo:

Si mi correlación de spearman resulta

```
[[ 1    -0.477498 ]
 [ -0.477498 1    ]]
```

Tras hacer el paso 22 se vuelve

```
[[ 1 1 ]
 [ 1 1 ]]
```

Y los blocs resultantes son

```
[[0, 1]]
```

Si mi correlación de spearman resulta

```
[[ 1    0.00132646 ]
 [ 0.00132646 1    ]]
```

Tras hacer el paso 22 se vuelve

```
[[ 1 0 ]
 [ 0 1 ]]
```

Y los blocs resultantes son [[0], [1]]

Entonces falla el programa, fue necesario pasarlo a [[0, 1]] para continuar

22. Guardar el Best estimated copula (BIC) que resulto del paso anterior
23. Convertir en una copula compuesta con la librería openturn
24. Obtener las marginales de la permutación inversa de cada bloque multiplicado por la cantidad de variables (dimensión)
25. Obtener la Copula P estimada de las marginales del paso anterior
26. Calcular la distribución P estimada de la composición entre nuestras marginales P estimadas (las originales) y la copula P estimada
27. Imprimir la Distribución P estimada bajo el nombre “Mejor distribución paramétrica estimada”
28. Con la librería openturn, simular 337 resultados de nuestra distribución P estimada. Nota: 337 es un número arbitrario, es común elegir el número correspondiente a la cardinalidad de la base de datos.
29. Repetir los pasos 10 y 11 de estadística descriptiva con nuestros resultados de la simulación
30. Guardar los resultados de la simulación de la cópula paramétrica en una gráfica
31. Para el modelado no paramétrico con kernel se estiman dos variables con kernel Smoothing (KS) Epanechnikov
32. Comparar lo obtenido con la Empírica de nuestros datos tanto en función de distribución como en los datos obtenidos de una simulación igual de tamaño.
33. Graficar bajo el nombre de “estimación no paramétrica de la cópula”
34. Ahora pasaremos a la estimación no paramétrica de Bernstein, igual con la librería openturn y a partir de una distribución compuesta con nuestras variables estimadas usando a kernel como parámetro
35. Las estadísticas de los pasos 10 y 11 son obtenidas a partir de simulaciones de la nueva distribución de Bernstein (paso anterior)
36. Se comparan las distribuciones, obteniendo las gráficas relacionadas con la Cópula Bernstein.

Nota: en cada cópula se duplico su rango de valores para facilitar su visualización y computo.

```
#Graficamos la copula
myGraph_KS = ot.Graph('Non-parametric estimation of the copula', label_var1_label_var2, True, 'topleft')
myGraph_KS.setLegendPosition('bottomleft')
myGraph_KS.add(rankCloud)
minPoint = [0.0]*2
maxPoint = [1.0]*2
pointNumber = [201]*2
graphCop = KernelSmoothing_copula_distribution.drawPDF(minPoint, maxPoint, pointNumber)
contour_estCop = graphCop.getDrawable(0)
contour_estCop.setDrawLabels(False)
nlev = 21
levels = ot.Point(nlev)
for i in range(nlev):
    levels[i] = 0.25 * nlev / (nlev - i)
contour_estCop.setLevels(levels)
contour_estCop.setLegend('Kernel smoothing copula')
contour_estCop.setColor('red')
contour_estCop.setDrawLabels(True)
myGraph_KS.add(contour_estCop)
view = otv.View(myGraph_KS)
view.save("Figures"+"Non-parametric_Kernel_Smoothing_copula_scatterplot.png", dpi=100)
```

