



# **ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**

**FACULTAD DE CIENCIAS**

**ESCUELA DE FÍSICA Y MATEMÁTICA**

## **“ESTRUCTURA DE LAS COMUNIDADES DE LEPIDOPTERAS EN FUNCIÓN DE VARIABLES CLIMÁTICAS Y HÁBITAT NATURAL Y ANTRÓPICO MEDIANTE ANÁLISIS DE INTERDEPENDENCIA, ALTOS MIRANDINOS VENEZUELA, 2013”**

**Trabajo de Titulación**

**Tipo: Proyecto de investigación**

Presentado para obtener el grado académico de:

**INGENIERA EN ESTADÍSTICA INFORMÁTICA**

**AUTORA: GINA ABIGAIL CHÁVEZ CASTILLO**

**TUTORA: ING. JESSICA ALEXANDRA MARCATOMA TIXI.**

Riobamba – Ecuador

2018

**© 2018, Gina Abigail Chávez Castillo**

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autora.

**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**ESCUELA DE FÍSICA Y MATEMÁTICA**

El Tribunal del Trabajo de Titulación, certifica que: El trabajo de investigación: **“Estructura de las comunidades de lepidopteras en función de variables climáticas y hábitat natural y antrópico mediante análisis de interdependencia, Altos Mirandinos Venezuela, 2013”**, de responsabilidad de la señorita Gina Abigail Chávez Castillo, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, quedando autorizada su presentación.

Firma

Fecha

Ing. Jessica Alexandra Marcatoma.Tixi

**DIRECTORA DE TESIS**

\_\_\_\_\_

Ing. Amalia Isabel Escudero Villa

**MIEMBRO DE TRIBUNAL**

\_\_\_\_\_

Yo, Gina Abigail Chávez Castillo, certifico que el presente trabajo de titulación es original y que todas las ideas, opiniones y contenidos expuestos en el presente son de exclusiva responsabilidad de la autora y el patrimonio intelectual del trabajo de titulación pertenecen a la ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO. Los contenidos en el documento se encuentran citados y referenciados.

Gina Abigail Chávez Castillo.

## **DEDICATORIA**

Dedico este trabajo con muchísimo amor a mi madre Silvia Esperanza Castillo, por ser la voz que me impulsaba, cuando creía que no era capaz. A ti madre bella te dedico la culminación de esta etapa de mi vida y aquí estoy, siendo tu orgullo y diciéndote la hermosa persona que eres.

Gracias por estar conmigo, ser mi inspiración y mi ejemplo de constancia y humildad.

## **AGRADECIMIENTO**

Quiero expresar mi gratitud a mis padres Héctor Vinicio Chávez Gómez y Silvia Esperanza Castillo por confiar en mí y darme la oportunidad de forjar mi propio camino. Gracias por ser mis pilares y los mentores de los valores y principios que rigen mi vida,

También, agradezco infinitamente a mis hermanos Jonathan y Sebastián que con sus palabras de apoyo me hacían sentir orgullosa de lo que soy y del ejemplo que les puedo dar. Espero algún día convertirme en su fortaleza para que puedan seguir adelante, los quiero.

A mi tutora de tesis Alexandra Marcatoma y a su esposo por haberme guiado con su experiencia, motivación y conocimiento, gracias por sus consejos, enseñanzas, apoyo y sobre todo por brindarme su amistad. A la Ing. Isabel Escudero por su colaboración y correcciones durante el proceso de elaboración del trabajo y a cada uno de los docentes de la carrera de Estadística Informática.

A mis amigos y a toda mi familia con los que compartí esta etapa universitaria gracias por regalarme recuerdos infinitos de alegría y diversión, los quiero.

Finalmente agradezco al PhD, Hugo Cerda y Carolina Ledesma integrantes del grupo CDETERRA; por haberme permitido realizar el presente estudio.

## ÍNDICE GENERAL

RESUMEN.....	XIV
SUMMARY .....	XV
INTRODUCCIÓN .....	1
CAPÍTULO I	
MARCO REFERENCIAL .....	2
1.1. Antecedentes.....	2
1.2. Planteamiento del problema .....	5
1.3. Justificación .....	5
1.3.1. Justificación Teórica .....	5
1.3.2. Justificación Aplicativa.....	6
1.4. Objetivos .....	6
1.4.1. Objetivos Generales.....	6
1.4.2. Objetivos Específicos .....	7
CAPÍTULO II	
MARCO TEÓRICO .....	8
2.1 Miranda.....	8
2.1.1 Historia .....	9
2.1.2 Situación Geográfica.....	9
2.1.3 Hidrografía.....	10
2.1.4 Clima.....	10
2.1.5 Economía.....	11
2.2 Lepidopteras.....	11
2.2.1 Morfología.....	13
2.2.2 Ciclo de Vida de las lepidopteras.....	15
2.2.3. Características Biológicas .....	17
2.2.4 Clasificación de las Lepidopteras .....	18
2.3 Análisis Multivariante .....	21
2.3.1 Definiciones básicas.....	21
2.3.2 Clasificación del Análisis Multivariante.....	21
2.3.3 Observaciones extremas.....	22
2.3.4 Normalidad multivariante.....	23
2.3.5 Análisis de Componentes Principales .....	24

2.3.6 <i>Análisis Factorial</i> .....	29
2.3.7 <i>Análisis de Correspondencias</i> .....	32
2.4 <i>Índice de Shannon</i> .....	46
<b>CAPÍTULO III</b>	
<b>METODOLOGÍA</b> .....	48
3.1 <i>Tipo y Diseño de Investigación</i> .....	48
3.2 <i>Unidad de Análisis</i> .....	48
3.3 <i>Población y Muestra</i> .....	48
3.4 <i>Operacionalización de variables</i> .....	49
<b>CAPÍTULO IV</b>	
<b>ANÁLISIS E INTERPRETACIÓN DE RESULTADOS</b> .....	51
4.1 <i>Análisis exploratorio de datos</i> .....	51
4.2 <i>Selección de Técnicas de Interdependencia</i> .....	61
4.3 <i>Técnicas de Interdependencia</i> .....	61
4.4 <i>Análisis Implicativo</i> .....	86
4.5 <i>Índice de Shannon</i> .....	88
<b>CONCLUSIONES</b> .....	89
<b>RECOMENDACIONES</b> .....	91
<b>BIBLIOGRAFÍA</b>	



## ÍNDICE DE FIGURAS

<b>Figura 1-2:</b> División Político Territorial Estado Miranda .....	8
<b>Figura 2-2:</b> Mariposa de cola de golondrina de ala de pájaro, <i>Troides (Papilionidae)</i> .....	10
<b>Figura 3-2:</b> Tagma cefálico vista lateral y vista frontal .....	13
<b>Figura 4-2:</b> Ciclo de vida de las <i>lepidopteras</i> .....	16
<b>Figura 5-2:</b> Larva polípoda eruciforme.....	16
<b>Figura 6-2:</b> Pupa normal ventra .....	17

## ÍNDICE DE TABLAS

<b>Tabla 1-2:</b> Perfiles-fila.....	33
<b>Tabla 2-2:</b> Perfiles-columna .....	34
<b>Tabla 1-3:</b> Operacionalización de variables .....	49
<b>Tabla 1-4:</b> Distribución estadística de frecuencias “Mes”.....	51
<b>Tabla 2-4:</b> Distribución estadística de frecuencias “Familia” .....	52
<b>Tabla 3-4:</b> Distribución estadística de frecuencias “Subfamilia”.....	53
<b>Tabla 4-4:</b> Distribución estadística de frecuencias “Tiempo atmosférico” .....	54
<b>Tabla 5-4:</b> Distribución estadística de frecuencias “Hospedero” .....	54
<b>Tabla 6-4:</b> Distribución estadística de frecuencias “Hábitat”.....	55
<b>Tabla 7-4:</b> Medidas de tendencia central y dispersión “Humedad” .....	56
<b>Tabla 8-4:</b> Medidas de tendencia central y dispersión “Temperatura”.....	57
<b>Tabla 9-4:</b> Medidas de tendencia central y dispersión “Presión atmosférica” .....	57
<b>Tabla 10-4:</b> Medidas de tendencia central y dispersión “Velocidad del viento promedio” .....	58
<b>Tabla 11-4:</b> Medidas de tendencia central y dispersión “Ráfaga de Viento”.....	59
<b>Tabla 12-4:</b> Medidas de tendencia central y dispersión “Pluviosidad” .....	60
<b>Tabla 13-4:</b> Observaciones extremas.....	62
<b>Tabla 14-4:</b> Matriz de covarianzas .....	66
<b>Tabla 15-4:</b> Matriz de correlación .....	67
<b>Tabla 16-4:</b> Autovalores .....	67
<b>Tabla 17-4:</b> Auto vectores .....	67
<b>Tabla 18-4:</b> Varianza explicada según el orden.....	68
<b>Tabla 19-4:</b> Componentes principales .....	69
<b>Tabla 20-4:</b> Aportación de las variables por componente .....	70
<b>Tabla 21-4:</b> Varianza explicada por las componentes.....	71
<b>Tabla 22-4:</b> Varianza acumulada .....	71
<b>Tabla 23-4:</b> Tabla de correspondencias “Mudable Familia y Mes”.....	72
<b>Tabla 24-4:</b> Perfiles de fila mudables “Familia y Mes” .....	72
<b>Tabla 25-4:</b> Perfiles de columna mudables “Familia y Mes” .....	73
<b>Tabla 26-4:</b> Resumen mudables “Familia y Mes”.....	73
<b>Tabla 27-4:</b> Puntos de fila generales mudables “Familia y Mes”.....	74
<b>Tabla 28-4:</b> Puntos de columnas generales.....	74
<b>Tabla 29-4:</b> Matriz de correspondencias mudables “Subfamilia y Hospedero”.....	76

<b>Tabla 30-4:</b> Matriz de correspondencias mudables “Familia y Hospedero” .....	77
<b>Tabla 31-4:</b> Perfiles fila “Familia y Hospedero”.....	77
<b>Tabla 32-4:</b> Perfiles de columna “Familia y Hospedero” .....	78
<b>Tabla 33-4:</b> Resumen “Familia y Hospedero”.....	78
<b>Tabla 34-4:</b> Varianza explicada según la dimensión.....	80
<b>Tabla 35-4:</b> Contribuciones de la dimensión 1 .....	82
<b>Tabla 36-4:</b> Contribuciones de la dimensión 2.....	83
<b>Tabla 37-4:</b> Cálculo del Índice de Shannon.....	87

## ÍNDICE DE GRÁFICOS

<b>Gráfico 1-4:</b> Porcentaje de lepidopteras por “Mes” .....	51
<b>Gráfico 2-4:</b> Porcentaje de lepidopteras por “Familia” .....	52
<b>Gráfico 3-4:</b> Porcentaje de lepidopteras por “Subfamilia” .....	53
<b>Gráfico 4-4:</b> Porcentaje de lepidopteras por “Tiempo atmosférico” .....	54
<b>Gráfico 5-4:</b> Porcentaje de lepidopteras por “Hospedero” .....	55
<b>Gráfico 6-4:</b> Porcentaje de lepidopteras por “Hábitat” .....	55
<b>Gráfico 7-4:</b> Variable “Humedad” .....	56
<b>Gráfico 8-4:</b> Variable “Temperatura” .....	57
<b>Gráfico 9-4:</b> Variable “Presión atmosférica” .....	58
<b>Gráfico 10-4:</b> Variable Velocidad del viento promedio” .....	59
<b>Gráfico 11-4:</b> Variable “Ráfaga de viento” .....	60
<b>Gráfico 12-4:</b> Variable Pluviosidad” .....	61
<b>Gráfico 13-4:</b> Variables climáticas .....	65
<b>Gráfico 14-4:</b> Relación entre variables climáticas .....	66
<b>Gráfico 15-4:</b> Diagrama de sedimentación de la varianza explicada .....	68
<b>Gráfico 16-4:</b> Diagrama bidimensional de las componentes principales .....	70
<b>Gráfico 17-4:</b> Mudables familia y mes.....	75
<b>Gráfico 18-4:</b> Mudables Familia y Hospedero.....	79
<b>Gráfico 19-4:</b> Contribución de las variables .....	80
<b>Gráfico 20-4:</b> Gráfico de variables.....	81
<b>Gráfico 21-4:</b> Variables de acuerdo a la tonalidad .....	81
<b>Gráfico 22-4:</b> Variables resaltadas.....	82
<b>Gráfico 23-4:</b> Círculo de correlaciones .....	83
<b>Gráfico 24-4:</b> Variables en todas las dimensiones .....	84
<b>Gráfico 25-4:</b> Calidad en el mapa de factores .....	84
<b>Gráfico 26-4:</b> Variables contribuyentes.....	85
<b>Gráfico 27-4:</b> Dendograma de similaridad.....	86
<b>Gráfico 28-4:</b> Grafo implicativo.....	86
<b>Gráfico 29-4:</b> Árbol cohesivo.....	87

## ÍNDICE DE ABREVIATURAS

1. **ACP** Análisis de Componentes Principales
2. **AC** Análisis de Correspondencias

## RESUMEN

El objetivo fue discriminar a través de técnicas de interdependencia el cambio en la estructura de las comunidades de lepidópteros en función de variables climáticas, hábitat natural y antrópico en Altos Mirandinos Venezuela, 2013. Las *lepidopteras* son guías fiables del estado y diversidad del medio ambiente. Sin embargo, los constantes cambios climáticos y el comportamiento antrópico se han convertido en factores influyentes de la presencia de las comunidades de estos seres vivos. La información proporcionada se obtuvo a través de una búsqueda directa con red entomológica manual dando como resultado el hallazgo de 1680 *lepidopteras*. El taxón con mayor abundancia de *lepidopteras* fue *Nymphalidae* (50,5%) además el mes con mayor presencia de *lepidopteras* fue marzo (19,8%). Las técnicas de interdependencia usadas fueron análisis de componentes principales, análisis de correspondencia y análisis factorial mixto. El Análisis de Componentes Principales (ACP) mostró que las dos primeras componentes explican el 92% de la variabilidad de la información demostrando una relación entre el clima y la presencia de las *lepidopteras*, por otro lado el Análisis de Correspondencias (AC) concluyó que la única variable que discrimina la abundancia de *lepidopteras* es la mudable mes, finalmente el Análisis Factorial Mixto evidenció un factor 1 denominado elementos meteorológicos con las variables temperatura, humedad y pluviosidad, y un factor 2 denominado fenómenos atmosféricos agrupó a las variables presión atmosférica, ráfaga de viento y velocidad promedio de viento; sin embargo las mudables de hábitat no aportaron mayor información acerca de la concurrencia de estos insectos no obstante las mudables tiempo atmosférico, familia (*Papilionidae*) y el hospedero (al vuelo) reflejaron la variabilidad de su presencia. Adicionalmente se resalta la importancia de realizar más estudios con apoyo estadístico en el área ecológica ya que dichos estudios permitirán implementar estrategias y políticas que incluyan la conservación de los diferentes insectos.

**Palabras Claves:** <ESTADÍSTICA>, <MARIPOSAS (*Lepidopteras*)>, <TÉCNICAS DE INTERDEPENDENCIA>, <COMPONENTES PRINCIPALES>, <FACTORIAL MIXTO>, <CORRESPONDENCIAS>.

## SUMMARY

The objective was to discriminate, through interdependent techniques, the change in the structure of the lepidoptera communities based on climatic variables, natural and anthropic habitat in Altos Mirandinos Venezuela, 2013. The *Lepidoptera* are reliable guides to the state and diversity of the environment. However, the constant climatic changes and the anthropic behavior have become influential factors of the presence of the communities of these living beings. The information provided was obtained through a direct search with a manual entomological network, resulting in the discovery of 1680 *Lepidoptera*. The taxon with the highest abundance of *Lepidoptera* was *Nymphalidae* (50.5%), and the month with the highest presence of *Lepidoptera* was March (19.8%). The interdependence techniques used were principal components analysis, correspondence analysis and mixed factorial analysis. The Principal Component Analysis (PCA) showed that the first two components explain 92% of the variability of the information demonstrating a relationship between the climate and the presence of the *Lepidoptera*, on the other hand the Correspondence Analysis (CA) concluded that the only variable that discriminates the abundance of *Lepidoptera* is the changeable month, finally the Mixed Factorial Analysis showed a factor 1 called meteorological elements with the variables temperature, humidity and rainfall, and a factor 2 called atmospheric phenomena grouped into the variables atmospheric pressure, burst of wind and average wind speed; However, the habitat did not provide more information about the concurrence of these insects, despite the changeable weather, family (*Papilionidae*) and host (in flight) reflected the variability of their presence. Additionally, the importance of carrying out more studies with statistical support in the ecological area is highlighted as these studies will allow the implementation of strategies and policies that include the conservation of different insects.

**Key Words:** <STATISTICS>, <BUTTERFLIES (*Lepidoptera*)>, <INTERDEPENDENCE TECHNIQUES>, <MAIN COMPONENTS>, <MIXED FACTORIAL, <CORRESPONDENCES>.

## INTRODUCCIÓN

Dentro de la biodiversidad ecológica se encuentran las *lepidopteras*, comúnmente conocidas como mariposas. Las *lepidopteras* son utilizadas como indicadores biológicos de la calidad medioambiental y de la salud de los ecosistemas debido a que son vulnerables a los cambios y se pueden encontrar en diversos hábitats, es decir, funcionan como organismos modelos para obtener pistas esenciales acerca del entorno, como: el cambio climático, la contaminación del aire y los efectos de las nuevas prácticas agrícolas (MUNN, 2013, p. 73), (Butterfly Conservation Europe, 2008, p. 2).

Una problemática común en los estudios ecológicos radica en la gran cantidad de variables que es necesario considerar, lo que hace prioritario el uso de técnicas que simplifiquen o sintetizen la información, permitiendo, al mismo tiempo una explicación de los fenómenos o estructuras estudiadas es así que el análisis multivariante se revela como una técnica de ordenación óptima para el tratamiento de datos ecológicos (Calabuig, 1980, p. 17).

La presente investigación utilizó una metodología exploratoria retrospectiva no experimental, contó con una fuente de datos secundarios proporcionada por el grupo de investigación CEDETERRA “Centro de Desarrollo de Tecnología para la reducción y racionalización de agroquímicos” captada de los Altos Mirandinos la zona poblada más alta del estado Miranda, Venezuela, los paquetes estadísticos de soporte fueron SPSS y R.

Las técnicas de interdependencia usadas fueron análisis de componentes principales, análisis de correspondencia y análisis factorial mixto a través de la presente investigación se busca contribuir a las aportaciones científicas halladas, con la determinación de la estructura que presentan las comunidades de *lepidopteras* en función de factores climáticos y antrópicos. Es por ello que analizar cómo interactúan las diferentes comunidades dará respuesta a las diferentes incógnitas que los ecólogos se plantean (Iván Andrés Díaz Concha, 2011, p. 1-23).



# CAPÍTULO I

## MARCO REFERENCIAL

### 1.1. Antecedentes

Estudiar la biodiversidad es uno de los objetivos fundamentales en el área biológica (Carvalho, 2009, pp.15-21); ésta resulta de un conjunto de fenómenos evolutivos y ecológicos, y se desenvuelve alrededor de toda la escala de organización de los seres vivos. La diversidad ecológica representa el número de especies y abundancia, tanto en el tiempo como en el lugar, ya que analiza las interrelaciones de las comunidades al interior de las poblaciones y en el ecosistema (Camero y Calderón, 2007, pp.95-110); valorando así el conocimiento de estos componentes. Las comunidades biológicas conceptualmente representan un conjunto de especies poblacionales con características únicas que coexisten e interactúan y funcionan en virtud de una compleja red de dependencia, estas comunidades están dictaminadas por tres componentes: composición, distribución espaciotemporal o estructura y función; el estudio de una comunidad en función de estos componentes no resulta redundante más bien enriquece los saberes ecológicos. Por otro lado, según Kresbs (1985) “*la estructura, la organización física o los patrones de un sistema ecológico se deben estudiar teniendo en cuenta que las comunidades biológicas poseen un conjunto de atributos que no residen en cada una de las poblaciones que las componen, sino que se manifiestan en la comunidad*” (Camero y Calderón, 2007, pp.95-110), (Botánico et al., 2007, pp.185-189).

El aumento demográfico en los últimos años ha generado una afectación en el reino animal y vegetal, debido a la tala de bosques, incendios forestales, contaminantes en los ríos, erosión, cosecha indiscriminada, introducción de especies, etc., estos cambios han ido de la mano con los desastres naturales y las variaciones climáticas, no solo locales sino también globales, perturbando las interacciones de los insectos de varias maneras provocando cambios en su fisiología, diversidad y abundancia (Ospina et al., 2015, pp.455-474).

Las *lepidopteras* son un grupo de insectos que desempeñan funciones vitales para el desarrollo del medio ambiente debido a que son los agentes bióticos transportadores del polen en el planeta, además cumplen funciones de reciclaje a través del consumo de materia orgánica dentro de los ecosistemas. Por otro lado son indicadores biológicos de la abundancia y la salubridad; es por ello que este grupo de animales han sido objeto de estudio desde hace varios años (Ortega,

2015, pp. 320-347), (Botánico et al., 2007, pp.185-189).

Hasta el momento, estudios como el propuesto en el Altiplano Cundiboyacense por Díaz y Bustamante (2011, pp.1-23), titulado “**Estructura y composición de la comunidad de insectos en una especie nativa (*Lupinus bogotensis*) y una invasora (*Ulex europaeus*) a la luz de Variables ecológicas: hora y temperatura en ambientes de borde de bosque y de camino**” hallaron que las especies exóticas invasoras representan una alteración a nivel estructural y funcional dentro de un ecosistema, y pueden tener un impacto negativo importante en el ecosistema donde han sido introducidas. El estudio de su efecto en el Altiplano Cundiboyacense es reciente, y en el tema de las relaciones planta-insecto los estudios son reducidos. El grupo taxonómico de los insectos representa un indicador importante por su abundancia en el campo, su taxonomía relativamente bien conocida y su importancia en cuanto a relaciones ecológicas dentro de un ecosistema; además se identificaron las diferencias en la actividad de los insectos en relación a las variables de hora y temperatura.

El estudio de Núñez y Barro (2003, pp.1-11) denominado “**Composición y Estructura de dos comunidades de mariposas (*Lepidoptera: papilionoidea*) en Boca de Canasi, La Habana, Cuba**” mostró aspectos de la composición y estructura de dos comunidades de mariposas presentes, en dos formaciones vegetales diferentes en un área natural relativamente bien conservada y un bosque siempre verde micrófilo (BSM). En el área fueron observadas 93 especies de mariposas (entre los dos hábitat), las cuales pertenecen a seis de las siete familias presentes en Cuba. La comunidad de mariposas del uveral fue la más rica en especies, la más diversa y la más equitativa con respecto a la comunidad del BSM.

Un estudio titulado “**Análisis de las comunidades de mariposas en los agrosistemas en Aragón, España (*Lepidoptera: Papilionoidea & Hesperioidea*)**” realizado por Castel (2005, pp.247-263) analizò la distribución faunística de las comunidades de mariposas en el agrosistema de la comarca Somontano de Barbastro (Sierras Marginales del Prepirineo Aragonés) muestreando diversos paisajes agrarios con distinta intensidad de gestión, desde áreas multiparceladas y de pluricultivo a otras de monocultivo, desde cultivos herbáceos a leñosos, en secano y regadío, analizando en cada uno de ellos la estructura faunística (número de especies y abundancia poblacional) y la diversidad, concluyendo que las comunidades de mariposas están relacionadas con la tipología vegetal y con la intensidad de la gestión antrópica aplicada en los distintos paisajes agrarios.

La investigación de Irazábal (2014, pp.1-74) titulada **“Patrones de estratificación de la comunidad de mariposas (*Ropalocera: Nymphalidae*) de un bosque húmedo tropical del Chocó ecuatoriano”** determinó que la composición y estructura de las comunidades de mariposas de cada estrato (dosel y sotobosque), varían significativamente a través de las estaciones húmeda y seca.

Según Checa (2014, pp. 1-13) en su estudio **“Microclimate variability significantly affects the composition, abundance and phenology of butterfly communities in a highly threatened neotropical dry forest”** pocos son las investigaciones que apuntalan la descripción de la dinámica poblacional de las comunidades de mariposas a partir de variables de microclima. Luego de la recolección de 93 especies de mariposas en una zona seca de Ecuador Checa encontró que la variación en el microclima y densidad de vegetación fueron factores predictores de la composición y abundancia de las mariposas.

En un estudio sobre la evolución de las técnicas estadísticas aplicadas en ecología desde 1992 hasta el 2010 concentrado en el artículo **“El pasado, presente y futuro en el uso de la estadística en estudios ecológicos en Bolivia”** escrito por Naoki (2012, pp.79-87) muestra que *“Muchos estudios trataron de medir más de una variable para caracterizar sus organismos o sistemas ecológicos y de evaluar simultáneamente el efecto de más de una variable o proceso. A pesar de su importancia y utilidad, los estudios descriptivos sin la aplicación de inferencias estadísticas adecuadas no aportan significativamente a los dos roles principales de los ecólogos: Identificar patrones inherentes a los sistemas naturales y elucidar aquellos procesos que los generan.”* y como en la gran mayoría de estudios ecológicos utilizan más de una variable surge la necesidad de generar estudios con técnicas multivariantes de interdependencia, para determinar las características que presentan las lepidópteras según el hábitat y un conjunto de variables climáticas que marcan la vida de este grupo de seres vivos, siendo esta una de las respuestas fundamentales para estudios ecológicos no analizados por los patrones estadísticos multivariados.

En otro estudio de Ríos (2013) titulado **“Estudio preliminar de la riqueza y composición de especies de lepidópteros diurnos, en un paisaje agrícola de los Altos Mirandinos (Cordillera de la Costa, Venezuela)”** demuestra la importancia de los paisajes agrícolas como hábitat para las *lepidopteras* diurnas además estudia la diversidad de estas especies utilizando los estimadores no paramétricos Ice y Chao.

La pérdida y desgaste de hábitat a causa de cambios en la combinación de factores ambientales y genéticos es la mayor amenaza para las *lepidopteras* e incluso para otras especies

taxonómicas, las poblaciones tienden a reducirse rápidamente debido a condiciones poco favorables impidiendo la adaptación de estas especies, por ello en busca de brindar registros biológicos dentro de los Altos Mirandinos, el presente proyecto servirá como información base para estudios futuros que permitan la toma de decisiones en la región y de manera especial en el área de estudio procurando la conservación de estos insectos (Moreno et al., 2015, pp.221-234).

## **1.2. Planteamiento del problema**

Entre los insectos, las *lepidopteras* se han establecido como taxón modelo de estudios para la caracterización de la biodiversidad en estimaciones de impacto ambiental, monitoreo de poblaciones animales y muchos otros estudios. Además ante la escasez de estudios sobre estructuras de comunidades de insectos que influyen en el ecosistema se busca dar respuesta a la siguiente pregunta (Ospina et al., 2015, pp.455-474).

¿Qué tipo de estructura presentan las comunidades de *lepidopteras* en función de variables climáticas, hábitat natural y antrópico de los Altos Mirandinos Venezuela, 2013?

## **1.3. Justificación**

### ***1.3.1. Justificación Teórica***

Actualmente el análisis multivariado es un recurso de acceso universal a la minería de datos ya que a través de este se consigue clasificar según pesos las características homogéneas de las unidades experimentales extraídas a partir de un conjunto de variables.

El análisis multivariante aporta de manera sustancial al área ecológica debido a que se manejan datos de comunidades de especies en diversos sitios y con el efecto de diversos factores ambientales, la aplicación de esta técnica estadística a la ecología es la oportunidad perfecta para resolver problemas ecológicos de manera natural, rutinaria y fructífera, dentro de esta se encuentran los análisis de interdependencia (análisis de componentes principales, análisis factorial, análisis de correspondencias) los mismos que ayudan a la clasificación de las comunidades de lepidópteras en función de variables climáticas y de hábitat (Gauch jr., 1982, p.298).

### **1.3.2. Justificación Aplicativa**

En el campo de la ecología la estadística se ha vuelto una de las herramientas principales de apoyo en cada uno de los estudios científicos pues una de las problemáticas de esta área es la gran cantidad de variables que se necesita analizar. Los ecólogos tienen dos tareas principales, identificar los patrones inherentes a los sistemas naturales <sup>1</sup> y dilucidar aquellos procesos que los generan, es por ello que el propósito de la presente investigación es conocer el tipo de estructura de las comunidades de *lepidopteras* según variables climáticas y de hábitat para detectar los climas influyentes y el tipo de hábitat donde se congrega en gran mayoría esta familia de insectos (Naoki, 2012, pp.79-87).

Las *lepidopteras* tienen una trascendencia significativa en el ámbito ecológico ya a que aparte de ser el grupo de herbívoros con la mayor porción de especies y biomasa<sup>2</sup> presentes en el planeta, de cumplir funciones de transferencia de polen y de servir de alimento para algunos depredadores como las aves, serpientes, sapos, etc. También son indicadores excelentes de sucesos antrópicas del medio ambiente (Ortega, 2015, pp.320-347).

La diversidad de *lepidopteras* es una fuente directa de aproximación a la diversidad vegetal dado que dependen directamente de las plantas. Entre los insectos, las *lepidopteras* prestan un servicio ecosistémico como controladores biológicos. Es así que las relaciones entre las variables climáticas y de hábitat brindan la posibilidad de tomar acciones inmediatas que permitan prevenir las afectaciones en las comunidades de *lepidopteras* causadas por el cambio climático evitando así desencadenar una posible desaparición de este grupo de insectos bioindicadores (Ortega, 2015, pp.320.347).

## **1.4. Objetivos**

### **1.4.1. Objetivos Generales**

Discriminar a través de técnicas de interdependencia el cambio en la estructura de las comunidades de lepidópteras en función de variables climáticas, hábitat natural y antrópico en Altos Mirandinos Venezuela, 2013.

---

<sup>1</sup> Organización relacionada de elementos que surge como una propiedad de la naturaleza.

<sup>2</sup> Cantidad de productos obtenidos por fotosíntesis, susceptibles a ser transformados en combustible útil para el hombre y expresada en unidades de superficie y volumen,

#### **1.4.2. *Objetivos Específicos***

- Describir las variables de las comunidades de *lepidopteras* que reposan en la base de datos proporcionada por CDETERRA.
- Usar técnicas estadísticas para la limpieza de datos.
- Seleccionar las técnicas de interdependencia adecuadas según la matriz de datos.
- Detectar las variables climáticas influyentes en el cambio de la estructura en las comunidades de *lepidopteras* a través de técnicas de interdependencia.
- Comparar los resultados entre las técnicas de interdependencia con el análisis estadístico implicativo (Gráfo de similaridad y cohesión)
- Generar el índice de diversidad para las comunidades de lepidópteras.

## CAPÍTULO II

### MARCO TEÓRICO

#### 2.1 Miranda

Con una superficie de 7.950 Km aproximadamente el Estado Miranda representa el 0,87% del territorio venezolano, limita al Norte con el Distrito Capital, Vargas y Mar Caribe, al Este con el Mar Caribe y Anzoátegui, al Oeste con el estado Aragua y al Sur con los estados Guárico y Aragua. Tiene como principales ciudades a Baruta, Caucagua, Cúa, Charallave, Guarenas, Guatire, Higuero, Río Chico Ocumare del Tuy, Petare, Santa Teresa, Santa Lucía y San Francisco de Yare y su capital Los Teques (Figura 1-2) (Instituto Nacional de Estadística, 2014, pp.1-97)(D'Elia, 2010, pp.1-41).



**Figura 1-2:** División Político Territorial Estado Miranda.

Fuente: (Instituto Nacional de Estadística, 2014, pp.1-97)

El Estado Miranda se encuentra dividido en cinco regiones:

- **Barlovento:** abarca los municipios de Acevedo, Andrés Bello, Brión, Buroz, Páez y Pedro Gual.
- **Laderas de la Cordillera del Litoral:** representa a los municipios de Plaza y Zamora.
- **Valles del Tuy:** compuesta por los municipios de Cristóbal Rojas, Independencia,

Lander, Paz Castillo, Simón Bolívar y Urdaneta.

- **Altos Mirandinos:** formado por los municipios de Carrizal, Guaicaipuro, y Los Salías.
- **Valle de Caracas:** conformado por los municipios de Baruta, Chacao, El Hatillo y Sucre.

### ***2.1.1 Historia***

Miranda fue uno de los territorios de la Provincia de Caracas que con su constante crecimiento dio lugar al desarrollo de diversos poblados entre ellos los valles de Caucagus y Carora, de este a oeste, y desde el mar Caribe hasta los llanos de Calabozo de norte a sur.

Gracias al triunfo federal en el año de 1863 la provincia se transformó en el Estado de Caracas evento que impulsó la creación del Distrito Federal en el año de 1864, este distrito se formó con los cantones de Caracas, La Guaira y Maiquetía mientras que los demás territorios se convirtieron en el Estado de Bolívar con su capital Petare. Tiempo después se vivió la separación del Distrito Federal y en el año de 1868 Caracas fue asignado como capital del Estado Bolívar y el territorio del Distrito fue absorbido por el mismo.

Apure, Guarico y Guzmán Blanco formaban parte del Estado del Centro y en el año de 1879 el Estado de Bolívar se convirtió en un territorio más de la zona central. Dos años después se ejecutó una nueva repartición territorial la cual se caracterizó por integrar el territorio de Miranda a Bolívar del Gran Estado Guzmán Blanco con su capital Villa de Cura.

Es así como a partir del año 1879 el Estado Miranda atravesó un constate cambio de nombre y capital y no fue hasta 1909 que Miranda quedó estructurado definitivamente. Actualmente el área del estado de Miranda conforma parte de la zona centro del país de Venezuela y constituye uno de los lugares preferidos por los habitantes de Margarita, en el oriente, y sobre todo de El Tocuyo, en el occidente para su poblamiento (Instituto Nacional de Estadística, 2014, pp.1-97).

### ***2.1.2 Situación Geográfica***

El estado Miranda cuenta con hermosos paisajes en su relieve entre los cuales destacan la Cordillera de la Costa paralela al Mar Caribe separada con una franja relativamente estrecha, cuenta con alturas superiores a los 2.500 m, entre las cuales destaca el pico Naiguatá (2765 m) con pendientes superiores al 69% su relieve es infranqueable; con similares características pero con menor elevación y un poco más al sur se encuentra la Sierra del Interior. Los valles Guarenas-Guatire y del Tuy que rodean a estas dos grandes cadenas montañosas han permitido



que la población logre establecer sus viviendas y obtener un papel económicamente activo, además, culminando la costa noroeste del estado se ubican zonas de tierras relativamente planas como la depresión Barloventeña que drena los principales ríos del sistema; con una gran diversidad de bosques (deciduos, semisecos) y especies (araguaney, indio desnudo, apamate, caoba, bucare, samán y otros) selváticas como la selva de Guatopo por debajo de los 1.000 metros de altura y cerca de ellas se encuentran las zonas de cacaotales de Barlovento con sus especies vegetales cactáceas columniformes, cactáceas bajas mezcladas con arbustos casi siempre espinosos, el yabo, divide, cují, yaqué, cariaquito negro, tuna y cardón. En las cercanías de Caracas, el Tuy Medio y en el valle Guarenas- Guatire dominan los matorrales (Instituto Nacional de Estadística, 2014, pp.1-97), (D'Elía, 2010, pp.1-41).

### ***2.1.3 Hidrografía***

Miranda posee un recorrido corto y escaso caudal en sus ríos, esta característica combinada con el asentamiento de la masa de población más grande del país, confiere a sus ríos una gran significación como fuentes de agua para consumo urbano; mientras que el río Tuy, recorre alrededor de 250 km, nace a 2.100 m de altitud, en el flanco sur de la cordillera del Litoral, fluye hacia el este por el abra de Tacatá hasta los valles del Tuy y Barlovento, y desemboca en el mar Caribe por la boca de Páparo.

El río El Jarillo y las quebradas Aguas Frías y La Negra, afluentes del Tuy por la margen izquierda, han sido represados, cerca de los Teques, en el embalse de Agua Fría, mientras que el Ocumare, fue represado cerca de Ocumare del Tuy, en el embalse de Ocumarito. El afluente más conocido es el río Guaire. El Tuy recibe los ríos Taguaza, Taguacita y Cuirá por el borde derecho. Entre los ríos más relevantes se tiene el Guarenas, río Grande o Caucagua, Capaya, El Guapo, El Curiepe, Cúpira (Instituto Nacional de Estadística, 2014, pp.1-97), (D'Elía, 2010, pp.1-41).

### ***2.1.4 Clima***

La altura, los vientos y la distancia del mar son factores que influyen directamente en la climatología de este territorio. En las zonas altas, como Los Teques, San Diego y Altos de Pipe, se registran temperaturas que oscilan entre los 18° y los 20°C, y con mucha frecuencia, es posible encontrar neblina. La cordillera de la Costa recibe vientos precipitados cargados de humedad que desencadenan lluvias abundantes, estos vientos provienen del mar Caribe y recorren el interior de la depresión de Barlovento la zona del Tuy medio, Guarenas, Guatire y el valle de Caracas registran temperaturas entre los 25° y 26°C, con períodos de lluvias entre los meses de mayo a noviembre. La zona de Barlovento registra temperaturas anuales de 28°C y su

período de lluvia abarca la mayor parte del año, dejando un promedio que oscila entre los 2.000 y 2.500 mm (Instituto Nacional de Estadística, 2014, pp.1-97), (D'Elía, 2010, pp.1-41).

### **2.1.5 Economía**

La riqueza de Miranda se destaca en sus suelos lo que ha permitido que la agricultura sea su principal potencial económico. En los valles del Tuy los cultivos que predominan en su producción son cacao, caña de azúcar, maíz, entre otros, mientras que en el territorio de Barlovento frutas y verduras son las preferidas. Otras de las comunidades que se caracterizan por su gran fertilidad de suelos son Guatire y Guarenas siendo el café el sustento primordial de la zona.

El Distrito Capital cuenta con un conjunto urbano manufacturero siendo así la manufactura otra de las fuentes económicas de Miranda. Además la arquitectura hotelera del litoral barloventeño ha impulsado la actividad turística sin quitarle importancia a las actividades financieras, comerciales y de servicios que se concentran en el área de Miranda (Instituto Nacional de Estadística, 2014, pp.1-97).

## **2.2 Lepidopteras**

El término *lepidoptera* proviene de la raíz latina *lepido* o de las antiguas raíces griegas *lepis* que significa escama y *pteron* que significa ala, es decir, las *lepidopteras* representan insectos con alas escamosas. Según Linnaeus las *lepidopteras* se dividen en tres grupos:

- Mariposas
- Patrones
- Micro y macro polillas.

Gracias a los esplendorosos colores de las mariposas producidos por la mezcla de pigmentos y la difracción de los rayos solares sobre las escamas esta especie se ha convertido en el símbolo de la belleza y elegancia del planeta (Figura 2-2) (Urretabizkaya et al., 2010a, pp.1-77).



**Figura 2-2:** Mariposa cola de golondrina de ala de pájaro, *Troides* (*Papilionidae*).

Fuente:(Grimaldi y Engel, 2005, pp.1-770)

Son el segundo orden de insectos más diverso y grande del ecosistema. Existen alrededor de 500.250 lepidópteros con 70.820 especies de mariposas, 3700 especies de patrones y aproximadamente 165,000 especies de polillas, incluyendo micro y macro polillas; este grupo de insectos se compone de 126 familias y 46 subfamilias las cuales se distinguen en características morfológicas, anatómicas, ecológicas y de comportamiento (Figura 3-2) (Khan, 2017, pp.1-17).

### **Posición taxonómica del orden *lepidoptera***

**Class:** Insects

**Infraclass:** Neoptera

**Subclass:** Pterygota

**Superorder:** Endopterygota

Unranked: Amphiesmenoptera

Unranked: Holometabola

**Order:** Lepidoptera Linnaeus, 1758

#### **Examples:**

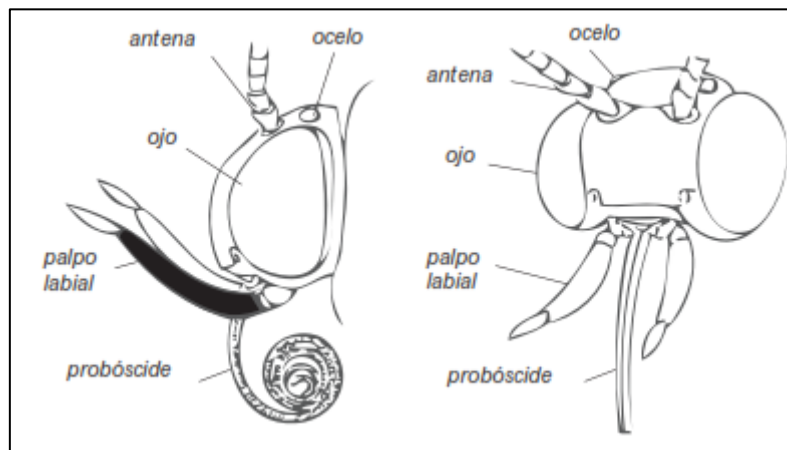
- Butterflies
- Skippers
- Micro-moths
- Macromoths (Khan, 2017, pp. 1-17).

### 2.2.1 Morfología

Las *lepidopteras* son insectos que poseen un sinnúmero de formas, tamaños, estructuras y diversos rasgos que hacen que cada *lepidoptera* sea única.

#### 2.2.1.1 Cabeza

Es pequeña, tiene forma redonda o elíptica y es el centro sensorial de la lepidóptera. La parte superior media de la cabeza recibe el nombre de ocelo, debajo se encuentran la antena, los ojos y más abajo la probóscide el cual se expande a los bordes de la mandíbula. En conjunto características como la forma, el tamaño de la cabeza, el color y la ubicación de los pelos en la cabeza brindan sustento a la identificación de especies de orugas con ayuda de un microscopio (Figura 3-2), (Khan, 2017, pp.1-17), (Urretabizkaya et. al., 2010a, pp.1-77).



**Figura 3-2:** Tagma cefálico vista lateral y vista frontal.  
Fuente:(Urretabizkaya et al., 2010a, pp.1-77)

#### 2.2.1.2 Aparato Bucal

Con piezas bucales de tipo sifón el aparato bucal presenta una estructura hueca larga y flexible con una formación de la trompa suctorial que abarca un tubo de alimento hermético. Además gracias a la ayuda de pequeños músculos presentes allí su longitud de probóscide puede aumentar casi 100 pliegues y mientras no necesiten alimentarse del néctar estas permanecen guardadas bajo su cabeza. Todas las *lepidopteras* conservan la estructura básica del aparato bucal es decir labio e hipofaringe o lengua con pares de mandíbulas y maxilares (Khan, 2017, p.3).