**Figura 9.1** | Código de la gráfica de la cópula en python

## Capítulo 10- Lista de Figuras

### Figuras

- Figura 1.1- Ejemplo de Regresión Lineal.** Edad contra Índice de aspecto. Imagen de autoría propia. (Hco\_Wco es la Altura entre el Diámetro basal)
- Figura 1.2- CVTM.** Mapa de elevación del límite del Cinturón Volcánico Mexicano y la zona de estudio está dentro del recuadro rojo. Lo limitado de negro es el CVTM, el límite rojo corresponde al CVMG y los triángulos rojos son los principales volcanes de faja Transmexicana. Fuente: Macías et al (2012) "Paricutín setenta años de historia"
- Figura 1.3- Modelo digital del terreno** del centro de México que muestra: a) Marco tectónico zona oeste. Esquema de la porción oeste de la FVTM (modificado de "Geología e historia eruptiva de algunos de los grandes volcanes activos de México" J.L.Macías,2005; DeMets y Stein, 1991; Lange y Carmichael, 1991 y Luhr et al., 1989). Las estructuras volcánicas son: 1) San Juan, 2) Sangangüey, 3) Tepetitlic, 4) Ceboruco, 5) Tequila, 6) Caldera de la Primavera. Complejo volcánico de Colima: 7) Cántaro, 8) Nevado de Colima y 9) volcán de Colima. Las abreviaciones son: RTZ-rift Tepic-Zacoalco, GCh-graben de Chapala, GCi graben de Citala;b) Regiones volcánicas que se han estudiado dentro del CVMG. Las referencias están insertadas en la imagen. Fuente: Gómez-Cálderón (2021)
- Figura 1.4- Localización** a) Las estructuras volcánicas con datos disponibles en la zona sureste del campo. Imagen de autoría propia. b) Imagen Satelital del área de estudio
- Figura 2.1- Morfologías 1.** Imagen por Simkin, T- and Siebert, L. 1994.
- Figura 2.2- Morfologías 2** "Estructuras y morfologías de formas volcánicas explosivas pequeñas." Por INSUGEO (Instituto Superior de Correlación Geológica)
- Figura 2.3- Medida H/W:** Wcr: Diámetro del cráter. Wco: Diámetro basal del cono. Hco: altura del cono
- Figura 2.4- Elongación:** Son los grados respecto al norte que se encuentra el eje mayor de la base. 1)La estructura volcánica 2) La base y sus diámetros, 3) La base tiende a tener un diámetro más grande que el resto, 4) Respecto a este diámetro (o eje mayor) se toman los grados conforme al Norte en sentido horario. Figura hecha por Esquivel A. Adriana para el presente trabajo.
- Figura 5.1- Elongación:** Esquema de corrección del ángulo de elongación
- Figura 6.1 |** Funciones de Distribución Acumulativa (CDF) de las variables de "Conos B" (sin datos menores a 20,000 años)  
a)Edad b)Índice de Aspecto
- Figura 6.2 |** Gráfica de la Cópula de "Conos B" (Sin edades menores a 80,000): a)Bernstein b)Kernel Smoothing c)Frank
- Figura 6.3|** Dispersión de Edad contra H/W de "Conos B" (sin edades menores a 20,000) usando cada ajuste de cópula a)Bernstein, b)Kernel, c)Frank)
- Figura 6.4 |** Funciones de Distribución Acumulativa (CDF) de las variables de "Conos Edad Ajustada: a)Edad b)Índice de Aspecto
- Figura 6.5|** Gráfica de la Cópula de "Conos Edad Ajustada": a)Bernstein b)Kernel Smoothing c)Frank
- Figura 6.6 |** Dispersión de Edad contra H/W de "Conos edad Ajustada" usando el ajuste de cópula ( a) Bernstein, b) Kernel, c) Frank)
- Figura 6.7|** Gráfico de dispersión Edad - Longitud
- Figura 6.8 |** Funciones de Distribución Acumulativa (CDF) de las variables de "Conos Edad vs Longitud": a)Edad b)Longitud
- Figura 6.9|** Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing, c) Frank en el conjunto de datos "Conos vs Longitud".
- Figura 6. 10|** Gráfica de la Cópula de a)Kernel (no-paramétrica) ajustada a los datos "Conos: Edad vs Longitud" b) La cópula de Frank (paramétrica) ajustada al mismo conjunto de datos.
- Figura 6. 11|** Gráfico de dispersión Edad - Latitud

**Figura 6.12** | Funciones de Distribución Acumulativa (CDF) de las variables de "Conos Edad vs Latitud": a) Edad b) Latitud

**Figura 6.13** | Gráfica de: a) la Cópula de Bernstein ajustada a los datos "Conos: Edad vs Latitud"; b) Simulación respectiva

**Figura 6.14** | Funciones de Distribución Acumulativa (CDF) de las variables de "Spatter Longitud vs Edad": a) Longitud; b) Edad

**Figura 6.15** | Funciones de Distribución Acumulativa (CDF) de las variables de "Domos Longitud vs Edad": a) Longitud; b) Edad

**Figura 6.16** | Spatter Edad vs Longitud". a) Gráfica de la cópula de Kernel; b) Datos originales contra los simulados con la Cópula de Kernel Smoothing

**Figura 6.17** | Gráfica de las Cópulas de: a) Bernstein; b) Kernel; c) Frank; en el conjunto "Domos Longitud vs Edad"

**Figura 6.18** | Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing c) Frank en el conjunto de datos "Domos Edad vs Longitud".