### 2.2.1.3 Antenas

Las antenas se ubican entre los ojos compuestos y muestran una gran variabilidad en cuanto a formas, tamaños, especies e incluso difieren por el sexo. La estructura básica de las antenas de *lepidoptera* suele ser filiforme, se altera en los capitales de la antena los cuales tienen forma de maza con un eje largo y una bombilla en el extremo.

En los patrones, la mayoría de las puntas de las antenas se cambian en una proyección angosta en forma de gancho mientras que las antenas de una polilla son plumosas o con filo de sierra. Las formas de las antenas ayudan a identificar las especies de lepidópteros (Urretabizkaya, Vasicek et al., 2010a, p. 11), (Khan, 2017, pp.3-4).

### 2.2.1.4 Ojos compuestos

Los ojos suelen estar emparejados, pueden ser de color dorado o incluso rojo como en algunas especies de patrones. Comparten características similares a los ojos de un artrópodo debido a que se componen de unidades repetitivas de hasta 17,000 omatidios, cada uno funciona como un receptor visual por separado y en conjunto proporcionan una vista mosaico de cualquier imagen, Cada ojo está conectado a un lente y este a su vez se encuentra unido a un nervio que va hacia el cerebro (Khan, 2017, p.4).

### 2.2.1.5 Tórax

El tórax es la segunda parte del cuerpo de la *lepidoptera*, se compone de tres segmentos articulados, el protórax, mesotórax y metatórax. Están cubiertos dorsalmente con tergitos, ventralmente con estepias y lateralmente con pleuritas (placas quitinosas). Entre los rasgos que ayuda a identificar a las especies de *lepidopteras* se tiene la presencia o ausencia y forma de las placas distorsionadas, la ubicación de las setas primarias, y la ubicación, color y forma del espiráculo protorácico (Urretabizkaya et al., 2010a, p. 12), (Khan, 2017, p. 4-5).

### 2.2.1.6 Patas

Los lepidópteros tienen tres pares de patas bien desarrolladas. Se localizan en cada segmento del tórax y se cubren con escamas. Cada pierna consta de nueve segmentos coxa, tibia, cinco segmentos tarsales con un pretarso y un par de garras curvas articuladas. La tibia de cada pierna contiene un órgano subgenual que detecta y amplifica pequeñas vibraciones (Urretabizkaya et al., 2010a, p. 12).

### 2.2.1.7 Alas

Las *lepidopteras* cuentan con un par de alas las cuales desempeñan actividades como buceo en paracaídas, equilibrio en paracaídas, etc. haciendo que vuelen y demuestren su ligereza en la naturaleza. Los fuertes muslos en el tórax mueven las alas hacia arriba y hacia abajo durante el vuelo. Ambos pares de alas están cubiertos con miles de escamas brillantes, coloridas y opacas. Las alas son membranas quitinosas, nutridas y soportadas por venas tubulares; la venación a más de ser un rasgo importante en el desarrollo filogenético de las *lepidopteras* cumple una función aerodinámica, pues permite que las *lepidopteras* modifiquen su sistema de vuelo y se adapten a diferentes entornos.

Muchas veces las alas se ven perturbadas por las condiciones variadas del medio ambiente causando una afectación directa en el tiempo, la velocidad, la alimentación, la búsqueda de lugares para mantenerse a salvo de los depredadores, etc.

Los patrones de vuelo son un ejemplo de éxito en la adaptación que desarrollan las *lepidopteras*. Cada cambio en la morfología de las alas está estrechamente relacionado con el material genético es por ello que los estudios de *lepidopteras* desempeñan un rol de gran importancia al momento de comprender el estado evolutivo y la taxonomía de este grupo de insectos (Khan, 2017, pp. 5-6)(Urretabizkaya et al., 2010a, p. 12)(Báez y Oromi, 2010, pp.302-318).

### 2.2.1.8 Abdomen

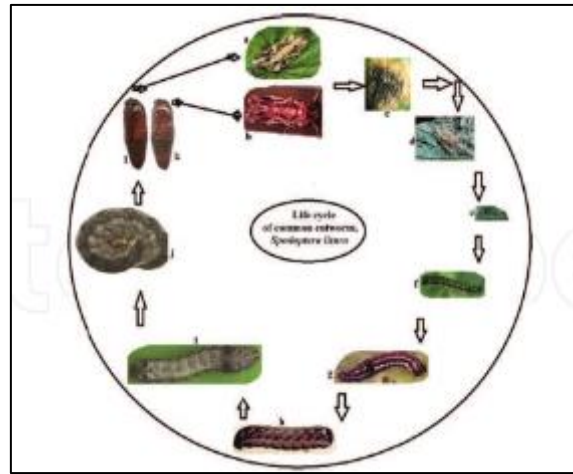
El abdomen es la tercera parte del cuerpo de las *lepidopteras* de aspecto cilíndrico contiene de 10 a 11 segmentos que se estrechan hacia el segmento final y proporciona membranas que permiten la articulación y el movimiento. En muchos lepidópteros existe un órgano complejo el tímpano ubicado en ambos lados del metatórax o en la base del abdomen el cual se encarga de percibir sonidos de alta frecuencia (Khan, 2017, p.6), (Urretabizkaya et al., 2010a, p. 12).

## 2.2.2 Ciclo de Vida de las lepidopteras

Su ciclo es Huevo-Larva-Pupa-Adulto (Figura 4-2) (Kauffmann, 2002, pp.1-3).

**Huevo:** Tienen un tamaño máximo de 1 mm vienen en diferentes formas como esféricas, hemisféricas, lenticulares, ovoides o fusiformes. Presenta un color amarillo verdoso y a medida que van creciendo este color se torna oscuro. Los huevos son dejados sobre hojas, tallos, flores y frutos, se sostienen debido a una sustancia colerética que la mariposa deja en el momento de la

de la oviposición y se solidifica al entrar en contacto con el aire (Urretabizkaya et al., 2010a, pp.12-13), (Báez y Oromi, 2010, pp. 302-318), (Kauffmann, 2002, pp. 1-3).



**Figura 4-2:** Ciclo de vida de las *lepidópteras*.  
Fuente: (Khan, 2017, p.8)

**Larva:** En esta etapa la lepidóptera crece, se desarrolla y almacena nutrientes que luego son utilizados para convertirse en adulto, luego del estado de pupa (Figura 5-2).

La cápsula cefálica presenta en su parte anterior, una estructura triangular denominada frontoclipeo, poseen de dos a cinco pares de patas falsas que brotan de su abdomen y 3 pares de patas que nacen de su tórax, tienen forma cilíndrica y una cabeza bien desarrollada con dos pequeñas antenas, su aparato bucal posee mandíbulas y maxilas además presenta un labio con estructura spinner que se encarga de hilar la seda. Pasan por cinco estadios larvales para alcanzar la siguiente etapa de pupa (Urretabizkaya et al., 2010a, p.13), (Báez y Oromi, 2010, p. 302-318), (Kauffmann, 2002, p.1-3).

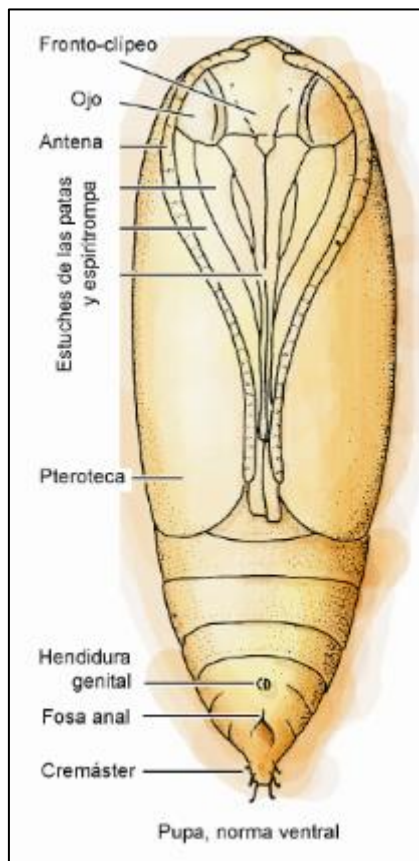


**Figura 5-2:** Larva polípoda eruciforme.  
Fuente: (Urretabizkaya et al., 2010, p.13)

**Pupa:** Más conocida como crisálida estos insectos buscan un lugar adecuado para la pupación y está directamente relacionado con la especie. Según la posición en que se fijan al sustrato las pupas pueden ser sucintas y suspensas, en algunos casos se desarrolla dentro de un capullo sedoso con restos vegetales incorporados y otras se desarrollan dentro de hojas envueltas que

sirven como protección, otras en camaritas de barro en el suelo o bien sin ninguna protección (Figura 6-2).

Ejemplos de pupas en forma de huevo como el bicho de cesto y en estado de larva invernante el gusano del brote del duraznero, el gusano de la pera y la manzana (Urretabizkaya et al., 2010, pp.13-14), (Kauffmann, 2002, pp.1-3), (Báez y Oromi, 2010, pp.302-318).



**Figura 6-2:** Pupa normal ventra.  
Fuente: (Báez y Oromi, 2010, pp.302-318)

**Adulto:** Al terminar la etapa de pupa, las presiones que ejerce el cuerpo desgarran la envoltura pupal dando lugar al último estadio de la *lepidoptera*. Al principio surge la porción cefálica y gradualmente va apareciendo el resto del cuerpo. Luego de un par de horas, se despliegan las alas debido a la presión que ejerce la hemolinfa, endureciendo el tegumento y fijándose los colores definitivos (Kauffmann, 2002, pp.1-3), (Urretabizkaya et al., 2010, p.14).

### 2.2.3. Características Biológicas

**Control endócrino:** Procesos como el crecimiento y la metamorfosis están sujetos a la interacción de un conjunto de hormonas las mismas que están involucradas en el manejo de los ritmos circadianos, el crecimiento, la regulación del desarrollo de la latencia, la regulación del



comportamiento migratorio y otras funciones fisiológicas.

Entre las principales hormonas tenemos la hormona del cerebro (BH) secretada por el proto cerebro que actúa sobre las glándulas ecdiales, la hormona eclosión secretada por células neurosecretoras medianas del cerebro que se almacena en el cuerpo cardíaco y se libera en la hemolinfa durante el cambio de pupa a imago y la hormona juvenil JH la cual es capaz de llevar el progreso juvenil de las especies.

**Polimorfismo:** Representa la diversidad de individuos que poseen diferencias morfológicas en su ciclo de vida dentro de la misma especie. El dimorfismo sexual es el más común, se caracteriza por diferenciar a las especies a partir del sexo (hembra, macho) y se encuentran en las familias *Pieridae*, *Nymphalidae*, *Papilionidae* y *Psychidae*. Entre otras formas de discriminar a los individuos se tienen los rasgos geográficos, estacionales, genéticos y ambientales, incluso el polimorfismo puede estar limitado a un sexo, típicamente la hembra.

**Feromonas:** Son sustancias bioquímicas secretadas por los individuos que se utilizan para la comunicación y afectan las actividades o el funcionamiento en otras especies. Existen sistemas de feromonas que cumple la función específica de atraer parejas.

**Migración:** Los lepidópteros migratorios son una parte esencial del ecosistema ya que desempeñan el papel de indicadores ambientales. Cualquier variación en el entorno puede causar efectos ambientales desfavorables y afectar de manera aguda a las lepidópteras haciendo que estas migren rápidamente en busca de hábitats con estaciones adecuadas. Sus destinos pueden ser las áreas subtropicales, tropicales y todos los continentes, excluida la Antártida (Khan, 2017, pp.1-17), (Urretabizkaya et al., 2010a, pp.1-77).

#### **2.2.4 Clasificación de las Lepidopteras**

Según Kristensen et al. (2007), Kristensen (1999) y Nieuwerkerken et al. (2011) citados en Moreno y García en su estudio titulado Orden *Lepidoptera* (2015) “se reconocen cuatro Subórdenes: *Zeugloptera*, con una familia, *Micropterigidae*, cuyos adultos conservan mandíbulas funcionales y venación homoneura (igual en alas anteriores y posteriores); *Aglossata*, con la familia *Agatiphagidae*, igualmente con venación homoneura, adultos con mandíbulas aunque posiblemente no funcionales, sin ocelos ni chaetosemata; *Heterobathmiina*, también con una única familia, *Heterobathmiidae*, homoneuros y con varias características propias; y finalmente *Glossata*, que incluye la enorme mayoría de las especies del Orden, con adultos que presentan una espiritrompa o probóscide enrollable formada por las gálea

*maxilares (sin embargo en algunos de los grupos basales los adultos presentan tanto venación homoneura como mandíbulas, si bien no funcionales).*

*La subdivisión de Glossata es compleja y comprende 40 superfamilias y 130-140 familias (cifras que varían de acuerdo con diferentes autores)”.*

### **Familias**

- *Hepialidae*
- *Nepticulidae*
- *Mompbidae*
- *Agonoxenidae*
- *Gelechiidac*
- *Argyresthiidae*
- *Sesiidac*
- *Cossidae*
- *Tortricidae*
- *Pyralidae*
- *Thyrididae*
- *Pierophoridae*
- *Noctuidae* (Solomon, 1995, p. 735).

### **Lepidopteras de los Altos Mirandinos**

#### **Familias**

*Hesperiidae*  
*Lycaenidae*  
*Nymphalidae*  
*Papilionidae*  
*Pieridae*  
*Riodinidae.*

#### **Subfamilias**

*Biblidinae*  
*Charaxinae*  
*Coliadinae*  
*Danainae*  
*Dismorphinae*  
*Eudaminae*

*Heliconiinae*  
*Hesperinae*  
*Heteropterinae*  
*Limenitidinae*  
*Nymphalinae*  
*Papilioninae*

*Pierinae*  
*Plyommatinae*  
*Pyrginae*  
*Riodininae*  
*Satyrinae*  
*Theclinae.*

## Especies

<i>Achlyodes pallida</i>	<i>Euptychoides eugenia</i>	<i>Nyctelius nyctelius</i>
<i>Actinote anteas bytia</i>	<i>eugenia</i>	<i>nyctelius</i>
<i>Actinote anteas bytia</i>	<i>Euptychoides saturnus</i>	<i>Oleria makrena makrena</i>
<i>Adelpha irmina irmina</i>	<i>Eurema albula marginella</i>	<i>Oleria phenomoe</i>
<i>Amphiselenis chama</i>	<i>Eurema elathea</i>	<i>phenomoe</i>
<i>Anthanassa drusilla halli</i>	<i>Eurema mexicana</i>	<i>Oressinoma typhla</i>
<i>Antoptus sp</i>	<i>bogotana</i>	<i>Pagyris cymothoe</i>
<i>Arawacus aetolus</i>	<i>Eurema phiale columbia</i>	<i>cymothoe</i>
<i>Arawacus dumenilii</i>	<i>Eurema salome limoneus</i>	<i>Papias subcostulata</i>
<i>Ascia monuste</i>	<i>Eurema xantochlora</i>	<i>Pargyris cymothoe</i>
<i>Athesis clearista clearista</i>	<i>xantochlora</i>	<i>cymothoe</i>
<i>Bolla zorilla</i>	<i>Euselasia venezolana</i>	<i>Parides erithalion Zeuxis</i>
<i>Calephelis laverna</i>	<i>Falga jeconia jeconia</i>	<i>Parides erithalion zeuxis</i>
<i>laverna</i>	<i>Forsterinaria sp</i>	<i>Pedaliodes manis ivica</i>
<i>Caligo oileus</i>	<i>Fountainea nessus</i>	<i>Phoebis sennae</i>
<i>Calycopis sp</i>	<i>Godyris kedema kedema</i>	<i>marcellina</i>
<i>Catasticta flisa briseis</i>	<i>Greta andromica</i>	<i>Pompeius pompeius</i>
<i>Celaenorrhinus eligius</i>	<i>andromica</i>	<i>Pronophila obscura</i>
<i>Charis sp</i>	<i>Heliconius charitonia</i>	<i>Pseudomaniola phaselis</i>
<i>Chioides catillus</i>	<i>bassieri</i>	<i>phaselis</i>
<i>Chlorostrymon simaethis</i>	<i>Heliconius clysonymus</i>	<i>Pteronymia artena beebei</i>
<i>Cobalopsis miaba</i>	<i>Heliconius erato hydara</i>	<i>Pteronymia latilla latilla</i>
<i>Cobalopsis nero</i>	<i>Heliopetes alana</i>	<i>Pteronymia oneida asopo</i>
<i>Consul fabius cecrops</i>	<i>Heliopetes arsalte</i>	<i>Pyrgus adepta</i>
<i>Contrafacia ahola</i>	<i>Hemiargus hanno</i>	<i>Pyrgus oileus</i>
<i>Corades enyo enyo</i>	<i>Hermeuptychia hermes</i>	<i>Pyrisitia proterpia</i>
<i>Cyanophrys sp</i>	<i>Hesperiidae sp marron</i>	<i>Pyrisitia venusta</i>
<i>Cybdelis mnasylyus</i>	<i>claro pareja</i>	<i>Quadrus lugubris lugubris</i>
<i>mnasylyus</i>	<i>Hyaliris coeno coeno</i>	<i>Quadrus u-lucida u-lucida</i>
<i>Cynea sp</i>	<i>Hylephila phyleus phyleus</i>	<i>Remella remus</i>
<i>Dalla c. cypselus</i>	<i>Hypanartia dione</i>	<i>Riodinidae</i>
<i>Dannaus plexippus</i>	<i>Hypanartia lethe</i>	<i>Saliana</i>
<i>Diaethria euclides</i>	<i>Hypothyris euclea euclea</i>	<i>Sostrata grippa</i>
<i>metiscus</i>	<i>Ithomia agnosia agnosia</i>	<i>Spathilepia clonius</i>
<i>Diaeus lacaena</i>	<i>Ithomia iphianassa</i>	<i>Taygetis thamyra</i>
<i>Dione junio</i>	<i>iphianassa</i>	<i>Tegosa anieta anieta</i>
<i>Dircenna jemina jemina</i>	<i>Junonia evarete</i>	<i>Telemiades nicomedes</i>
<i>Dismorphia crisia foedora</i>	<i>Leodonta dysoni dysoni</i>	<i>Thespies othna othna</i>
<i>Dismorphia medora</i>	<i>Leptophobia eleusis</i>	<i>Thoon nr canta</i>
<i>medora</i>	<i>Leptotes cassius</i>	<i>Thyridia psidii aedesia</i>
<i>Dryas iulia alcionea</i>	<i>Lucida lucia scopas</i>	<i>Urbanus dorantes</i>
<i>Dryas iulia alcionea</i>	<i>Lymanopoda caucana</i>	<i>dorantes</i>
<i>Epargyreus aspina</i>	<i>Magneuptychia nebulosa</i>	<i>Urbanus doryssus</i>
<i>Epargyreus exadeus</i>	<i>Mechanitis polymnia</i>	<i>Urbanus viterboana</i>
<i>Episcada hymenaea</i>	<i>veritabilis</i>	<i>Vanessa virginiensis</i>
<i>hymenaea</i>	<i>Mestra hercilia</i>	<i>Vettius coryna catargyra</i>
<i>Eresia carne carne</i>	<i>Milanion leucaspis</i>	<i>Vettius marcus</i>
<i>Eueides aliphera aliphera</i>	<i>Morpho helenor corydon</i>	<i>Xenophanes tryxus</i>
	<i>Nisoniades bessus</i>	<i>Ypthimoides renata</i>

## 2.3 Análisis Multivariante

El análisis multivariante está compuesto por una serie de métodos estadísticos que resumen, clasifican y relacionan un conjunto de variables generalmente correlacionadas que caracterizan a una muestra de individuos.

Gracias a la disponibilidad de estas técnicas y al constante avance de la tecnología se ha logrado desarrollar procedimientos más allá del análisis univariante con el objetivo principal de simplificar la información presentada en las variables con una mínima pérdida de información (Rencher, 2012, p.16).

### 2.3.1 Definiciones básicas

#### Matriz de Datos

Es un arreglo de datos números o variables organizados en dimensiones  $n \times p$  es decir  $n$  filas y  $p$  columnas. Los nombres de las matrices estarán representados por letras mayúsculas en negrita.

Consideremos que sobre los individuos  $w_1, \dots, w_n$  se han observado las variables  $X_1, \dots, X_p$ .

Donde  $x_{ij} = X_j(w_i)$  representa las observaciones de la variable  $X_j$  sobre el individuo  $w_i$ . (Peña, 2002, pp. 68-69).