**Figura 6.19** | Funciones de Distribución Acumulativa (CDF) de las variables de "Spatter Latitud vs Edad": a) Latitud; b) Edad

**Figura 6.20** | Funciones de Distribución Acumulativa (CDF) de las variables de "Domos Latitud vs Edad": a) Latitud; b) Edad

**Figura 6.21** | Gráfica de las Cópulas de: a) Bernstein; b) Kernel; c) Frank; en el conjunto "Spatter Latitud vs Edad"

**Figura 6.22** | Resultado de los datos originales contra los simulados con la Cópula de: a) Kernel Smoothing b) Frank en el conjunto de datos "Spatter Edad vs Latitud".

**Figura 6.23** | Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing c) Frank en el conjunto de datos "Domos Edad vs Latitud".

**Figura 6.24** | Funciones de Distribución Acumulativa (CDF) de las variables de "Spatter Altura vs Edad": a) Altura; b) Edad

**Figura 6.25** | Funciones de Distribución Acumulativa (CDF) de las variables de "Domos Altura vs Edad": a) Altura; b) Edad

**Figura 6.26** | Gráfica de la cópula de Kernel en los conjuntos: a) Spatter b) Domo; en "Altura vs Edad"

**Figura 6.27** | Resultado de los datos originales contra los simulados con la Cópula de Kernel Smoothing en el conjunto de: a) Spatter, b) Domos; en "Edad vs Longitud".

**Figura 6.28** | Gráfica de la cópula de Frank en los conjuntos: a) Spatter b) Domo; en "Altura vs Edad"

**Figura 6.29** | Funciones de Distribución Acumulativa (CDF) de las variables de "Spatter Diámetro Basal vs Edad": a) DiamBasal; b) Edad

**Figura 6.30** | Funciones de Distribución Acumulativa (CDF) de las variables de "Domos Diámetro Basal vs Edad": a) DiamBasal; b) Edad

**Figura 6.31** | "Spatter Diámetro Basal vs Edad" con la Cópula de Bernstein: a) Gráfica de la cópula; b) Datos originales contra los simulados

**Figura 6.32** | Gráfica de la cópula de: a) Bernstein, b) Kernel, c) Frank; para el conjunto de Domos "Diámetro Basal vs Edad"

**Figura 6.33** | Resultado de los datos originales contra los simulados con la Cópula de: a) Bernstein, b) Kernel Smoothing c) Frank en el conjunto de datos "Domos Diámetro Basal vs Edad".

**Figura 6.34** | Funciones de Distribución Acumulativa (CDF) de las variables de "Conos Localización vs Elongación": a) Latitud b) Longitud c) Elongación

**Figura 6.35** | Gráfico de dispersión de latitud-longitud de los conos.

**Figura 6.36** | Gráfica de las Cópulas de Frank: a) En el conjunto Latitud vs Elongación; b) En el conjunto Longitud vs Elongación.

**Figura 6.37** | Gráfica de las Cópulas de: a) Bernstein en el conjunto Latitud vs Elongación; b) Kernel en el conjunto Latitud vs Elongación; c) Bernstein en el conjunto Longitud vs Elongación; d) Kernel en el conjunto Longitud vs Elongación.

**Figura 6.38** | Resultado de los datos originales contra los simulados con la Cópula de Kernel para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación

**Figura 6.39** | Resultado de los datos originales contra los simulados con la Cópula de Bernstein para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación

**Figura 6.40**| Funciones de Distribución Acumulativa (CDF) de las variables de “Spatter Localización vs Elongación”: a) Latitud b) Longitud c) Elongación

**Figura 6.41**| Gráficos de las cópulas: a) De Bernstein para Latitud-Elongación, b) De Bernstein para Longitud-Elongación, c) De Frank para Latitud- Elongación, d) De Kernel para Longitud-Elongación, e) De Frank para Latitud-Elongación, f) De Frank para Longitud- Elongación.