Además, indicaremos que:

- $x_i$  la fila  $i$ -ésima de  $X$ , que utilizaremos como vector columna.
- $X_j$  es la columna  $j$ -ésima de  $X$ .
- $\bar{x} = (\bar{x}_1, \dots, \bar{x}_j, \dots, \bar{x}_p)'$  es el vector columna de las medias de las variables, la cual se expresa de la siguiente forma:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

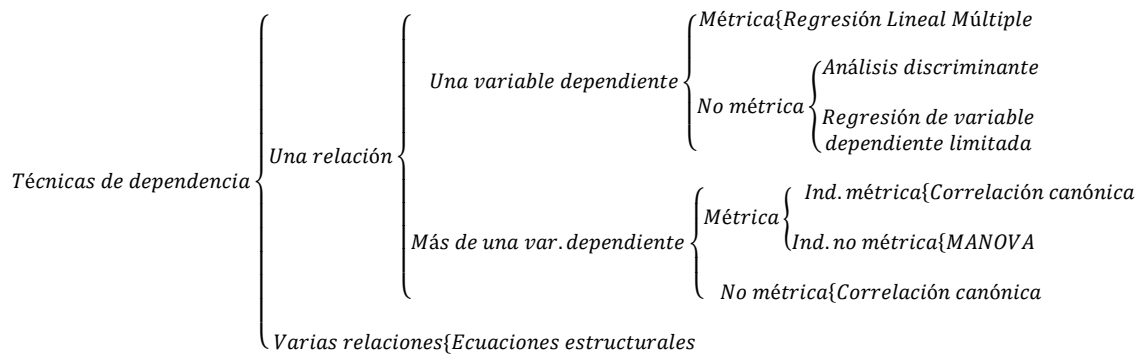
### 2.3.2 Clasificación del Análisis Multivariante

Como lo mencionan Jiménez y Manzano (2005, pp.4-10), en su libro “Análisis Multivariante Aplicado” las técnicas multivariantes pueden dividirse en dos grandes grupos:

- Técnicas explicativas o de dependencia
- Técnicas descriptivas o de interdependencia.

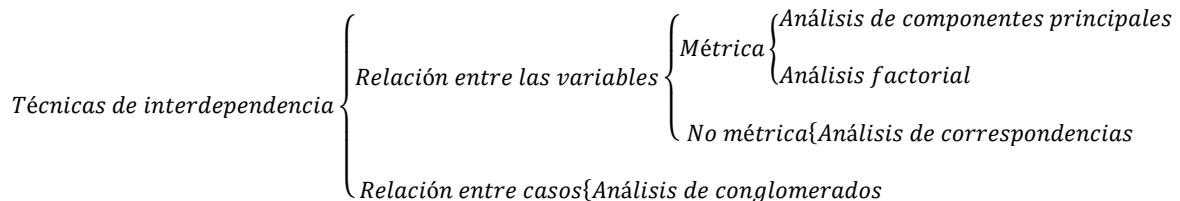
## Técnicas explicativas o de dependencia

El objetivo de estas técnicas radica en estudiar la posible relación que existe entre variables dependientes e independientes, es decir explicar si el conjunto de variables independientes influyen de manera individual o conjunta en las variables dependientes (Jiménez y Manzano, 2005).



## Técnicas descriptivas o de Interdependencia

Estas técnicas permiten determinar cómo y por qué están relacionadas las variables entre sí y se utilizan cuando no es posible diferenciar las variables dependientes de las independientes (Jiménez y Manzano, 2005).



### 2.3.3 Observaciones extremas

Se consideran observaciones extremas aquellas que distan del patrón estructural de un conjunto de datos haciendo que el investigador las considere como posibles observaciones ilegítimas que arrojan falsos resultados en el estudio.

Entre las causas que originan esta problemática tenemos:

- Observaciones extremas que surgen de un error de procedimiento, ejemplo: entrada errónea de datos. En caso de encontrarse con este tipo de observaciones de inmediato se debe proceder con la eliminación o recodificación a datos ausentes.
- Observaciones que ocurren a causa de un acontecimiento extraordinario, es decir que existe una explicación de la ocurrencia de este evento y el investigador deberá tomar la decisión de eliminar o no la observación extrema apoyándose en su sentido común.

- Observaciones extraordinarias que no tienen explicación y el investigador se sentiría inclinado a omitirlas del análisis a menos que no sean representadas para la población (Silva y Barroso, 2004, pp. 1-173).

Además Rousseeuw y Hubert (2011, pp.73-79) en su estudio “**Robust statistics for outlier detection**” establecen el método robusto de detección de valores extremos con un factor de corrección de 1.483 definido en la siguiente ecuación:

$$\frac{|x_i - med(x)|}{(Meda(x) * 1.483)} > 2.5$$

donde **med(x)** es la mediana de las observaciones (Estimador robusto del centro de los datos), Meda es la mediana de las desviaciones absolutas  $|x_i - med(x)|$  (Medida robusta de la dispersión). Este método puede verse como una estandarización robusta de los datos (Peña, 2002, pp.122-125).

#### 2.3.4 Normalidad multivariante

La mayoría de las técnicas multivariadas trabajan bajo la suposición de que el conjunto de datos posee normalidad conjunta, sin embargo verificar normalidad univariante para cada una de las variables no es un condicionante de variabilidad multivariante conjunta (si la distribución conjunta es normal multivariante cada una de las marginales es una normal univariante, pero no a la inversa).

Por tal razón luego de corroborar el supuesto de normalidad de cada una de las variables por naturaleza univariante, es necesario también constatar la hipótesis de normalidad multivariante, por ello Mardia (1970, pp.519-530) en su estudio “**Measures of Multivariate Skewness and Kurtosis with Applications**” propuso un test para contrastar si la simetría y la curtosis multivariantes del conjunto de variables observadas implican o no la hipótesis de normalidad.

El proceso de elaboración de dicho contraste se sujeta a las medidas muestrales de asimetría y curtosis multivariante.

$$\text{Asimetría: } G_{1/p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [(X_i - \bar{X})' S^{-1} (X_j - \bar{X})]^3$$

$$\text{Curtosis: } G_{2/p} = \frac{1}{n} \sum_{i=1}^n [(X_i - \bar{X})' S^{-1} (X_i - \bar{X})]^2$$

donde:

$n$ : representa el número total de observaciones.

$X_i$  y  $X_j$ : son vectores columna que contienen los valores de todas las variables para las observaciones  $i$  –ésima y  $j$  –ésima, respectivamente.

$\bar{X}$ : es el correspondiente vector columna de medias muestrales.

$S^{-1}$ : es la inversa de la matriz de varianzas y covarianzas muestrales.

Luego se calcula el estadístico  $K_p^2$  para contrastar simetría y mesocurtosis multivariante:

$$k_p^2 = [G_{1,p}]^2 + [G_{2,p}]^2 \sim \chi_2^2$$

por último, el criterio de decisión es:

$$\text{Si } K_2^2 > \chi_2^2 \quad \text{se rechaza } H_0$$

teniendo como hipótesis nula que el conjunto de datos sigue una distribución normal multivariante (Mangin et al., 2006, pp. 1-530).

### **2.3.5 Análisis de Componentes Principales**

Es una técnica de reducción de datos que tiene como objetivo principal sintetizar la información de un grupo de variables originales a través de la obtención de sus combinaciones lineales, dichas combinaciones reciben el nombre de componentes principales. Estas componentes recogen tanta información como sea posible con una mínima pérdida de información y en un pequeño número de componentes (Closas et al., 2013, pp.65-92) (Zamora y Esnaola, 2007, pp.121-159).

#### *2.3.5.1 Definición y obtención de las componentes principales*

Sea  $X = [X_1, \dots, X_p]$  una matriz de datos multivariantes las componentes principales son variables compuestas incorrelacionadas que se van construyendo según el orden de importancia de acuerdo a la variabilidad total (Cuadras, 2007, pp.1-249).

El Análisis de Componentes Principales (ACP) es una técnica estadística que no necesita cumplir con el supuesto de normalidad multivariante en los datos.

#### **Construcción de las componentes principales**

Partiendo de  $X = (X_1, \dots, X_n)'$  con matriz de covarianzas  $Cov(X) = \Sigma$  semidefinida positiva y con  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$ , las raíces características correspondientes. Se definen las siguientes combinaciones lineales:

$$\begin{cases} Y_1 = l'_1 X = l'_{11} X_1 + \dots + l'_{1p} X_p \\ \vdots \\ Y_p = l'_p X = l'_{p1} X_1 + \dots + l'_{pp} X_p \end{cases}$$

con el vector aleatorio  $Y = (Y_1, \dots, Y_p)'$ . Dadas dos cualesquiera de sus componentes,  $i$  y  $j$ , se tiene:

$$Var(Y_i) = l'_i \Sigma l_i, Cov(Y_i, Y_j) = l'_i \Sigma l_j$$

para todo vector  $X$ , independientemente de su distribución (Gutierrez y Gutierrez-Sanchez, 2005, pp. 1-37).

Considerando  $Y_1$  como la combinación lineal de varianza máxima tal que  $Var(Y_i) = l'_i \Sigma l_i$  con restricción de vectores  $l$  unitarios para todas las componentes principales, por lo tanto  $l'_i l_i = 1$ .

La primera componente principal será la combinación lineal  $Y_1 = l'_1 X$  con varianza máxima  $Var(Y_1)$  y restricción  $l'_1 l_1 = 1$ .

La segunda componente principal será la combinación lineal  $Y_2 = l'_2 X$  con varianza máxima  $Var(Y_2)$  y restricción  $l'_2 l_2 = 1$  y además incorrelada con  $Y_1$ , es decir:

$$Cov(l'_1 X, l'_2 X) = 0$$

manteniendo el mismo procedimiento se construirá las  $p$  combinaciones lineales  $Y_1, \dots, Y_p$  con máxima  $Var(l'_i X)$  y bajo las siguientes condiciones:

- $l'_i l_i = 1$
- $Cov(l'_i X, l'_k X) = 0$  para  $k < i$ .

### 2.3.5.2 Cálculo del Primer Componente

La primera componente principal se almacenará en:

$$Y_1 = e'_1 X, e'_1 e_1 = 1$$

tal que:

$$Var(Y_1) = Var(l'X), \max Var(l'X) = Var(e'_1 X) = e'_1 \Sigma e_1$$

estando ante el Problema de Lagrange de maximización condicionada:

$$\left\{ \begin{array}{l} \text{máx}\{l' \Sigma l\} \\ l' l = 1 \end{array} \right\} \Rightarrow \Phi_1(l) = l' \Sigma l - \lambda(l' l - 1) \Rightarrow \frac{\partial \Phi_1(l)}{\partial l} = 2 \Sigma l - 2 \lambda l = 0 \Rightarrow (\Sigma - \lambda I) l = 0$$



asumiendo que  $\Sigma_{p \times p}$  tiene autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$ , con auto vectores asociados  $e_1, e_2, \dots, e_p$  y como  $l' \Sigma l = \lambda l' l = 1, Var(l' \Sigma l) = \lambda$  y es claro que tomando  $l = e_1$ , correspondiente al mayor autovalor, dando como resultado la primera componente principal

$$Y_1 = e_1' \text{ con } Var(Y_1) = \lambda_1.$$

### 2.3.5.3 Cálculo del segundo componente

Según la definición anterior, la segunda componente principal será la combinación lineal  $Y_2 = l' X$ , incorrelada con  $Y_1$  con varianza máxima. Por tal razón (Gutierrez y Gutierrez-Sanchez, 2005, pp. 1-37):

$$\begin{aligned} & \text{máx}\{l' X l\}, \text{ con } l' l = 1, l' \Sigma e_1 = 0 \\ \Phi_2(l) = l' \Sigma l - \lambda(l' l - 1) - 2v(l' \Sigma e_1) & \Rightarrow \frac{\partial \Phi_2(l)}{\partial l} = 2 \Sigma l - 2 \lambda l - 2v \Sigma e_1 \end{aligned}$$

la solución será obtener la raíz  $\lambda_2$ , el segundo autovalor mayor con el correspondiente auto vector  $e_2$  de modo que la segunda componente principal estará definida por:

$$Y_2 = e_2' X \text{ con } Var(Y_2) = \lambda_2.$$

### 2.3.5.4 Cálculo de la $(r + 1)$ -ésima componente principal $(1 \leq r + 1 \leq p)$

Para este propósito tenemos:

$$\begin{aligned} Y_{r+1} &= l' X; l' l = 1; l' \Sigma e_i = 0, i = 1, \dots, r \\ \Phi_{r+1}(l) &= l' \Sigma l - \lambda(l' l - 1) - 2 \sum_{i=1}^r v_i l' \Sigma e_i \end{aligned}$$

se demuestra que, siendo  $\lambda_i \neq 0, i = 1, \dots, r$  y  $v_i = 0, i = 1, \dots, r$  su maximización es:

$$\{2 \Sigma l - 2 \lambda l = 0, \Sigma l - \lambda l = 0, (\Sigma - \lambda I) l = 0\}$$

con  $\lambda_{r+1} \neq 0$ , se puede seleccionar  $\lambda = \lambda_{r+1}, l = e_{r+1}$  dando como resultado la  $(r + 1)$  -ésima componente principal que se define de la siguiente manera:

$$Y_{r+1} = e_{r+1}' X, Var(Y_{r+1}) = \lambda_{r+1}.$$

En el caso de que se tenga  $\lambda_{r+1} = 0, \lambda_i \neq 0, i \neq r + 1$ , se genera una combinación lineal de  $\alpha_{r+1}$  y  $\alpha_i$  en la que  $\alpha_i \neq 0$ .

Finalmente  $A = (e_1, e_2, \dots, e_p)$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ ,  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$ , como  $A'A = I$ , se cumple que  $A'\Sigma A = \Lambda$ .

### Elaboración de las $p$ componentes principales conjuntas.

Esta técnica se presenta como una alternativa globalizada al método anterior antes explicado obteniendo los mismos resultados (Gutierrez y Gutierrez-Sanchez, 2005, pp. 1-37).

### Lema de maximización

Partiendo de una  $A$  con dimensiones  $p \times p$  definida positiva, con autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$  y autovalores normalizados  $e_1, \dots, e_p$ , y sea  $x$  un vector con dimensiones  $p \times 1$ , arbitrario no nulo entonces se cumple que:

$$\text{máx} \frac{x'Ax}{x'x} = \lambda_1, \quad \text{alcanzado en } x = e_1,$$

$$\text{mín} \frac{x'Ax}{x'x} = \lambda_p, \quad \text{alcanzado en } x = e_p,$$

$$\text{máx} \frac{x'Ax}{x'x} = \lambda_{k+1}, \quad \text{alcanzado en } x = e_{k+1} \quad k = 1, 2, \dots, p-1.$$

Se justifica el siguiente teorema:

### Teorema de componentes principales

Siendo  $X = (X_1, \dots, X_p)'$  un vector aleatorio con matriz de covarianza conocida  $\Sigma$  definida positiva y real y sean  $(\lambda_i, e_i)$  los autovalores-auto vectores de  $\Sigma$  con  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$ .

La  $Y_i$   $i$ -ésima componente principal antes mencionada viene dada por:

$$Y_i = e_i'X = e_{i1}X_1 + \dots + e_{ip}X_p, \quad i = 1, \dots, p,$$

con varianza  $\text{Var}(Y_i) = e_i'\Sigma e_i = \lambda_i$  y covarianza  $\text{Cov}(Y_i, Y_j) = e_i'\Sigma e_j = 0, i \neq j$ .

Se pueden obtener componentes principales no únicas en el caso de que se presenten autovalores iguales  $\lambda_k$  y  $e_k$  asociados no únicos.

Sea la matriz  $\mathbb{P} = (e_1, \dots, e_p)$  caracterizada por poseer a los autovectores normalizados  $e_i$  como sus columnas y dada una matriz  $B$ , con dimensiones  $p \times p$  definida positiva con descomposición espectral  $B = \sum_{i=1}^p \lambda_i e_i e_i'$ , entonces:

$$B = \mathbb{P}\Lambda\mathbb{P}' \sum_{i=1}^p \lambda_i e_i e_i' \text{ sujeto a } \mathbb{P}\mathbb{P}' = I \text{ y } \lambda = \text{diag}(\lambda_1, \dots, \lambda_p).$$

Al ser  $B^{-1} = \mathbb{P}\Lambda^{-1}\mathbb{P}' = \sum_{i=1}^p \sqrt{\lambda_i} e_i e_i'$  se puede aplicar la raíz cuadrada de  $B$  y utilizar esta descomposición precisando la matriz  $B^{1/2}$  en el caso en que  $\lambda_p > 0$  (Gutierrez y Gutierrez-Sanchez, 2005, pp. 1-37).

### 2.3.5.6 Propiedades de las Componentes Principales

Las propiedades de la nueva matriz de componentes principales son las siguientes:

- Conservan la variabilidad inicial: la suma de las varianzas de los componentes principales es igual a la suma de las varianzas de las variables originales.
- La proporción de variabilidad explicada por cada componente es el cociente entre sus varianzas.
- Las covarianzas entre componentes principales y las variables  $X$  se obtienen a partir del producto de las coordenadas del vector propio por el valor propio.
- La correlación entre un componente principal y una variable  $X$  es proporcional al coeficiente de esa variable, y el coeficiente de proporcionalidad es el cociente entre la desviación típica de la componente y la variable.

$$Corr(y_i; x_j) = \frac{Cov(y_i, x_j)}{\sqrt{Var(y_i)Var(x_j)}}$$

- Las  $r$  componentes principales ( $r < p$ ) proporcionan la predicción lineal óptima con  $r$  variables del conjunto de variables  $X$ .
- Si estandarizamos los componentes principales, a través del cociente de cada componente para su desviación típica, se obtiene la estandarización multivariante de los datos originales. Estandarizando los componentes  $Y$  por sus desviaciones típicas, se obtienen las nuevas variables:

$$Y_c = YD^{-1/2} = XAD^{-1/2}$$

donde:

$D^{-1/2}$  es una matriz con las inversas de las desviaciones típicas de las componentes (Peña, 2002, pp.149-153).

## Variabilidad explicada por las componentes principales

Con  $\lambda_i$  como varianza de la componente principal  $Y_i$  y  $tr(\mathbf{S}) = \sum_{i=1}^p \lambda_i$  de variación total se tiene que:

- $\lambda_i$  es el valor de variabilidad que contribuye la variación total de  $Y_i$ .
- Con  $q < p$  tal que  $Y_1, \dots, Y_q$  se cumple que  $\sum_{i=1}^q \lambda_i$  será el valor de que aporte a la variación total  $tr(\mathbf{S})$ .
- Las primeras componentes principales estarán sujetas a explicar la variabilidad de las variables de acuerdo al siguiente porcentaje:

$$P_m = 100 \frac{\lambda_1 + \dots + \lambda_m}{\lambda_1 + \dots + \lambda_p}$$

si el porcentaje de variabilidad de las dos primeras componentes principales tiene un valor del 90% significa que estas componentes explican una gran parte de la variabilidad de las variables y se puede sustituir  $X_1, X_2, \dots, X_p$  por las dos primeras componentes principales  $Y_1, Y_2$ . (Cuadras, 2007, pp.64-66).

### 2.3.6 Análisis Factorial

Es una técnica multivariante de reducción de datos que tiene como objetivo principal identificar aquellas variables latentes (factores) que ocasionan que las variables observadas se correlacionen.

#### Modelo de Análisis Factorial

##### Elementos del Modelo:

- El vector  $X$  ( $p \times 1$ ) con variables  $X_i$   $i = 1, \dots, p$ , de la población.
- El vector de factores comunes  $F$  ( $m \times 1$ ) el cual contiene los factores comunes.
- El vector de errores o de factores específicos  $\epsilon$ , ( $p \times 1$ ), que vincula los errores  $\epsilon_i, i = 1, \dots, p$  a cada variable  $X_i$ .
- La matriz de factor loadings  $L$ , ( $p \times m$ ) de rango  $m$  con los “loadings” de la variable  $X_i$  con respecto a  $F_j$  el factor común.

### Características de los elementos del Modelo

Los vectores  $X, F$  y  $\epsilon$  comparten la característica de ser aleatorios mientras que la matriz  $L$  es no aleatoria sin ninguna restricción a sus elementos  $l_{ij}$ , por ejemplo, no se estipula un valor nulo a determinados factores-loadings lo que hace que el modelo sea no restringido (“unrestricted”).

Finalmente,  $X$  es observable y el vector de factores  $F$  y el vector de errores  $\epsilon$  son no observables.

### Hipótesis probabilísticas sobre los elementos del Modelo

Se verifica:

$E[X] = \mu$  ( $p \times 1$ ) como el de vector media poblacional

$Cov(X) = \Sigma$  con dimensiones  $p \times p$  representa a la matriz definida positiva de covarianzas de la población:

$$E[F] = 0; E[\epsilon] = 0; Cov[F] = E[FF'] = \Phi, m \times m$$

$Cov[\epsilon] = \psi$ ,  $p \times p$  con  $\psi = diag(\psi_1, \dots, \psi_p)$ , siendo  $\psi_i, i = 1, \dots, p$ , las varianzas específicas o unicidades (Gutierrez y Gutierrez-Sanchez, 2005, pp. 1-62):

$$Cov[F; \epsilon] = E[F\epsilon'] = 0$$

concluyendo, que los vectores de factores  $F$  y errores  $\epsilon$  son incorrelados, centrados y presentan matrices de covarianzas  $\Phi$  y  $\psi$  respectivamente, con  $\epsilon_i$  heterocedásticos e incorrelados.

$\Phi = Cov(F)$  se presenta como una matriz no necesariamente diagonal, con los elementos de  $F$  correlados.

Tomando en cuenta las consideraciones anteriores  $X, F, \epsilon$  y  $L$  conforman un modelo lineal que se define como:

$$X = \mu + LF + \epsilon$$

el modelo de Análisis Factorial implica un conjunto de relaciones lineales del tipo:

$$\begin{cases} X_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \epsilon_1 \\ \vdots \\ X_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \epsilon_p \end{cases}$$

en donde  $(F_1, \dots, F_m)$  en conjunto explican una parte de todas las variables  $X_i, i = 1, \dots, p$ , y la

parte no explicada mediante los factores recae en los  $\epsilon_i$  asociados a la respectiva  $X_i$ .