**Figura 6.42**| Resultado de los datos originales contra los simulados con la Cópula de Bernstein para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación, en Spatter

**Figura 6.43**| Resultado de los datos originales contra los simulados con la Cópula de Kernel para los conjuntos: a) Latitud vs Elongación, b) Longitud vs Elongación, en Spatter

**Figura 7.1**| Grafica de la regresión lineal del conjunto “Conos B” al comparar Índice de Aspecto con Edad

**Figura 7.2**| Grafica de la regresión cuadrática del conjunto “Conos B” al comparar Índice de Aspecto con Edad

**Figura 7.3**| Grafica de la regresión cúbica del conjunto “Conos B” al comparar Índice de Aspecto con Edad

**Figura 7.4**| Grafica de la regresión lineal del conjunto “Conos Edad Ajustada” al comparar Índice de Aspecto con Edad

**Figura 7.5**| Grafica de la regresión cuadrática del conjunto “Conos Edad Ajustada” al comparar Índice de Aspecto con Edad

**Figura 7.6**| Grafica de la regresión cúbica del conjunto “Conos Edad Ajustada” al comparar Índice de Aspecto con Edad

**Figura 7.7**| Grafica de la regresión lineal del conjunto “Latitud-Elongación” al comparar Localización contra Elongación

**Figura 8.1** | Foto satelital del área de estudio con las localizaciones de las estructuras sobrepuestas (círculos azules) y el cambio de la elongación (barras naranjas) conforme a la Longitud yendo de dirección Norte-Sur en el este del campo a dirección Suroeste - Noreste en el Oeste del campo para: a) Conos, b) Spatter

**Figura 9.1** | Código de la gráfica de la cópula en Python

## Tablas

**Tabla 6.1 – Estadísticas de Conos** Resultados de medidas estadísticas con la lista original de datos de Conos: “Conos Edad Ajustada” (todos los datos) y “Conos B” (sin datos menores a 20,000 años). En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

**Tabla 6.2 – Estadísticas de los datos Simulados** a con la cópula de Kernel y la Cópula de Bernstein del conjunto Conos B. En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

**Tabla 6.3 – Estadísticas de los datos Simulados** a con la cópula de Kernel y la Cópula de Bernstein del conjunto Conos Edad Ajustada. En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

**Tabla 6.4 – Estadísticas de Conos con la cópula de Bernstein** ajustada a datos de “Conos edad ajustada” (todos los datos) y “Conos B” (sin datos menores a 20,000 años). En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

**Tabla 6.5 – Estadísticas de Conos con la cópula de Kernel** ajustada a datos de “Conos edad ajustada” (todos los datos) y “Conos B” (sin datos menores a 20,000 años). En cada columna las estadísticas están en orden [Índice de Aspecto, Edad]

**Tabla 6.6** | Estadísticas de Conos vs longitud de datos originales

**Tabla 6.7** | Estadísticas de Conos vs longitud con la cópula de Kernel, Bernstein y Frank.

**Tabla 6.8** | Estadísticas de los Conos en Edad vs Latitud: univariado

**Tabla 6.9** | Estadísticas de los Conos en Edad vs Latitud: con cópulas no paramétricas

**Tabla 6.10** | Estadísticas de Spatter y Domos para el conjunto de Longitud vs Edad

**Tabla 6.11** | Estadísticas de Spatter y Domos con la cópula de Kernel ajustada a datos de Longitud-Edad

**Tabla 6.12| Estadísticas de Spatter y Domos con la cópula de Bernstein ajustada a datos de Longitud-Edad**

**Tabla 6. 13| Estadísticas de Spatter y Domos para el conjunto de Latitud vs Edad**

**Tabla 6.14| Estadísticas de Spatter y Domos con la cópula de Kernel ajustada a datos de Latitud-Edad**

**Tabla 6.15| Estadísticas de Spatter y Domos con la cópula de Bernstein ajustada a datos de Latitud-Edad**

**Tabla 6.16| Estadísticas de Spatter y Domos con la cópula de Frank ajustada a datos de Latitud-Edad**

**Tabla 6.17| Estadísticas de Spatter y Domos para el conjunto de Altura vs Edad**

**Tabla 6.18| Estadísticas de Spatter y Domos con la cópula de Kernel Smoothing para el conjunto de Altura vs Edad**

**Tabla 6.19| Estadísticas de Spatter y Domos con la cópula de Bernstein para el conjunto de Altura vs Edad**

**Tabla 6.20| Estadísticas de Spatter y Domos para el conjunto de DiamBasal vs Edad**

**Tabla 6.21| Estadísticas de Spatter y Domos con la cópula de Kernel Smoothing para el conjunto de DiamBasal vs Edad**

**Tabla 6.22| Estadísticas de Spatter y Domos con la cópula de Bernstein para el conjunto de DiamBasal vs Edad**

**Tabla 6.23| Estadísticas de Conos Resultados** de medidas estadísticas de los conjuntos de datos Elongación-Latitud y Elongación-Longitud

**Tabla 6. 24| Estadísticas de Conos con la cópula de Kernel** ajustada a datos de Latitud-Elongación y Longitud-Elongación

**Tabla 6.25| Estadísticas de Conos con la cópula de Bernstein** ajustada a datos de Latitud-Elongación y Longitud-Elongación

**Tabla 6.26| Estadísticas de Spatter** de los datos Latitud-Elongación y Longitud-Elongación

**Tabla 6. 27| Estadísticas de Spatter con la cópula de Bernstein** ajustada a datos de Latitud-Elongación y Longitud-Elongación

**Tabla 6. 28| Estadísticas de Spatter con la cópula de Kernel** ajustada a datos de Latitud-Elongación y Longitud-Elongación

**Tabla 8.2|** Tabla de los la tipología usada, cada pareja de variables comparada, las marginales estimadas a cada variable, los p-values de las copulas no paramétricas, el parámetro de la cópula paramétrica y el promedio de los coeficientes de correlación de alguna de las cópulas no paramétricas. En verde, las comparaciones donde hubo dependencia, el coeficiente de correlación respectivo es azul si la dependencia es positiva y amarillo si es negativa. En blanco, las comparaciones donde no hay dependencia evidente, sus p-values en rojo. En naranja las comparaciones donde hubo p-values menores al nivel de significancia, sin embargo, el coeficiente de correlación promedió cercano a 0.

**Tabla 8.1|** Tabla de los la tipología usada, cada pareja de variables comparada, las marginales estimadas a cada variable, los p-values de las copulas no paramétricas, el parámetro de la cópula paramétrica y el promedio de los coeficientes de correlación de alguna de las cópulas no paramétricas.