Formalmente, el conjunto de ecuaciones lineales mencionado anteriormente presenta características múltiples al tener  $m$  “regresores”  $F_j, j = 1, \dots, m$ ; y características multivariantes al existir  $p$  variables  $X_i, i = 1, \dots, p$  “dependientes” que básicamente representa un modelo de regresión lineal múltiple multivariante sin embargo hay una diferencia primordial en cuanto al Análisis Factorial y Regresión Lineal pues los  $F_j$  son aleatorios y no observables.

### 2.3.6.1 Resultados sobre el Modelo de Análisis Factorial

Igualdad estructural de la matriz de covarianzas  $\Sigma$  en el Modelo de Análisis Factorial

$$E[(X - \mu)(X - \mu)'] = \Sigma = E[(LF + \epsilon)(LF + \epsilon)'] = E[LFF'L' + LF\epsilon' + \epsilon F'L' + \epsilon\epsilon'] =$$

$$LE[FF']L + LE[F\epsilon'] + E[\epsilon F']L' + E[\epsilon\epsilon'] = L\Phi L' + \psi$$

De esta manera la matriz de covarianzas de la población,  $\Sigma$  se estructura en términos de la terna  $(L; \Phi, \psi)$ , mediante  $\Sigma = L\Phi L' + \psi$ .

En consecuencia, modelizar la población  $X = (X_1, \dots, X_p)$ , en función del Análisis Factorial, implica encontrar las matrices  $L; \Phi$  y  $\psi$  con las condiciones exigidas en el modelo (Gutierrez y Gutierrez-Sanchez, 2005, pp. 1-62).

Por otro lado la matriz  $\Sigma$  ( $p \times m$ ) entre  $X$  y  $F$  es:

$$Cov[X; F] = E[(X - \mu)F'] = E[(LF + \epsilon)F'] = E[LFF'] + E[\epsilon F'] = LE[FF'] = L\Phi$$

suponiendo  $\Sigma$  con estructura  $\Sigma = L\Phi L' + \psi$  mediante la terna  $(L; \Phi, \psi)$  se verifica que si  $C$  es una matriz cualquiera con dimensiones  $m \times m$  y característica no singular, se cumple que la terna  $(LC; C^{-1}\Phi C^{-1})$  estructura  $\Sigma$  de manera semejante.

Justamente:

$$\Sigma = L\Phi L' + \psi = LCC^{-1}\Phi C^{-1}C'L' + \psi = (LC)(C^{-1}\Phi C^{-1})(C'L') + \psi =$$

$$(LC)(C^{-1}\Phi C^{-1})(LC)' + \psi = L^*\Phi^*L^{*'} + \psi$$

Por otra parte  $LC$  y  $C^{-1}\Phi C^{-1}$  cumplen las condiciones exigidas a  $L$  y  $\Phi$ .

$$(LC)$$

- es  $(p \times m)$ ;  $C^{-1}\Phi C^{-1}$ ,

- es  $m \times m$  y
- es la matriz de covarianzas de  $F^* = C^{-1}F$ ,  $m \times m$ .

En realidad,  $E[C^{-1}F] = C^{-1}E[F] = 0$ ; y  $Cov[C^{-1}F] = E[C^{-1}FF'C^{-1}] = C^{-1}E[FF']C^{-1} =$

$$C^{-1}\Phi C^{-1}$$

Lo que significa que lugar de:

$$X - \mu = LF + \epsilon; \quad (L; F; \epsilon)$$

podemos considerar un modelo equivalente

$$X - \mu = LCC^{-1}F + \epsilon = (LC)(C^{-1}F) + \epsilon$$

y los “nuevos factor-loadings” ( $LC$ ) y los “nuevos factores” ( $C^{-1}F$ ), factorizan a la matriz de covarianzas de manera semejante.

En otras palabras, si existe un modelo factorial con la dupla ( $L; F$ ) cualquier otro modelo definido en la forma ( $L^* = LC; F^* = C^{-1}F; \epsilon$ ) estructura la matriz de covarianza de manera semejante por lo tanto la terna ( $L^*; F^*; \epsilon$ ) también sería una solución equivalente (Gutierrez y Gutierrez-Sanchez, 2005, pp. 1-62).

### 2.3.7 Análisis de Correspondencias

El Análisis de Correspondencias (AC) es una técnica estadística multivariante que estudia la relación existente entre las categorías de dos variables no métricas y las representa en un espacio multidimensional reducido. Esta técnica estadística se asemeja al Análisis de Componentes Principales y Coordenadas Principales para variables cualitativas (Closas et al., 2013, pp. 65-92).

Dada una tabla de contingencia que cataloga una misma población con clase  $i$  de la característica  $I$  y clase  $j$  de la característica  $J$  medida en un número de individuos  $k_{ij}$  se busca transformarla teniendo en cuenta su naturaleza y características, de tal manera que se le puedan aplicar métodos generales del análisis multivariante.

#### Transformación de una tabla de contingencia bruta

Dada una tabla de contingencia con elementos  $k_{ij}$  consideramos la siguiente transformación

$$K = \sum_{ij} k_{ij}$$

$$f_{ij} = \frac{k_{ij}}{K}$$

$$f_{i.} = \sum_{j=1}^p f_{ij} = \sum_{j=1}^p \frac{k_{ij}}{K}$$

$$f_{.j} = \sum_{i=1}^n f_{ij} = \sum_{i=1}^n \frac{k_{ij}}{K}$$

### Nubes de puntos en $\mathbb{R}^p$ y en $\mathbb{R}^n$

Supóngase una matriz de datos con dimensiones  $n \times p$ , con  $(n)$  filas y  $(p)$  columnas fijas se definen los Perfiles-fila y Perfiles-columna.

**Perfiles-fila**  $\left\{ \left( \frac{f_{ij}}{f_{i.}} \right); j = 1, \dots, p \right\}_i$  se tienen  $n$  perfiles-filas, cada uno afectado de la masa  $f_{i.}$

**Perfiles-columna**  $\left\{ \left( \frac{f_{ij}}{f_{.j}} \right); i = 1, \dots, p \right\}_j$  donde la masa de cada perfil-columna es  $f_{.j}$  (Gutierrez y

R., 2005, pp. 1-60).

#### 1. Tabla de Perfiles-Fila

**Tabla 1-2:** Perfiles-fila.

	1	...	j	...	p	Masas
1	$\frac{f_{11}}{f_{1.}}$	...	$\frac{f_{1j}}{f_{1.}}$	...	$\frac{f_{1p}}{f_{1.}}$	$f_{1.}$
:	$f_{1.}$		$f_{1.}$		$f_{1.}$	
:			:			
i	$\frac{f_{i1}}{f_{i.}}$	...	$\frac{f_{ij}}{f_{i.}}$	...	$\frac{f_{ip}}{f_{i.}}$	$f_{i.}$
:	$f_{i.}$		$f_{i.}$		$f_{i.}$	
:			:			
n	$\frac{f_{n1}}{f_{n.}}$	...	$\frac{f_{nj}}{f_{n.}}$	...	$\frac{f_{np}}{f_{n.}}$	$f_{n.}$
	$f_{n.}$		$f_{n.}$		$f_{n.}$	

Fuente: (Gutierrez y R., 2005, pp. 1-60)

#### 2. Tabla de Perfiles-Columna



**Tabla 2-2:** Perfiles columna.

	1	...	j	...	p
1	$\frac{f_{11}}{f_{.1}}$	...	$\frac{f_{1j}}{f_{.j}}$	...	$\frac{f_{1p}}{f_{.p}}$
:	$\frac{f_{i1}}{f_{.1}}$	...	$\frac{f_{ij}}{f_{.j}}$	...	$\frac{f_{ip}}{f_{.p}}$
:			:		
i	$\frac{f_{i1}}{f_{.1}}$	...	$\frac{f_{ij}}{f_{.j}}$	...	$\frac{f_{ip}}{f_{.p}}$
:			:		
n	$\frac{f_{n1}}{f_{.1}}$	...	$\frac{f_{nj}}{f_{.j}}$	...	$\frac{f_{np}}{f_{.p}}$
Masas	$f_{.1}$	...	$f_{.j}$	...	$f_{.p}$

Fuente: (Gutierrez y R., 2005, pp. 1-60)

Es así que se tienen 2 nubes de puntos: la primera nube conformada por  $n$  puntos en  $\mathbb{R}^p$  de coordenadas:

$$\left( \frac{f_{i1}}{f_{.1}}, \frac{f_{i2}}{f_{.2}}, \dots, \frac{f_{ij}}{f_{.j}}, \dots, \frac{f_{ip}}{f_{.p}} \right); i = 1, \dots, n$$

$$\left( \frac{f_{1j}}{f_{.j}}, \frac{f_{2j}}{f_{.j}}, \dots, \frac{f_{ij}}{f_{.j}}, \dots, \frac{f_{nj}}{f_{.j}} \right); j = 1, \dots, p$$

y estos puntos están afectados de masa:  $f_{.i}$  y  $f_{.j}$  respectivamente.

Por otro lado, los  $n$  puntos-fila anteriores se encuentran en un subespacio  $p - 1$  dimensional de  $\mathbb{R}^p$ , con relación baricéntrica:

$$\sum_{j=1}^p \left( \frac{f_{ij}}{f_{.i}} \right) = 1; i = 1, \dots, n$$

se cumple también para los puntos-columna:

$$\sum_{i=1}^n \left( \frac{f_{ij}}{f_{.j}} \right) = 1; j = 1, \dots, p$$

de modo que estos  $p$  puntos en  $\mathbb{R}^n$ , se mantienen en un subespacio  $(n - 1)$  dimensional de  $\mathbb{R}^n$ .

### Distancias en $\mathbb{R}^p$ y $\mathbb{R}^n$

Para determinar la distancia entre dos perfiles-fila, correspondientes a los individuos  $i$  e  $i'$ , se puede hacer uso de la distancia  $\chi^2$  (chi-cuadrado):

$$d^2(i, i') = \sum_{j=1}^p \left( \frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \frac{1}{f_{.j}}$$

y para la distancia entre perfiles-columna,  $j$  y  $j'$ :

$$d^2(j, j') = \sum_{i=1}^n \left( \frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2 \frac{1}{f_{i.}}$$

la distancia chi-cuadrado no es la única herramienta a considerar más bien es usual en el análisis de correspondencias clásico y una de sus ventajas es el cumplimiento del Principio de Equivalencia Distribucional el cual sostiene que si dos puntos-fila se confunden en uno solo, en  $\mathbb{R}^p$  o en  $\mathbb{R}^n$  y se consideran englobados en un solo punto con masa, la suma de las masas de ambos puntos-fila son invariantes a las distancias entre los demás en  $\mathbb{R}^p$  y  $\mathbb{R}^n$ .

La distancia  $\chi^2$  es la natural considerando el espacio de puntos de los perfiles fila (o el de las columnas) como un espacio euclídeo ponderado con métrica definida por la matriz  $\mathbb{D}_c^{-1} = \text{diag}\left(\frac{1}{f_{.j}}\right)$ , estando los puntos del correspondiente espacio afectados por las masas (ponderaciones) asociadas a cada punto, y antes indicadas (Gutierrez y R., 2005, pp. 1-60).

Respecto al Principio de equivalencia distribucional la distancia entre puntos-fila o puntos-columna dada por la métrica  $\chi^2$  en  $\mathbb{R}^p$  y  $\mathbb{R}^n$  respectivamente se define con las ecuaciones de distancias:

$$d^2(i, i') = \sum_{j=1}^p \left( \frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \frac{1}{f_{.j}}$$

$$d^2(j, j') = \sum_{i=1}^n \left( \frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2 \frac{1}{f_{i.}}$$

en términos de los Espacios Euclídeos Ponderados, estas distancias elevadas al cuadrado no son más que las distancias entre puntos de los euclídeos  $\mathbb{R}^p$  y  $\mathbb{R}^n$  ponderados elevados al cuadrado dadas por:

$$d^2(\bar{i}, \bar{i}') = (\bar{i} - \bar{i}')' \mathbb{D}_{\mathbb{R}^p}^{-1} (\bar{i} - \bar{i}') ; \mathbb{D}_{\mathbb{R}^p} = \text{diag}(f_{.j})$$

$$d^2(\bar{j}, \bar{j}') = (\bar{j} - \bar{j}')' \mathbb{D}_{\mathbb{R}^n}^{-1} (\bar{j} - \bar{j}') ; \mathbb{D}_{\mathbb{R}^n} = \text{diag}(f_{i.})$$

(donde  $\bar{i}$  y  $\bar{i}'$  son puntos-vectores  $\mathbb{R}^p$  y  $\bar{j}$  y  $\bar{j}'$  de  $\mathbb{R}^n$  y  $\mathbb{D}_{\mathbb{R}^p}^{-1}$  y  $\mathbb{D}_{\mathbb{R}^n}^{-1}$  definen las métricas en  $\mathbb{R}^p$  y  $\mathbb{R}^n$ ).

En el caso de que los puntos-fila  $i$  e  $i'$  se hallen superpuestos en  $\mathbb{R}^p$  se tiene que:

$$\frac{f_{ij}}{f_i} = \frac{f_{i'j}}{f_{i'}} \quad j = 1, \dots, p$$

lo que implica que existen perfiles iguales en  $\mathbb{R}^p$  y estos coinciden con el de un punto-fila,  $i^*$  tal que:

$$\frac{f_{ij}}{f_i} = \frac{f_{i'j}}{f_{i'}} = \frac{f_{i^*j}}{f_{i^*}}$$

entonces si consideramos un punto-fila  $i^*$ , con masa la suma de las masas,  $f_i + f_{i'} = f_{i^*}$  se concluye que:

$$\frac{f_{ij} + f_{i'j}}{f_i + f_{i'}} = \frac{f_{i^*j}}{f_{i^*}}$$

de modo que al tener denominadores iguales ( $f_i + f_{i'} = f_{i^*}$ ), la suma de las sumas de  $i$  e  $i'$ , equivale a considerar el punto-suma de las filas correspondiente a los puntos  $i$  e  $i'$  en la matriz de correspondencias  $f_{ij}$ .

De este modo si se sustituye dos filas ( $i$  e  $i'$ ) de la matriz original (de las  $f_{ij}$ ) por una única fila cuyo origen fue la suma de ambas filas dada por la suma de las masas correspondientes.

¿La superposición realizada tiene algún efecto en las distancias entre puntos-fila en  $\mathbb{R}^p$ , o/y la distancia entre puntos-columna en  $\mathbb{R}^n$ ?

1. Las distancias entre puntos-fila no se ven perturbadas por efecto de la superposición pues la distancia al cuadrado  $x^2$  no se altera si se consideran dos puntos-fila arbitrarios, ya que las ponderaciones  $f_{.j} = \sum_i f_{ij}$  no cambian.

2. En cuanto a las distancias al cuadrado en  $\mathbb{R}^n$ ;  $d^2(j, j')$  contiene los sumandos correspondientes a las dos filas que se fusionaron en una única. Estos dos sumandos se expresan de la siguiente manera:

$$\frac{1}{f_i} \left( \frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2 + \frac{1}{f_{i'}} \left( \frac{f_{i'j}}{f_{.j}} - \frac{f_{i'j'}}{f_{.j'}} \right)^2$$

pero esta expresión es igual a:

$$f_i \left( \frac{f_{ij}}{f_i f_{.j}} - \frac{f_{ij'}}{f_i f_{.j'}} \right)^2 + f_{i'} \left( \frac{f_{i'j}}{f_i f_{.j}} - \frac{f_{i'j'}}{f_i f_{.j'}} \right)^2 = f_i B + f_{i'} B = (f_i + f_{i'}) B = f_{i^*} B$$

las cantidades entre paréntesis se conservan iguales dado que  $\frac{f_{ij}}{f_i} = \frac{f_{i'j}}{f_{i'}}$ . Sin embargo en el proceso de en la superposición los dos términos se reemplazan por:

$$\frac{1}{f_{i^*}} \left( \frac{f_{i^*j}}{f_j} - \frac{f_{i^*j'}}{f_{j'}} \right)^2 = f_{i^*} \left( \frac{f_{i^*j}}{f_{i^*} \cdot f_j} - \frac{f_{i^*j'}}{f_{i^*} \cdot f_{j'}} \right)^2$$

que es igual a  $f_{i^*} B$ .

Continuando con el proceso de superposición al calcular  $d^2(j, j')$  se sustituye los dos sumandos por su suma y se observa que los demás sumandos de  $d^2(j, j')$  no se alteran por la superposición y  $d^2(j, j')$  es invariante.

Para finalizar se puede replicar el proceso de superposición para los puntos-columna, así mismo esta superposición de puntos columna no afecta a la geometría de las filas, ni al resto de distancias  $x^2$ .

### **Ajustes a la nube de perfiles en $\mathbb{R}^p$ y $\mathbb{R}^n$**

Una vez definidas las nubes de puntos se procede al ajuste de los correspondientes subespacios óptimos en ambas nubes (Gutierrez y R., 2005, pp. 1-60).

### **Ajuste, en $\mathbb{R}^p$ , de la nube de perfiles-fila**

Los  $n$  puntos, perfiles-fila, en  $\mathbb{R}^p$  con coordenadas

$$\left\{ \left( \frac{f_{ij}}{f_i} \right); j = 1, \dots, p \right\} ; i = 1, \dots, n$$

se manejarán a través de las coordenadas:

$$\left\{ \frac{f_{ij}}{f_i \cdot \sqrt{f_j}}; j = 1, \dots, p \right\} ; i = 1, \dots, n$$

las cuales se trasladan del origen del sistema de referencia al centro de gravedad de la respectiva nube de puntos en  $R^p$ .

Dicho centro de gravedad viene dado por:

$$\left\{ \sum_{i=1}^n f_i \cdot \left( \frac{f_{ij}}{f_i \cdot \sqrt{f_j}} \right) = \sqrt{f \cdot j} ; j = 1, \dots, p \right\}$$

por tanto, al trasladar el origen al centro de gravedad, las coordenadas de los puntos-perfiles fila, en  $R^p$ , pasan a ser:

$$\left\{ \frac{f_{ij}}{f_i \cdot \sqrt{f \cdot j}} - \sqrt{f \cdot j} ; j = 1, \dots, p \right\} ; i = 1, \dots, n$$

al proyectar esta nube de puntos transformados sobre el subespacio vectorial definido por el vector unitario  $\mu$ , uno de los puntos de ella, digamos el  $i$ -ésimo, con  $i = 1, \dots, n$ , proporcionar una proyección  $\hat{\Psi}_i$  dada así:

$$\hat{\Psi}_i = \sum_{j=1}^p \left( \frac{f_{ij}}{f_i \cdot \sqrt{f \cdot j}} - \sqrt{f \cdot j} \right) u_j$$

en donde  $u_j$  es la  $j$ -ésima componente del vector unitario  $\mu$  en  $R^p$ . Pero el punto  $i$ -ésimo está dotado de una masa  $f_i$ . Por tanto, la suma ponderada por  $f_i$  de todas las proyecciones al cuadrado, valdrá:

$$\sum_{i=1}^n f_i \cdot \Psi_i^2$$

por tanto, el subespacio definido por aquel vector  $u$  tal que se verifica:

$$\text{Max}_{\mu} \sum_{i=1}^n f_i \cdot \Psi_i^2$$

define la primera componente principal según la metodología del Análisis Factorial General. Continuando dicha metodología, se llega a la conclusión de que la matriz de covarianzas a diagonalizar para obtener todas las componentes principales es la de término general:

$$t_{jj'} = \sum_{i=1}^n f_i \cdot \left( \frac{f_{ij}}{f_i \cdot \sqrt{f \cdot j}} - \sqrt{f \cdot j} \right) \left( \frac{f_{ij'}}{f_i \cdot \sqrt{f \cdot j'}} - \sqrt{f \cdot j'} \right)$$

que define la matriz  $p \times p$ , digamos  $T$ .