## Capítulo 11- Bibliografía

- <sup>1</sup> Koenker, R. (2005). Quantile regression. Cambridge university press. Pág. 18.
- <sup>2</sup> Gómez Martínez Y.,2020, "Métodos metaheurísticos de optimización global para la simulación estocástica basada en cópulas de propiedades petrofísicas a partir de atributos sísmicos elásticos" Tesis, Maestro en Ciencias de la Tierra
- <sup>3</sup> Joe, H. (2014) Dependence Modeling with Copulas. Chapman and Hall/CRC Monographs on Statistics and Applied Probability. CRC Press
- <sup>4</sup> Nelsen, R.B. (2006) An introduction to copulas. Springer Science & Business Media
- <sup>5</sup> Hernandez Maldonado V.M.,2014. Simulación Estocástica Espacial de Propiedades Petrofísicas usando cópulas de Béstein. Tesis, Dr en ingeniería.
- <sup>6</sup> Gómez Calderón D. Alberto, 2018. Relación morfométrica y estructural del vulcanismo mono/poligenético en la porción sureste del Campo Volcánico Michoacán Guanajuato. Tesis, Licenciado en Geociencias.
- <sup>7</sup> Haraldur Sigurdsson, et al. The Encyclopedia of Volcanoes (2 ed.). (Valentine & Connor,2015 es el capítulo 23)
- <sup>8</sup> Perez-Lopez Raul, 2011, Scaling laws of the size-distribution of monogenetic volcanoes within the Michoacán-Guanajuato Volcanic Field (Mexico).
- <sup>9</sup> Velasco-Tapia F. 2001, Estado actual de la investigación geoquímica en el campo monogenético de la Sierra de Chichinautzin: análisis de información y perspectivas.
- <sup>10</sup> Gómez-Vasconcelos, M.G., 2018. El volcán Parícutín en el campo volcánico Michoacán-Guanajuato: una revisión. Ciencia Nicolaita. Este volumen.
- <sup>11</sup> Smithsonian Institution, Global Volcanism Program. (s.f.). Volcanoes of the World.
- <sup>12</sup> Gómez Calderón D. Alberto, 2021. "Coexistencia de magmas shoshoníticos y calciocalinos en la Región Volcánica Graben del Cahulote de Santa Ana –Pedernales, Michoacán, México. Petrografía, geoquímica e implicaciones". Tesis, Maestro en Ciencias de la Tierra
- <sup>13</sup> Hasenaka, T., Carmichael, I.S.E., 1985, J. Volcanol. Geotherm. Res. 25, 105-124
- <sup>14</sup> Demant, A. (1981). Interpretación geodinámica del vulcanismo del Eje Neovolcánico Transmexicano. Revista Mexicana de Ciencias Geológicas, 5(2), 217-222
- <sup>15</sup> Connor, C. B. (1987). Structure of the Michoacán-Guanajuato volcanic field, Mexico. Journal of Volcanology and Geothermal Research, 33(1-3), 191-200.
- <sup>16</sup> Marcial Basilio Violeta J., 2018, Cartografía y estratigrafía del área volcánica La Trampa-Coapa, Campo Volcánico Michoacán-Guanajuato. Pág 11.
- <sup>17</sup>Gómez Calderón D. Alberto, 2021. "Coexistencia de magmas shoshoníticos y calciocalinos en la Región Volcánica Graben del Cahulote de Santa Ana –Pedernales, Michoacán, México. Petrografía, geoquímica e implicaciones". Tesis, Maestro en Ciencias de la Tierra.
- <sup>18</sup> Corona Chávez, 2018, Revisitando los volcanes de la región del Parícutín: morfología, petrología y edad. Pág 37 y 47
- <sup>19</sup> Mikusinski, J. (1959). "Operational Calculus." Pergamon Press.
- <sup>20</sup> Sklar, A. (1959). "Fonctions de répartition à n dimensions et leurs marges." Publications de l'Institut de Statistique de l'Université de Paris, 8, 229-231.
- <sup>21</sup> Gómez-Tuena, A., Mori, L., & Straub, S. M. (2018). Geochemical and petrological insights into the tectonic origin of the Transmexican Volcanic Belt. Earth-Science Reviews, 183, 153- 181.
- <sup>22</sup> Ferrari, L., Orozco-Esquivel, T., Manea, V., & Manea, M. (2012). The dynamic history of the Trans-Mexican Volcanic Belt and the Mexico subduction zone. Tectonophysics, 522, 122-149
- <sup>23</sup> Gómez Calderón D. Alberto, 2021. "Coexistencia de magmas shoshoníticos y calciocalinos en la Región Volcánica Graben del Cahulote de Santa Ana –Pedernales, Michoacán, México. Petrografía, geoquímica e implicaciones". Tesis, Maestro en Ciencias de la Tierra.