**Nota 2.** Es fácil ver que esta matriz  $T$  que resuelve el problema puede ponerse como  $T = X'X$ , donde  $X$  es una matriz  $n \times p$ , con término general  $x_{ij}$  dado por:

$$x_{ij} = \frac{f_{ij} - f_i \cdot f_j}{\sqrt{f_i \cdot f_j}}$$

en efecto, se tiene para la proyección del  $i$ -ésimo punto sobre  $\mu$  :

$$\begin{aligned}\widehat{\Psi}_i &= \sum_{j=1}^p \left( \frac{f_{ij}}{f_{i\cdot} \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j}} \right) u_j \\ &= \left( \frac{f_{i1}}{f_{i\cdot} \sqrt{f_{\cdot 1}}} - \sqrt{f_{\cdot 1}}, \dots, \frac{f_{ij}}{f_{i\cdot} \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j}}, \dots, \frac{f_{ip}}{f_{i\cdot} \sqrt{f_{\cdot p}}} - \sqrt{f_{\cdot p}} \right)_{1 \times p} (\mu_1, \dots, \mu_p)_{p \times 1}\end{aligned}$$

podemos considerar globalmente las proyecciones de todos los puntos, en  $i = 1, \dots, n$ , englobándolas en la expresión matricial:

$$\begin{pmatrix} \frac{f_{11}}{f_{1\cdot} \sqrt{f_{\cdot 1}}} - \sqrt{f_{\cdot 1}} & \dots & \frac{f_{1j}}{f_{1\cdot} \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j}} & \dots & \frac{f_{1p}}{f_{1\cdot} \sqrt{f_{\cdot p}}} - \sqrt{f_{\cdot p}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{i1}}{f_{i\cdot} \sqrt{f_{\cdot 1}}} - \sqrt{f_{\cdot 1}} & \dots & \frac{f_{ij}}{f_{i\cdot} \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j}} & \dots & \frac{f_{ip}}{f_{i\cdot} \sqrt{f_{\cdot p}}} - \sqrt{f_{\cdot p}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{n1}}{f_{n\cdot} \sqrt{f_{\cdot 1}}} - \sqrt{f_{\cdot 1}} & \dots & \frac{f_{nj}}{f_{n\cdot} \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j}} & \dots & \frac{f_{np}}{f_{n\cdot} \sqrt{f_{\cdot p}}} - \sqrt{f_{\cdot p}} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_j \\ \vdots \\ \mu_p \end{pmatrix} = H\mu$$

en donde  $H_{n \times p} \mu_{p \times 1}$  es el vector columna n-dimensional que contiene las n proyecciones  $\widehat{\Psi}_i$   $i = 1, \dots, n$  de los n puntos de la nube de puntos. Por lo tanto, si se considera:

$$(H\mu)'(H\mu)$$

se obtiene la suma de los cuadrados de todas las proyecciones  $\widehat{\Psi}_i$ . Para obtener, finalmente, la suma ponderada de las proyecciones al cuadrado que constituye la Inercia de la nube de puntos, tendremos que efectuar la operación (Gutierrez y R., [sin fecha], pp. 1-60):

$$\sum_{i=1}^n f_{i\cdot} \widehat{\Psi}_i^2 = (\sqrt{f_{\cdot i}} \cdot H\mu)' (\sqrt{f_{\cdot i}} \cdot H\mu) = (H^* \mu)' (H^* \mu)$$

en donde la matriz  $H_{n \times p}^*$  tiene por elementos:

$$h_{ij}^* = \left( \frac{f_{ij}}{f_{i\cdot} \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j}} \right) \sqrt{f_{\cdot i}} \quad i = 1, \dots, n ; j = 1, \dots, p.$$

**Nota 3.** Por otra parte, la matriz  $X$  será la  $H^*$ , de tal manera que se considere  $(H^*)'(H^*)$  en virtud de que la inercia está afectada por las masas de los perfiles  $f_{i\cdot}$ .

En efecto el termino  $j, j'$  de la matriz  $T = (H^*)'(H^*)$  que es de dimensión  $p \times p$ , vendrá dada por:

Fila  $j$ -ésima de  $(H^*)' \times$  columna  $j$  – ésima de  $H^* =$

$$\left(\frac{f_{1j}}{f_{1\cdot}\sqrt{f_j}} - \sqrt{f_j}\right)\sqrt{f_1}\left(\frac{f_{1j'}}{f_{1\cdot}\sqrt{f_{j'}}} - \sqrt{f_{j'}}\right)\sqrt{f_1} + \dots + \left(\frac{f_{ij}}{f_{i\cdot}\sqrt{f_j}} - \sqrt{f_j}\right)\sqrt{f_i}\left(\frac{f_{ij'}}{f_{i\cdot}\sqrt{f_{j'}}} - \sqrt{f_{j'}}\right)\sqrt{f_i} + \dots + \left(\frac{f_{nj}}{f_{n\cdot}\sqrt{f_j}} - \sqrt{f_j}\right)\sqrt{f_n}\left(\frac{f_{nj'}}{f_{n\cdot}\sqrt{f_{j'}}} - \sqrt{f_{j'}}\right)\sqrt{f_n}.$$

Esta expresión del término  $(j, j')$  puede ponerse en la forma:

$$\sum_{i=1}^n \sqrt{f_i}\sqrt{f_i}\left(\frac{f_{ij}}{f_{i\cdot}\sqrt{f_j}} - \sqrt{f_j}\right)\left(\frac{f_{ij'}}{f_{i\cdot}\sqrt{f_{j'}}} - \sqrt{f_{j'}}\right) = t_{jj'}$$

calculados por la teoría general los autovalores  $\lambda_\alpha$  de la matriz simétrica  $T$  la proyección de un punto  $i$  sobre el eje factorial  $\alpha$ -ésimo  $v_\alpha$  vendrá dada por:

$$\widehat{\Psi}_{\alpha i} = \sum_{j=1}^n \left(\frac{f_{ij}}{f_{i\cdot}\sqrt{f_j}} - \sqrt{f_j}\right) u_{\alpha j}$$

en donde  $u_\alpha$  es el autovector de  $T$  asociado a  $\lambda_\alpha$ , con  $\alpha = 1, \dots, p$

### **Simplificación en los cálculos de las proyecciones $\widehat{\Psi}_{\alpha i}$ .**

Hay, pues, que calcular los autovalores de la matriz  $T = (H^*)'(H^*) = X'_{pxn}X_{nxp}$ .

Pueden comprobarse los hechos siguientes:

1. El vector  $u_p = (\sqrt{f_{\cdot 1}}, \dots, \sqrt{f_{\cdot j}}, \dots, \sqrt{f_{\cdot p}})'$  es un vector de  $T$ , con respecto al autovalor 0. En efecto, esto quiere decir que  $(T - \lambda I)u = 0$  es cierto para:

$$T_{pxp}u_{p(px1)} = 0_{px1}$$

es decir

$$T = \left(\sqrt{f_{\cdot 1}}, \dots, \sqrt{f_{\cdot j}}, \dots, \sqrt{f_{\cdot p}}\right)' = 0_{px1}$$

lo cual es fácilmente comprobable.

2. El hecho de que  $u_p = (\sqrt{f_{\cdot j}}; j = 1, \dots, p)$  sea autovalor de  $T$  trae consecuencias interesantes que permiten simplificar los cálculos para obtener las proyecciones de los puntos sobre los ejes factoriales. En efecto, nótese en primer lugar que dado cualquiera otro de los auto vectores  $u_\alpha$ , y denotando por  $u_{\alpha j}$  a su  $j$ -ésima componente, se verifica:

$$\sum_{j=1}^p u_{\alpha j} \sqrt{f_{\cdot j}} = 0$$

en segundo lugar, si consideramos la proyección del punto  $i$ -ésimo de la nube sobre el eje  $\alpha$ ,  $\widehat{\Psi}_{\alpha i}$ , esta expresión se puede simplificar. En efecto:

$$\widehat{\Psi}_{\alpha i} = \sum_{j=1}^n \left( \frac{f_{ij}}{f_{i \cdot} \sqrt{f_j}} - \sqrt{f_j} \right) u_{\alpha j} = \sum_{j=1}^n \frac{f_{ij}}{f_{i \cdot} \sqrt{f_j}} u_{\alpha j} - \sum_{j=1}^n \sqrt{f_j} u_{\alpha j}$$

donde:

$$\widehat{\Psi}_{\alpha i} = \sum_{j=1}^n \left( \frac{f_{ij}}{f_{i \cdot} \sqrt{f_j}} \right) u_{\alpha j}.$$

3. Por otra parte, puede comprobarse lo siguiente: si  $u_{\alpha}$  es un autovector de  $T$ , distinto del antes denotado  $u_p$ , lo es también de  $T^* = X^* X^*$ , donde:

$$X^* = (x_{ij}^*) \quad x_{ij}^* = \frac{f_{ij}}{\sqrt{f_i f_j}}$$

y con respecto al mismo autovalor.

Así pues  $X^*$  no es centrada en contraposición de la  $X$  que lo era. Pero es más fácil manejar  $X^*$ , de tal forma que más simple realizar el análisis de correspondencias sobre  $T^* = X^* X^*$  que sobre  $T = X' X$ .

Obsérvese que hay que tener en cuenta, no obstante, que ello es cierto para todos los auto vectores excepto para el denotado  $u_p$ , que lo era de  $T$  respecto del autovalor 0. Este auto vector  $u_p$  lo es también de la nueva matriz,  $T^*$ , pero en cambio lo es respecto del autovalor unidad.

**Nota 4.** Si se define  $x_{ij}^* = \frac{f_{ij}}{\sqrt{f_i f_j}}$  en lugar de  $x_{ij}^* = \frac{f_{ij} - f_i f_j}{\sqrt{f_i f_j}}$  se comprueba que  $T^* = X^{*'} X^*$  cumple con respecto a los autovalores la propiedad dicha.

Primero veamos qué forma tiene  $T^*$ :

$$T^* = \begin{pmatrix} \frac{f_{11}}{\sqrt{f_1 f_1}} & \cdots & \frac{f_{i1}}{\sqrt{f_i f_1}} & \cdots & \frac{f_{n1}}{\sqrt{f_n f_1}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{1j}}{\sqrt{f_1 f_j}} & \cdots & \frac{f_{ij}}{\sqrt{f_i f_j}} & \cdots & \frac{f_{nj}}{\sqrt{f_n f_j}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{1p}}{\sqrt{f_1 f_p}} & \cdots & \frac{f_{ip}}{\sqrt{f_i f_p}} & \cdots & \frac{f_{np}}{\sqrt{f_n f_p}} \end{pmatrix} \begin{pmatrix} \frac{f_{11}}{\sqrt{f_1 f_1}} & \cdots & \frac{f_{1j'}}{\sqrt{f_1 f_{j'}}} & \cdots & \frac{f_{1p}}{\sqrt{f_1 f_p}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{i1}}{\sqrt{f_i f_1}} & \cdots & \frac{f_{ij'}}{\sqrt{f_i f_{j'}}} & \cdots & \frac{f_{ip}}{\sqrt{f_i f_p}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{n1}}{\sqrt{f_n f_1}} & \cdots & \frac{f_{nj'}}{\sqrt{f_n f_{j'}}} & \cdots & \frac{f_{np}}{\sqrt{f_n f_p}} \end{pmatrix}$$



Luego el término genérico  $jj'$  de la matriz  $T^* = X^{*'}X^*$  se expresa así:

$$t_{jj'}^* = \frac{f_{1j}f_{1j'}}{f_1\sqrt{f_j}\sqrt{f_{j'}}} + \dots + \frac{f_{ij}f_{ij'}}{f_i\sqrt{f_j}\sqrt{f_{j'}}} + \dots + \frac{f_{nj}f_{nj'}}{f_n\sqrt{f_j}\sqrt{f_{j'}}$$

es decir:

$$t_{jj'}^* = \sum_{i=1}^n \frac{f_{ij}f_{ij'}}{f_i\sqrt{f_j}\sqrt{f_{j'}}$$

obsérvese que también puede ser expresado así:

$$t_{jj'}^* = \sum_{i=1}^n \frac{f_{ij}f_{ij'}}{f_i^2\sqrt{f_j}\sqrt{f_{j'}}} f_i = \sum_{i=1}^n f_i \left( \frac{f_{ij}}{f_i\sqrt{f_j}} \right) \left( \frac{f_{ij'}}{f_i\sqrt{f_{j'}}$$

por tanto, podemos decir que  $X^*$  es una matriz no centrada (respecto de la  $X$  que si lo era).

En segundo lugar, todos los auto vectores  $u_\alpha$  de la matriz  $T$ , excepto el  $u_p$ , lo son también de  $T^*$  y respecto de los mismos autovalores. En cambio, el  $u_p$  es autovector también de  $T^*$  pero en lugar de ser respecto del autovalor nulo de  $T$  lo es del autovalor unidad de  $T^*$ .

En efecto, supongamos un auto vector  $u_\alpha \neq u_p$  de componentes  $u_{\alpha j}$ . Si  $u_\alpha$  es autovector de  $T$ , respecto del autovalor  $\lambda_\alpha$  y  $u_\alpha$  lo es de  $T^*$  respecto del mismo autovalor, entonces:

$$T u_\alpha = \lambda_\alpha u_\alpha$$

$$T^* u_\alpha = \lambda_\alpha u_\alpha$$

luego se trata de comprobar que  $T u_\alpha = T^* u_\alpha$ . Se tiene:  $T u_\alpha$  es un vector ( $p \times 1$ ), cuya  $j$ -ésima componente se obtendrá multiplicando la  $j$ -ésima fila de  $T$  por el vector  $u_\alpha$ , de componentes  $u_{\alpha j'}$ , o sea:

$$\sum_{j'=1}^p t_{jj'} u_{\alpha j'}$$

análogamente se llegaría a:

$$\sum_{j'=1}^p t_{jj'} u_{\alpha j'} = \sum_{j'=1}^p t_{jj'}^* u_{\alpha j'}$$

en efecto:

$$\begin{aligned}
& \sum_{j'=1}^p \left\{ \sum_{i=1}^n f_i \cdot \left[ \frac{f_{ij}}{f_i \sqrt{f \cdot j}} - \sqrt{f \cdot j} \right] \left[ \frac{f_{ij'}}{f_i \sqrt{f \cdot j'}} - \sqrt{f \cdot j'} \right] \right\} u_{\alpha j'} \\
&= \sum_{j'=1}^p \left\{ \sum_{i=1}^n f_i \cdot \left[ \frac{f_{ij}}{f_i \sqrt{f \cdot j}} - \sqrt{f \cdot j} \right] \left[ \frac{f_{ij'}}{f_i \sqrt{f \cdot j'}} - \sqrt{f \cdot j'} \right] \right\} u_{\alpha j'} - \sum_{i=1}^n f_i \frac{f_{ij}}{f_i \sqrt{f \cdot j}} \sqrt{f \cdot j'} u_{\alpha j'} \\
&\quad - \sum_{i=1}^n f_i \sqrt{f \cdot j} \frac{f_{ij'}}{f_i \sqrt{f \cdot j'}} u_{\alpha j'} + \sum_{i=1}^n f_i \sqrt{f \cdot j} \sqrt{f \cdot j'} u_{\alpha j'}
\end{aligned}$$

y permutando las sumas queda solo el primer término del segundo, con lo que se prueba

$$\sum_{j=1}^p u_{\alpha j} \sqrt{f \cdot j} = 0.$$

Se tiene:

$$\begin{aligned}
1.- \sum_{j'=1}^p \left\{ \sum_{i=1}^n f_i \frac{f_{ij}}{f_i \sqrt{f \cdot j}} \sqrt{f \cdot j'} \right\} u_{\alpha j'} &= \sum_{j'=1}^p \left\{ \sum_{i=1}^n f_i \frac{f_{ij}}{\sqrt{f \cdot j}} \sqrt{f \cdot j'} \right\} u_{\alpha j'} = \\
& \sum_{i=1}^n \frac{f_{ij}}{\sqrt{f \cdot j}} \sqrt{f \cdot j'} \left( \sum_{j'=1}^p \sqrt{f \cdot j'} u_{\alpha j'} \right) = 0 \\
\sum_{j'=1}^p \left\{ f_i \sqrt{f \cdot j} \frac{f_{ij'}}{\sqrt{f \cdot j'}} \right\} u_{\alpha j'} &= \sum_{j'=1}^p \left\{ \sum_{i=1}^n \sqrt{f \cdot j} \frac{f_{ij'}}{\sqrt{f \cdot j'}} \right\} u_{\alpha j'} = \sum_{j'=1}^p \sqrt{f \cdot j} \frac{\sqrt{f \cdot j'}}{f \cdot j'} \left( \sum_{i=1}^n f_{ij'} \right) u_{\alpha j'} = \\
& \sum_{j'=1}^p \sqrt{f \cdot j} \sqrt{f \cdot j'} u_{\alpha j'} = \sqrt{f \cdot j} \sum_{j'=1}^p \sqrt{f \cdot j'} u_{\alpha j'} = 0
\end{aligned}$$

$$2.- \sum_{j'=1}^p \sum_{i=1}^n \left\{ f_i \sqrt{f \cdot j} \sqrt{f \cdot j'} \right\} u_{\alpha j'} = \sum_{i=1}^n f_i \sqrt{f \cdot j} \left( \sum_{j'=1}^p \sqrt{f \cdot j'} u_{\alpha j'} \right) = 0.$$

**Nota 5.** En cuanto al estudio del vector  $u_p = (\sqrt{f \cdot j} ; j = 1, \dots, p)$ , se tiene lo siguiente:

$$\sum_{j'} t_{jj'} \sqrt{f \cdot j'} = \text{elemento } j - \text{ésimo del vector } Tu_p = \lambda_{\alpha} u_p = 0$$

en efecto:

$$\begin{aligned}
& \sum_{j'=1}^p \sum_{i=1}^n f_i \left[ \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j'}} \right] \left[ \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} - \sqrt{f_{\cdot j'}} \right] \sqrt{f_{\cdot j'}} \\
&= \sum_{j'=1}^p \left[ \sum_{i=1}^n f_i \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} \frac{f_{ij'}}{f_i \sqrt{f_{\cdot j'}}} \sqrt{f_{\cdot j'}} - \sum_{i=1}^n f_i \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} \sqrt{f_{\cdot j'}} \sqrt{f_{\cdot j'}} \right. \\
&\quad \left. - \sum_{i=1}^n f_i \sqrt{f_{\cdot j}} \frac{f_{ij'}}{f_i \sqrt{f_{\cdot j'}}} \sqrt{f_{\cdot j'}} + \sum_{i=1}^n f_i \sqrt{f_{\cdot j}} \sqrt{f_{\cdot j'}} \sqrt{f_{\cdot j'}} \right] \\
&= \sum_{j'=1}^p \left[ \sum_{i=1}^n \frac{f_{ij} f_{ij'}}{f_i \sqrt{f_{\cdot j'}}} - \sum_{i=1}^n \frac{f_{ij}}{\sqrt{f_{\cdot j}}} f_{\cdot j'} - \sum_{i=1}^n f_{ij'} \sqrt{f_{\cdot j}} + \sum_{i=1}^n f_i \sqrt{f_{\cdot j}} f_{\cdot j'} \right] \\
&= \sum_{i=1}^n \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} \left( \sum_{j'=1}^p f_{ij'} \right) \\
&\quad - \sum_{i=1}^n \frac{f_{ij}}{\sqrt{f_{\cdot j}}} \left( \sum_{j'=1}^p f_{j'} \right) - \sum_{i=1}^n \sqrt{f_{\cdot j}} \left( \sum_{j'=1}^p f_{ij'} \right) + \sum_{i=1}^n f_i \sqrt{f_{\cdot j}} \left( \sum_{j'=1}^p f_{j'} \right) \\
&= \sum_{i=1}^n \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} f_i - \sum_{i=1}^n \frac{f_{ij}}{\sqrt{f_{\cdot j}}} - \sum_{i=1}^n \sqrt{f_{\cdot j}} f_i + \sum_{i=1}^n f_i \sqrt{f_{\cdot j}} = 0
\end{aligned}$$

así pues, el vector  $u_p$  es autovector de  $T$ , respecto del autovalor 0. En cambio, este autovector  $u_p$ , que lo es también de  $T^*$ , lo es respecto del autovalor 1 de dicha matriz  $T^*$  (Gutierrez y R., 2005, pp. 1-60).

Es decir, se verifica:

$$T^* u_p = u_p \quad ; \quad \sum_{j'=1}^p t_{jj'}^* u_{\alpha j'} = u_{pj}$$

o sea, en virtud de  $\sum_{j'=1}^p \sum_{i=1}^n \frac{f_{ij} f_{ij'}}{f_i \sqrt{f_{\cdot j}} \sqrt{f_{\cdot j'}}} \sqrt{f_{\cdot j'}} = \sqrt{f_{\cdot j}}$ .

En efecto:

$$\sum_{j'=1}^p \sum_{i=1}^n \frac{f_{ij} f_{ij'}}{f_i \sqrt{f_{\cdot j}} \sqrt{f_{\cdot j'}}} \sqrt{f_{\cdot j'}} = \sum_{j'=1}^p \sum_{i=1}^n \frac{f_{ij} f_{ij'}}{f_i \sqrt{f_{\cdot j}}} = \sum_{i=1}^n \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} \left( \sum_{j'=1}^p f_{ij'} \right) = \sum_{i=1}^n \frac{f_{ij}}{f_i \sqrt{f_{\cdot j}}} f_i =$$

$$\sum_{i=1}^n \frac{f_{ij}}{\sqrt{f_{\cdot j}}} = \frac{1}{\sqrt{f_{\cdot j}}} \sum_{i=1}^n f_{ij} = \frac{f_{\cdot j}}{\sqrt{f_{\cdot j}}} = \sqrt{f_{\cdot j}}.$$

### Ajuste a la nube de puntos-columna, en $R^n$ , constituida por los perfiles-columna

Para construir el ajuste en  $R^n$ , basta con efectuar una permutación de los índices generales  $i, j$ . Para el Análisis de Correspondencias Simple, es obvio: estamos aplicando técnicas factoriales (el Análisis Factorial General o el Análisis de Componentes Principales) a una matriz inicial de datos que constituye estadísticamente hablando, una tabla de contingencia, en la que el papel de las filas y columnas es intercambiable, y al contrario que ocurre en otras tablas, como por ejemplo en las tablas de medidas (variables-observaciones) a las que les hemos aplicado el Análisis de Componentes Principales, los papeles de la/columna son intercambiables, es decir; el ajuste en  $R^n$  se puede obtener intercambiando  $i, j$  simplemente.

Ahora consideramos los puntos  $j$  de la nube en  $R^n$ , que tienen por coordenadas:

$$\left\{ \frac{f_{ij}}{f_{\cdot j} \sqrt{f_{\cdot i}}} \mid i = 1, \dots, n; f_{\cdot j} \right\} \mid j = 1, \dots, p$$

en donde  $f_{\cdot j}$  es la masa de cada punto.

El centro de gravedad de los  $p$  puntos anteriores es el punto de coordenadas  $\sqrt{f_{\cdot i}} \mid i = 1, \dots, n$ . A partir de esta nube y centro de gravedad todo el desarrollo es paralelo al realizado en el caso de la nube en  $R^p$ .

En concreto el Análisis Factorial General (o el Análisis de Componentes Principales si se quiere) se aplicará en este caso a una matriz  $S$  (la antes denotada  $T$ ), cuyos autovalores y auto vectores resuelven la cuestión. Esta matriz  $S$  tendrá como elemento genérico:

$$S_{ii'} = \sum_{j'=1}^p f_{i'} \left[ \frac{f_{ij}}{f_j \sqrt{f_{\cdot i}}} - \sqrt{f_{\cdot i}} \right] \left[ \frac{f_{i'j}}{f_j \sqrt{f_{\cdot i'}}} - \sqrt{f_{\cdot j'}} \right] \mid i, i' = 1, \dots, n$$

$S$  es, evidentemente, de dimensión  $n \times n$ , mientras que  $T$  lo era de  $p \times p$ .

Como todo lo del caso anterior en  $R^n$ , es análogo. Por ejemplo, el escribir  $S$  como producto de la forma  $X'X$ , que ahora será en términos de una matriz análoga a la dada por [1.18] pero permutando  $i$  por  $j$ , llamémosla  $X^*$ , y por otro lado, según el Análisis Factorial General y/o Análisis de Componentes Principales considerando  $XX'$  en lugar de  $X'X$ . En definitiva:

$$S_{n \times n} = H'H^* = X_{n \times p}^* X_{p \times n}'.$$

**Nota 7.** Obsérvese que:

$$H_{n \times p}^* = ((h_{ij}^*)); \quad h_{ij}^* = \frac{f_{ij} - f_j f_i}{\sqrt{f_{\cdot j} f_{\cdot i}}} = h_{ij} = x_{ij}$$

por lo que  $H^* = X$ . Basta tener en cuenta la observación de 2 anterior. En definitiva, en lugar de considerar  $T = X'X$ , consideramos  $S = XX'$ .

Igual que en el caso de la nube en  $R^n$ , aquí puede implicarse el cálculo de autovalores y autovectores, porque, como es fácil comprobar, podemos calcularlo sobre una matriz no centrada,  $S^*$  más fácil de manejar, que tiene los mismos autovalores y autovectores que la matriz  $S$  y que es del tipo no centrado (Gutierrez y R., [sin fecha], pp. 1-60):

$$S^* = X^{**} X^{**'}$$

en donde  $X^{**} = (x_{ij}^{**})$ , siendo:

$$x_{ij}^{**} = \frac{f_{ij}}{\sqrt{f_{\cdot i} f_{\cdot j}}}$$

## 2.4 Índice de Shannon

Una de las herramientas ecológicas que permite medir la diversidad de las especies es el índice de Shannon. Este índice está sujeto a una analogía entre la cantidad de información en un mensaje de código con la diversidad de una comunidad.

El cálculo de este índice se realiza a través de la siguiente igualdad:

$$H' = - \sum_{i=1}^S p_i \log_2 p_i$$

$$p_i = \frac{n_i}{N}$$

donde:

$S$ : Número de especies

$n_i$ : Número de individuos de la especie  $i$

$p_i$ : Proporción de individuos de la especie  $i$  respecto al total de individuos

$N$ : Representa el número de todos los individuos de todas las especies.

Una vez obtenido el valor del índice de Shannon este se compara de acuerdo al siguiente supuesto; los valores menores a 2 representan baja diversidad mientras que los valores mayores que 3 simbolizan alta diversidad. Generalmente los valores del índice se encuentran en el intervalo (1,5; 3,5).

El índice de Shannon aumenta a medida que aumenta la riqueza y la uniformidad de la comunidad (Moreno, 2001, pp.1-84).

## CAPÍTULO III

### METODOLOGÍA

#### 3.1 Tipo y Diseño de Investigación

La presente investigación utiliza una metodología exploratoria, retrospectiva no experimental:

- Exploratoria. Ayuda a caracterizar a las variables climáticas y de hábitat destacando que en estos estudios el propósito central no es la identificación de relaciones causales pero su información es útil para la proposición de las mismas.
- Retrospectivo: La fuente de datos utilizada es histórica (2013).
- No experimental: Se fundamenta en la observación de fenómenos tal y como se dan en su contexto natural (Sampieri et al.,1991, pp.57-73).

Los datos fueron proporcionados por el grupo de investigación CDETERRA “Centro de Desarrollo de Tecnología para la Reducción y Racionalización de Agroquímicos” captada de los Altos Mirandinos Venezuela.

Las técnicas estadísticas fueron Análisis de Componentes Principales, Análisis Factorial Mixto, Análisis de Correspondencias y Análisis implicativo en los paquetes estadísticos SPSS y R.

#### 3.2 Unidad de Análisis

Las unidades de análisis fueron las *lepidopteras*.

#### 3.3 Población y Muestra

La investigación no realiza el cálculo de la muestra ya que la información proporcionada por el grupo de investigación CDETERRA ya obtuvo el tamaño muestral a través de una búsqueda directa con red entomológica manual, cuyo resultado fue 1680 *lepidopteras* en la zona de Altos Mirandinos.

### 3.4 Operacionalización de variables

**Tabla 1-3:** Operacionalización de variables.

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE	ESCALA DE MEDIDA	CATEGORIA
Mes	Cada uno de los doce períodos de tiempo, de entre 28 y 31 días, en que se divide el año.	Cualitativa	Nominal	Febrero Marzo Abril Mayo Junio Julio Agosto Septiembre Octubre Noviembre
Especie	Categoría o división establecida teniendo en cuenta determinadas cualidades. Condiciones o criterios de clasificación.	Cualitativa	Nominal	<i>Achlyodes pallida</i> <i>Actinote antea bytia</i> <i>Actinote antea bytia</i> <i>Adelpha irmina irmina</i> <i>Amphiselenis chama</i> <i>Anthanassa drusilla halli</i> (Ver página 20)
Familia	Categoría taxonómica de clasificación de los seres vivos inferior a la de orden y superior a la de género.	Cualitativa Politómica	Nominal	<i>Hesperiidae</i> <i>Lycaenidae</i> <i>Nymphalidae</i> <i>Papilionidae</i> <i>Pieridae</i> <i>Riodinidae</i>
Subfamilia	Subcategoría taxonómica de clasificación de los seres vivos que tiene un rango inferior al de familia.	Cualitativa Politómica	Nominal	<i>Biblidinae</i> <i>Charaxinae</i> <i>Coliadinae</i> <i>Danainae</i> <i>Dismorphinae</i> <i>Eudaminae</i> <i>Heliconiinae</i> <i>Hesperinae</i> <i>Heteropterinae</i> <i>Limnithidinae</i> <i>Nymphalinae</i> <i>Papilioninae</i> <i>Pierinae</i> <i>Plyommatainae</i> <i>Pyrginae</i> <i>Riodininae</i> <i>Satyrinae</i> <i>Theclinae</i>
Tiempo Atmosférico	Se utiliza para designar a la variedad de fenómenos que se suceden en la atmósfera.	Cualitativa Dicotómica	Nominal	Nublado Soleado



Hospedero	Organismo que alberga a otro en su interior o que lo porta sobre él.	Cualitativa Politómica	Nominal	Al vuelo Libando en compositae y sobre compositae Posada en hierba, pasto, piperaceae y en vegetación
Hábitat	Es el ambiente que ocupa una población biológica.	Cualitativa Dicotómica	Nominal	Bosque Cultivo
Temperatura (°C)	Grado o nivel térmico de un cuerpo o de la atmósfera en el interior.	Cuantitativa	Intervalo	
Humedad (%)	Condición o estado de húmedo en el exterior.	Cuantitativa	Razón	
Presión Atmosférica (Hpa)	Es la presión de un fluido medido con referencia al vacío perfecto o al cero absoluto.	Cuantitativa	Razón	
Velocidad Del viento promedio (cm / s)	Corriente de aire que se produce en la atmósfera al variar la presión.	Cuantitativa	Razón	
Ráfaga de Viento (cm / s)	Ráfaga de Viento	Cuantitativa	Razón	
Pluviosidad (Hpa)	Es la diferencia o resta entre la presión atmosférica y la presión manométrica.	Cuantitativa	Razón	

Realizado por: Abigail Chávez

## CAPÍTULO IV

### ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

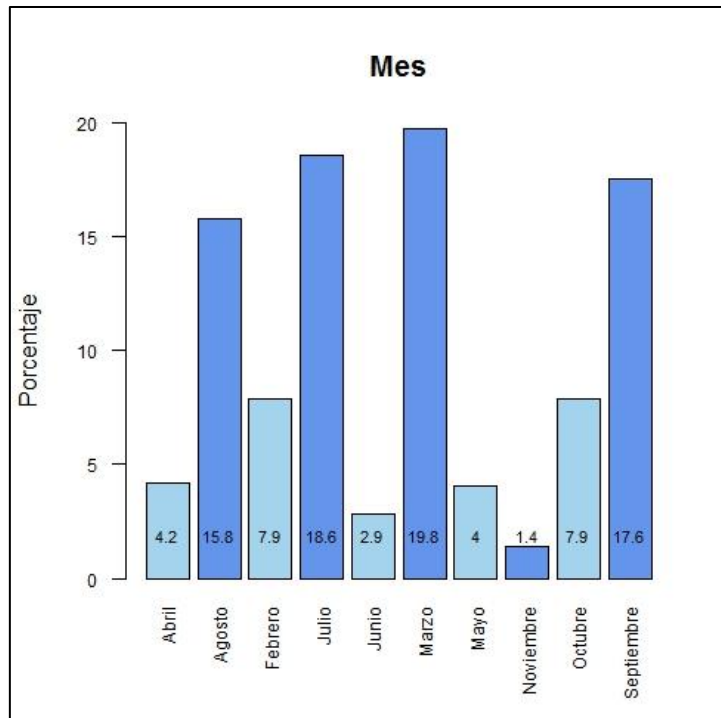
#### 4.1 Análisis exploratorio de datos

Mudable estadística: Mes

**Tabla 1-4:** Distribución estadística de frecuencias “Mes”.

Mudable Mes	Número de <i>lepidopteras</i>	Porcentaje
Abril	70	0,04
Agosto	265	0,16
Febrero	133	0,08
Julio	312	0,19
Junio	48	0,03
Marzo	332	0,20
Mayo	68	0,04
Noviembre	24	0,01
Octubre	133	0,08
Septiembre	295	0,18

Realizado por: Abigail Chávez



**Gráfico 1-4:** Porcentaje de lepidopteras por “Mes”.

Realizado por: Abigail Chávez

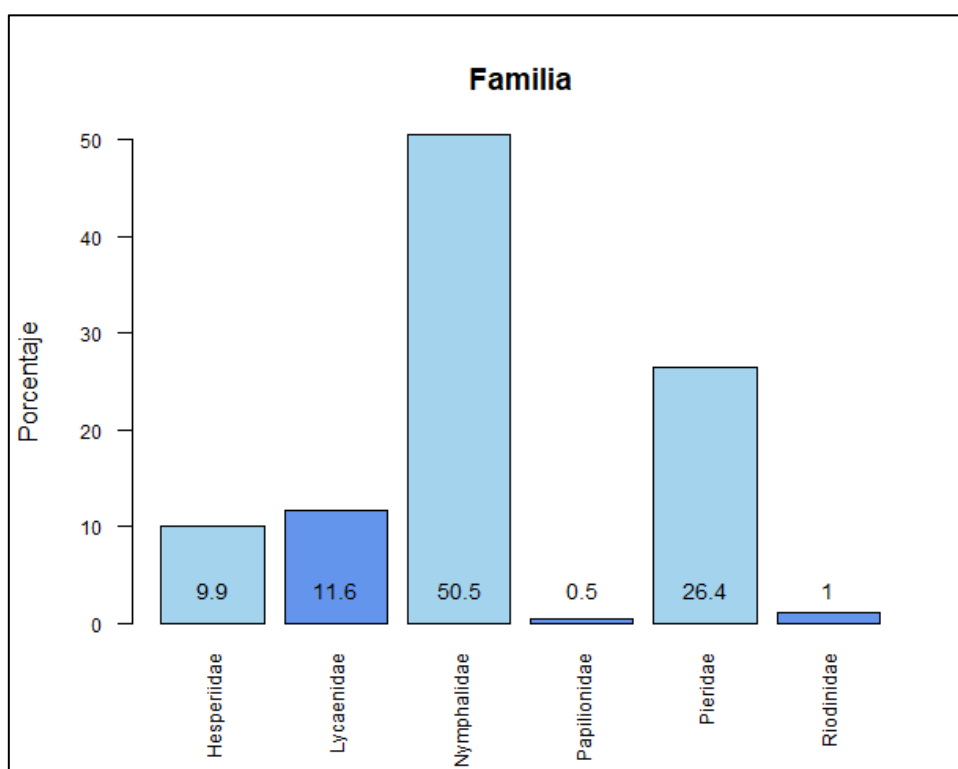
Los meses que asociaron mayor presencia de *lepidopteras* en los Altos Mirandinos fueron marzo (19,8%), julio (18,6%), septiembre (17,6%) y agosto (15,8%), en los demás meses del año la presencia de estos insectos se redujo notablemente.

**Mudable estadística: Familia**

**Tabla 2-4:** Distribución estadística de frecuencias “Familia”.

Mudable Familia	Número_de_ <i>lepidopteras</i>	Porcentaje_F
Hesperiidae	167	0,10
Lycaenidae	195	0,12
Nymphalidae	849	0,51
Papilionidae	8	0,005
Pieridae	444	0,26
Riodinidae	17	0,01

Realizado por: Abigail Chávez



**Gráfico 2-4:** Porcentaje de lepidopteras por “Familia”.

Realizado por: Abigail Chávez

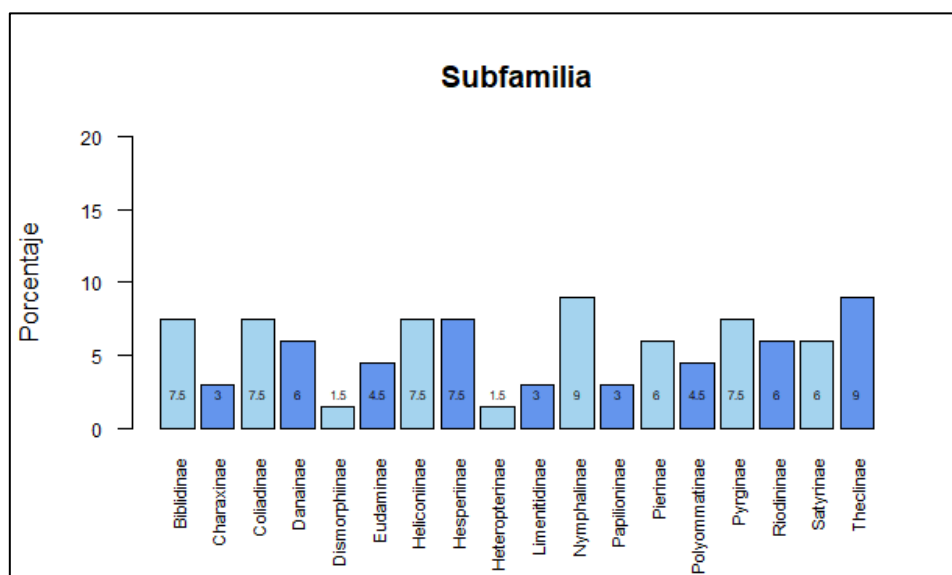
Las familias que conservaron mayor presencia de *lepidopteras* en los Altos Mirandinos fueron la familia *Nymphalidae* (50,5%), *Pieridae* (26,4%), *Lycaenidae* (11,6%) y *Hesperiidae* (9,9%), mientras que en las demás familias el número de insectos se redujo notablemente.

**Mudable estadística: Subfamilia**

**Tabla 3-4:** Distribución estadística de frecuencias “Subfamilia”.

Mudable Subfamilia	Número_de_ <i>lepidopteras</i>	Porcentaje_S
<i>Biblidinae</i>	5	0,075
<i>Charaxinae</i>	2	0,030
<i>Coliadinae</i>	5	0,075
<i>Danainae</i>	4	0,060
<i>Dismorphiinae</i>	1	0,015
<i>Eudaminae</i>	3	0,045
<i>Heliconiinae</i>	5	0,075
<i>Hesperiinae</i>	5	0,075
<i>Heteropterinae</i>	1	0,015
<i>Limnitiidinae</i>	2	0,030
<i>Nymphalinae</i>	6	0,090
<i>Papilioninae</i>	2	0,030
<i>Pierinae</i>	4	0,060
<i>Polyommatainae</i>	3	0,045
<i>Pyrginae</i>	5	0,075
<i>Riodininae</i>	4	0,060
<i>Satyrinae</i>	4	0,060
<i>Theclinae</i>	6	0,090

Realizado por: Abigail Chávez



**Gráfico 3-4:** Porcentaje de lepidopteras por “Subfamilia”.

Realizado por: Abigail Chávez

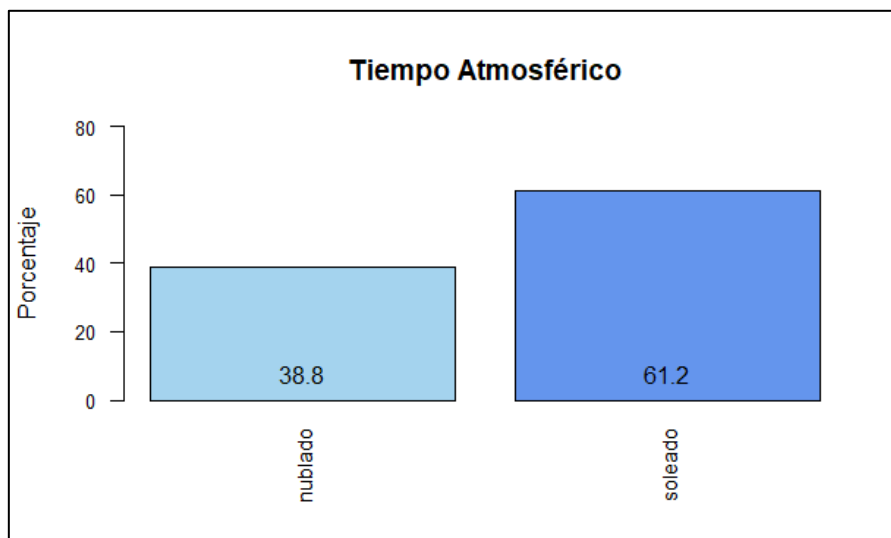
Las subfamilias que presentaron mayor presencia de *lepidopteras* en los Altos Mirandinos fueron *Nymphalidae* (6%), *Theclinae* (6%), *Biblidinae* (5%), *Pyrginae* (5%), *Danainae* y *Pierinae* (4%) mientras que en las demás subfamilias, el número de insectos se redujo notablemente.

**Mudable estadística:** Tiempo atmosférico

**Tabla 4-4:** Distribución estadística de frecuencias “Tiempo atmosférico”.

Mudable Tiempo atmosférico	Número_de_ <i>lepidopteras</i>	Porcentaje_T
nublado	26	38,8
soleado	41	61,2

Realizado por: Abigail Chávez



**Gráfico 4-4:** Porcentaje de lepidopteras por “Tiempo atmosférico”.

Realizado por: Abigail Chávez

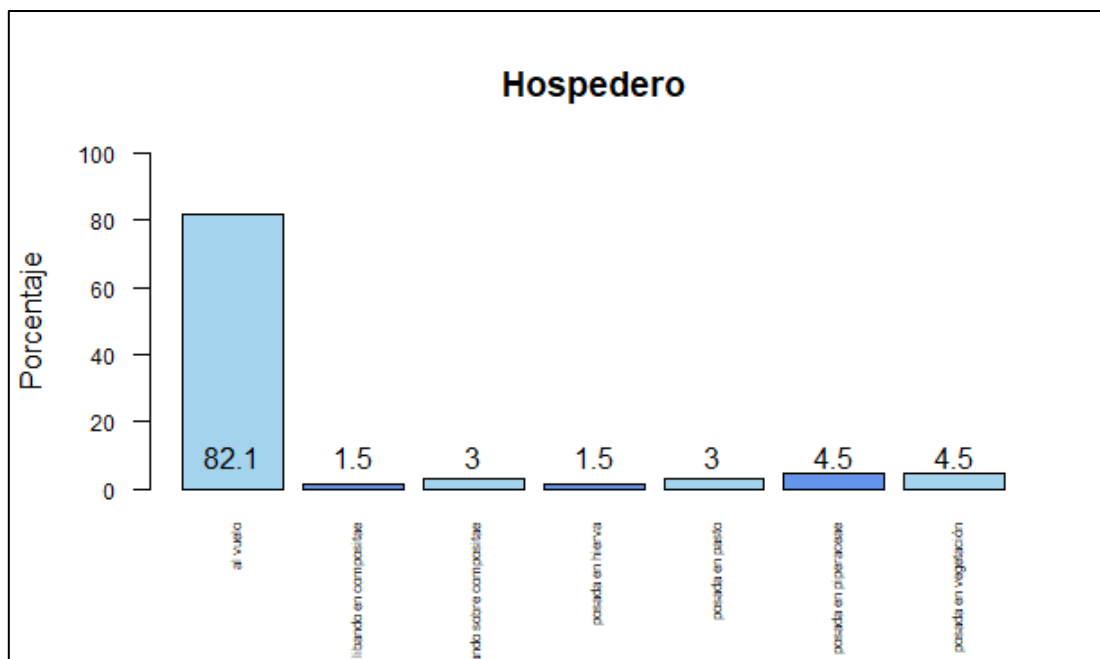
El tiempo atmosférico con mayor presencia de *lepidopteras* en los Altos Mirandinos fue el tiempo soleado (61,2%) seguido del tiempo atmosférico nublado (38,8%).