- 
- <sup>24</sup> Jackson et al., 2005, "Glossary of Geology" de la Sociedad Geológica de América
- <sup>25</sup> Istituto Nazionale di Geofisica e Vulcanologia. (s.f.). Osservatorio Vesuviano. Recuperado el 26 de mayo de 2023, de <https://www.ov.ingv.it/en/>
- <sup>26</sup> United States Geological Survey. (s.f.). Mount St. Helens. U.S. Department of the Interior. Recuperado el 26 de mayo de 2023, de <https://www.usgs.gov/volcanoes/mount-st-helens>
- <sup>27</sup> National Geographic Society. (s.f.). Mount Pinatubo. National Geographic. Recuperado el 26 de mayo de 2023, de <https://www.nationalgeographic.org/encyclopedia/mount-pinatubo/>
- <sup>28</sup> Erik Klemetti, 2018, artículo sobre la duración de las erupciones volcánicas en la revista Wired
- <sup>29</sup> Simon A. Carn, 2013, Volcanic Eruptions and Their Repose, Unrest, Precursors, and Timing
- <sup>30</sup> Keith Putirka, 2015, "The Encyclopedia of Volcanoes"
- <sup>31</sup> Departamento de Petrología y Geoquímica (Universidad Complutense de Madrid), 2014, Atlas de rocas ígneas.
- <sup>32</sup> Velazco Tapia 2003, El campo volcánico chichinautzin. Pag 27
- <sup>33</sup> Gómez Alberto, 2019, El volcán Jorullo a 260 años de su nacimiento
- <sup>34</sup> Alejandro José Toselli, 2020, Elementos básicos de petrología ígnea
- <sup>35</sup> José Luis Macías, 2005, Geología e historia eruptiva de algunos de los grandes volcanes activos de México
- <sup>36</sup> Haraldur Sigurdsson, et al. The Encyclopedia of Volcanoes (2nd ed., pp. 359-367). Academic Press.
- <sup>37</sup> Erdely, A., 2003, Construcción de medidas de dependencia por medio de cópulas. Tesis UNAM.
- <sup>38</sup> Wood, C. A. (1980). Bulletin of Volcanology, 43(1), 5-9.
- <sup>39</sup> Johnson, Robert y Patricia Kuby. Estadística elemental. 11a edición. 2012.
- <sup>40</sup> Wasserman, L. (2004). All of statistics: A concise course in statistical inference. Springer Science & Business Media.
- <sup>41</sup> Freedman, D., Pisani, R., & Purves, R. (2007). Statistics (4th ed.). W. W. Norton & Company.
- <sup>42</sup> Dennis G. Zill y Warren S. Wright., 2009, Cálculo de Varias Variables
- <sup>43</sup> Erdely, A. "Construcción de medidas de dependencia por medio de cópulas" Tesis UNAM. 2003.
- <sup>44</sup> Nelsen, R. B. (2006). An introduction to copulas. Springer Science & Business Media.
- <sup>45</sup> Schwarz, G. (1978). Estimating the dimension of a model.
- <sup>46</sup> V. H. Le, M. A. Díaz-Viera, D. Vázquez-Ramírez, R. del Valle-García, A. Erdely, D. Grana, Bernstein copula-based spatial cosimulation for petrophysical property prediction conditioned to elastic attributes, Journal of Petroleum, Science and Engineering 193, ISSN: 0920-4105(print),1873-4715(online), (<https://doi.org/10.1016/j.petrol.2020.107382>) 2020.
- <sup>47</sup> The Estimation of Copulas: Theory and Practice, 2007, Arthur Charpentier
- <sup>48</sup> Nolan, J. P., & Pollard, D. (1987). Product kernel estimators of multivariate densities. The Annals of Statistics, 15(3), 729–737.
- <sup>49</sup> Nelsen, R. B. (2006). An introduction to copulas. Springer Science & Business Media.
- <sup>50</sup> Joe, H. (2014). Dependence modeling with copulas
- <sup>51</sup> R.J.Secor, 1999, Aconcagua: A climbing guide.
- <sup>52</sup> M. A. Díaz-Viera, A. Erdely, T. Kerdan, R. del-Valle-García and F. Mendoza-Torres, "Bernstein copula-based spatial stochastic simulation of petrophysical properties using seismic attributes as secondary variable", Geostatistics Valencia 2016, Editors: Gómez-Hernández, J.J., Rodrigo-Illarri, J., Rodrigo-Clavero, M.E., Cassiraga, E., Vargas-Guzmán, J.A. (Eds.), Quantitative Geology and Geostatistics, Vol. 19, Geostatistics Valencia 2016, ISBN 978-3-319-46818-1, 1057 pages, ([https://doi.org/10.1007/978-3-319-46818-8\\_33](https://doi.org/10.1007/978-3-319-46818-8_33)) pp. 487-504, March 14, 2017.
- <sup>53</sup> Van Huong Le (2021). "Copula Based Modeling For Petrophysical Property Prediction Using Seismic Attributes As Secondary Variables" Tesis, Doctor en ciencias de la Tierra.
- <sup>54</sup> Gómez Calderón D. Alberto, 2021. "Coexistencia de magmas shoshoníticos y calciocalinos en la Región Volcánica Graben del Cahulote de Santa Ana –Pedernales, Michoacán, México. Petrografía, geoquímica e implicaciones". Tesis, Maestro en Ciencias de la Tierra
- <sup>55</sup> Roberto Porto Lorena, 2021, Distribución Normal Truncada. Recuperado el 28 de Junio de 2023, de <https://roberto-portolorena.medium.com/distribuci%C3%B3n-normal-truncada-95d88f99a9ac>

---

<sup>56</sup> Statologos,2023, Distribución normal inversa: definición y ejemplo. Recuperado el 28 de Junio de 2023, de <https://statologos.com/distribucion-normal-inversa/>

<sup>57</sup> Martín A. Díaz-Viera, Farhid M. Elisea Guerrero, 2023,CURSO "Ciencia de Datos Aplicada a Ciencias de la Tierra",Posgrado de Ciencias de la Tierra, UNAM.