**Mudable estadística:** Hospedero

**Tabla 5-4:** Distribución estadística de frecuencias “Hospedero”.

Mudable Hospedero	Número_de_ <i>lepidopteras</i>	Porcentaje_H
al vuelo	55	0,821
libando en compositae	1	0,015
libando sobre compositae	2	0,030
posada en hierba	1	0,015
posada en pasto	2	0,030
posada en piperaceae	3	0,045
posada en vegetación	3	0,045

Realizado por: Abigail Chávez



**Gráfico 5-4:** Porcentaje de lepidópteros por “Hospedero”.

Realizado por: Abigail Chávez

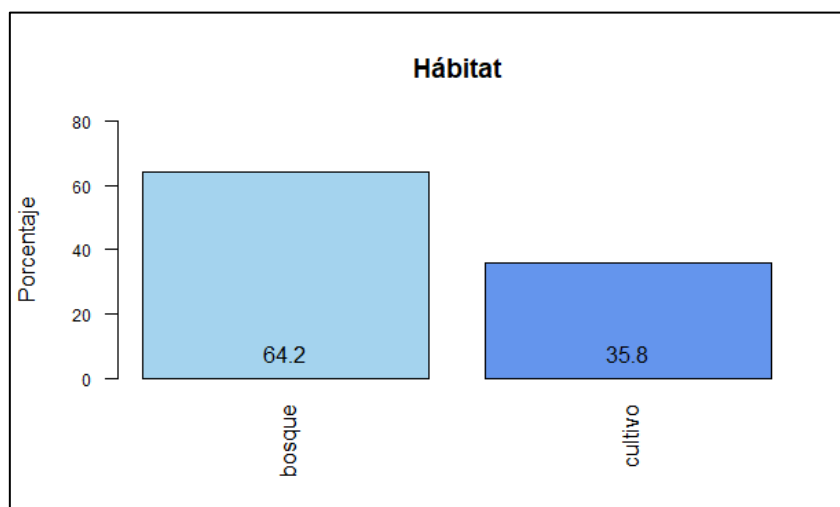
El hospedero que conservó mayor presencia de *lepidopteras* en los Altos Mirandinos fue al vuelo (82,1%) mientras que las demás categorías de hospederos se redujeron notablemente.

**Mudable estadística: Hábitat**

**Tabla 6-4:** Distribución estadística de frecuencias “Hábitat”.

Mudable Hábitat	Número_de_ <i>lepidopteras</i>	Porcentaje_TH
bosque	43	0,642
cultivo	24	0,358

Realizado por: Abigail Chávez



**Gráfico 6-4:** Porcentaje de lepidópteros por “Hábitat”.

Realizado por: Abigail Chávez

El hábitat que presentó mayor presencia de *lepidopteras* en los Altos Mirandinos fue el hábitat bosque (64,2%) mientras que en la categoría cultivo se redujeron notablemente.

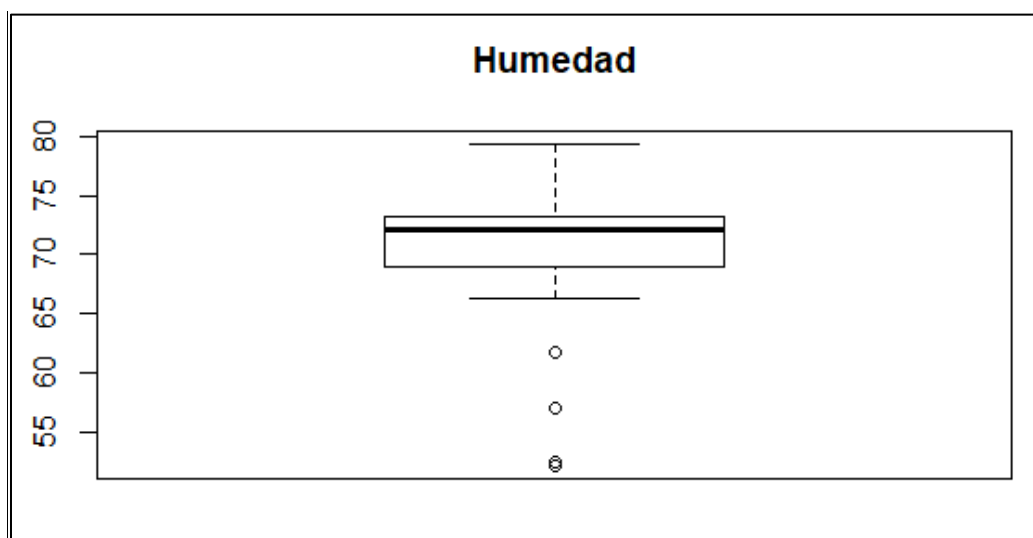
**Variable estadística: Humedad**

**Tabla 7-4:** Medidas de tendencia central y dispersión “Humedad”.

Media	70,60
Mediana	72,11
Desviación	4,69
Max	79,35
Min	52,15
Rango	27,21
CV	6,64

Realizado por: Abigail Chávez

La humedad promedio de la zona de Altos Mirandinos a lo largo del 2013 fue de 70,60 % con una desviación estándar de 4,69% y un rango de fluctuación entre 79,35% y 52,15%, el coeficiente de variación indicó que los datos de humedad eran heterogéneos.



**Gráfico 7-4:** Variable “Humedad”.

Realizado por: Abigail Chávez

La variable humedad presentó una distribución asimétrica positiva con la presencia de outliers.

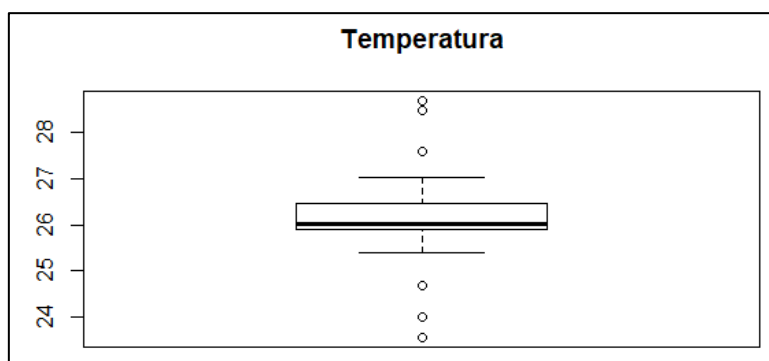
**Variable estadística:** Temperatura

**Tabla 8-4:** Medidas de tendencia central y dispersión “Temperatura”.

Media	26,13
Mediana	26,02
Desviación	0,73
Max	28,67
Min	23,58
Rango	5,09
CV	2,79

Realizado por: Abigail Chávez

En promedio la variable temperatura tomó el valor de 26,13 °C con una desviación estándar de 0,73 °C y un rango de fluctuación entre 28,67 °C y 23,58 °C, el coeficiente de variación indicó que la temperatura en la región de los Altos Mirandinos fue heterogénea.



**Gráfico 8-4:** Variable “Temperatura”.

Realizado por: Abigail Chávez

La variable Temperatura presentó una distribución asimétrica negativa, contó con la presencia de seis datos outliers sin embargo el 95% de los datos contenidos en la caja no presentaron mayor variabilidad.

**Variable estadística:** Presión atmosférica

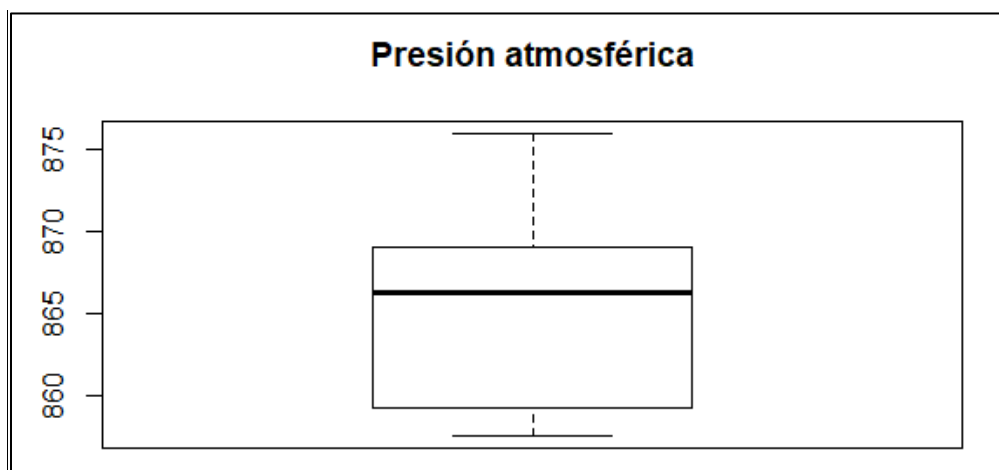
**Tabla 9-4:** Medidas de tendencia central y dispersión “Presión atmosférica”.

Media	865,26
Mediana	866,30
Desviación	4,74
Max	875,92
Min	857,64
Rango	18,29
CV	0,55

Realizado por: Abigail Chávez



En promedio la variable presión atmosférica tomó el valor de 865,26 Hpa con una desviación estándar de 4,74 Hpa y un rango de fluctuación entre 875,92 Hpa y 857,64 Hpa, el coeficiente de variación indicó que las presiones atmosféricas eran homogéneas.



**Gráfico 9-4:** Variable “Presión atmosférica”.  
Realizado por: Abigail Chávez

La variable presión atmosférica presentó una distribución asimétrica positiva, no se observó la presencia de datos outliers, los datos contenidos en la caja no presentaron mayor variabilidad.

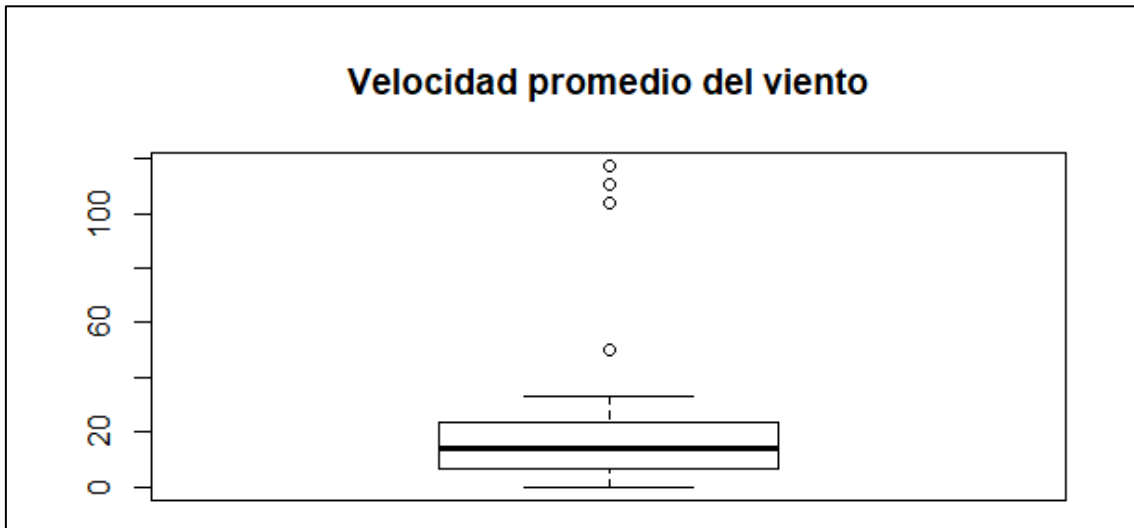
**Variable estadística:** Velocidad del viento promedio

**Tabla 10-4:** Medidas de tendencia central y dispersión “Velocidad del viento promedio”.

Media	18,66
Mediana	14,05
Desviación	22,78
Max	117,36
Min	0,00
Rango	117,36
CV	122,07

Realizado por: Abigail Chávez

La variable Velocidad del viento promedio presentó una media de 18,66 cm/s con una desviación estándar de 22,78 cm/s y un rango de fluctuación de 117,36 cm/s a 0,00 cm/s, el coeficiente de variación indicó que la velocidad del viento promedio fue heterogénea.



**Gráfico 10-4:** Variable “Velocidad del viento promedio”.  
Realizado por: Abigail Chávez

La variable velocidad del viento promedio presentó una distribución asimétrica negativa, contó con la presencia de cuatro datos outliers sin embargo el 95% de los datos contenidos en la caja no presentarán mayor variabilidad.

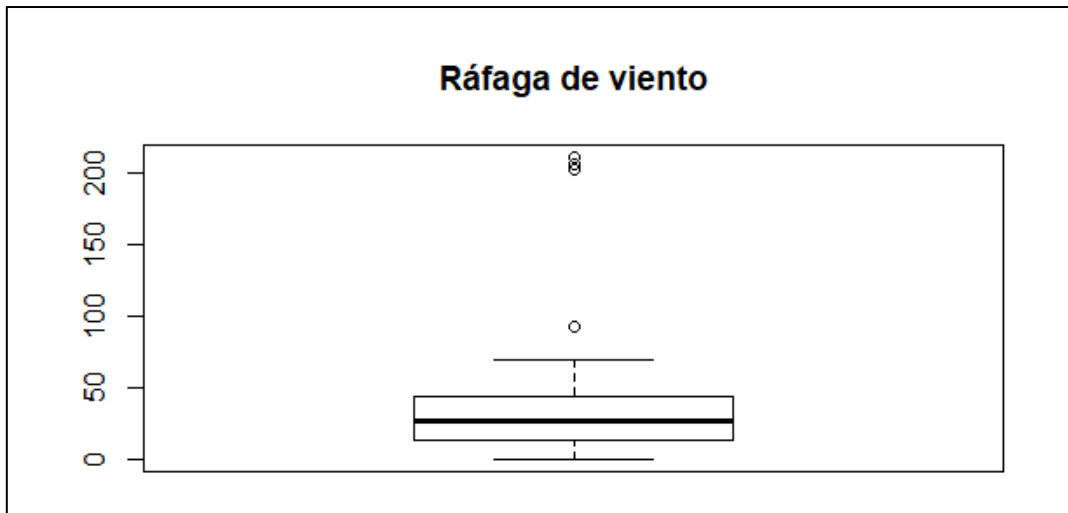
**Variable estadística:** Ráfaga de viento

**Tabla 11-4:** Medidas de tendencia central y dispersión “Ráfaga de viento”.

Media	35,73
Mediana	26,29
Desviación	42,47
Max	211,00
Min	0,00
Rango	211,00
CV	118,87

Realizado por: Abigail Chávez

En promedio la variable ráfaga de viento tomó el valor de 35,73 cm/s con una desviación estándar de 42,47 cm/s y un rango de fluctuación entre 211,00 cm/s y 0,00 cm/s, el coeficiente de variación indicó que los datos de la variable Ráfaga de viento fueron heterogéneos.



**Gráfico 11-4:** Variable “Ráfaga de viento”.  
Realizado por: Abigail Chávez

La variable ráfaga de viento presentó una distribución asimétrica negativa, contó con la presencia de cuatro datos outliers sin embargo el 95% de los datos contenidos en la caja no presentaron mayor variabilidad.

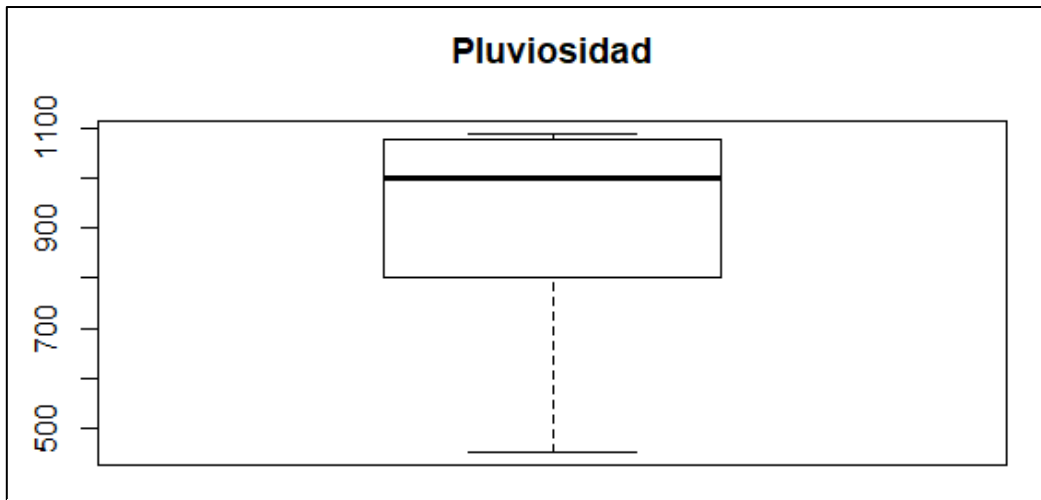
**Variable estadística:** Pluviosidad

**Tabla 12-4:** Medidas de tendencia central y dispersión “Pluviosidad”.

Media	926,77
Mediana	999,98
Desviación	168,79
Max	1088,24
Min	452,92
Rango	635,32
CV	18,21

Realizado por: Abigail Chávez

En promedio la variable pluviosidad tomó el valor de 926,77 Hpa con una desviación estándar de 999,98 Hpa y un rango de fluctuación entre 1088,24 Hpa y 452,92 Hpa, el coeficiente de variación indicó que los datos fueron heterogeneos.

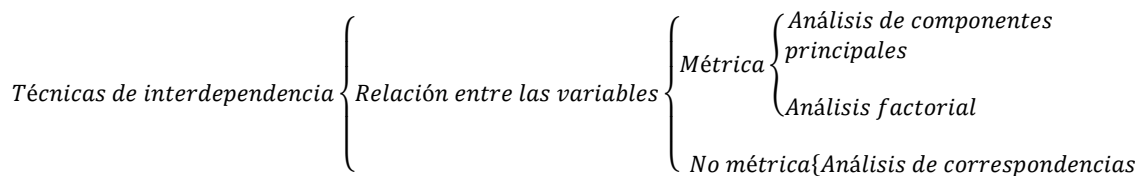


**Gráfico 12-4:** Variable “Pluviosidad”.  
Realizado por: Abigail Chávez

La variable pluviosidad presentó una distribución asimétrica positiva, no se observó la presencia de datos outliers.

#### 4.2 Selección de Técnicas de Interdependencia

Con base al objetivo específico planteado en la investigación y según la naturaleza de las variables las técnicas de interdependencia que ayudan a detectar variables influyentes son (Jiménez y Manzano, 2005, p. 531):



El análisis de componentes principales estudió a las variables estadísticas:  $X_1$ : temperatura,  $X_2$ : humedad,  $X_3$ : presión atmosférica,  $X_4$ : velocidad del viento promedio,  $X_5$ : ráfaga de viento,  $X_6$ : pluviosidad, el análisis de correspondencia estudió a las mudables estadísticas:  $X_7$ : subfamilia,  $X_8$ : tiempo atmosférico,  $X_9$ : hospedero,  $X_{10}$ : hábitat y el análisis factorial mixto estudió a las variables estadísticas y mudables estadísticas.

#### 4.3 Técnicas de Interdependencia

Para ejecutar las técnicas de interdependencia, se analizó si el conjunto de variables estadísticas

siguen una distribución normal multivariante seguido por la detección de observaciones extremas en el dominio de dichas variables y la independencia entre las observaciones.

### Normalidad Multivariante

**H<sub>0</sub>**: El conjunto de variables estadísticas ( $X_1$ : temperatura,  $X_2$ : humedad,  $X_3$ : presión atmosférica,  $X_4$ : velocidad del viento promedio,  $X_5$ : ráfaga de viento,  $X_6$ : pluviosidad) siguen una distribución normal multivariante.

**H<sub>1</sub>**: El conjunto de variables estadísticas ( $X_1$ : temperatura,  $X_2$ : humedad,  $X_3$ : presión atmosférica,  $X_4$ : velocidad del viento promedio,  $X_5$ : ráfaga de viento,  $X_6$ : pluviosidad) no siguen una distribución normal multivariante.

A un nivel de significancia del 5% con un valor  $34319,24 > 11,0705$  se cumple que  $K_p^2 > \chi^2_{\alpha, k-1}$  por lo que se concluye que el conjunto de variables estadísticas ( $X_1$ : temperatura,  $X_2$ : humedad,  $X_3$ : presión atmosférica,  $X_4$ : velocidad del viento promedio,  $X_5$ : ráfaga de viento,  $X_6$ : pluviosidad) no siguieron una distribución normal multivariante.

### Detección de observaciones extremas

$$Z(X_i) = \frac{|x_i - med(x)|}{Meda(x) * 1.483} > 2.5$$

**Tabla 13-4:** Observaciones extremas.

Z(X <sub>1</sub> )	Z(X <sub>2</sub> )	Z(X <sub>3</sub> )	Z(X <sub>4</sub> )	Z(X <sub>5</sub> )	Z(X <sub>6</sub> )
1,22	0,11	1,00	1,71	1,72	1,00
3,55	3,21	0,52	0,21	0,16	2,54
1,15	1,17	1,66	0,75	0,63	0,41
1,61	2,00	3,22	1,71	1,72	0,91
0,16	0,00	1,09	1,71	1,72	0,95
1,43	1,91	3,13	1,71	1,72	0,91
13,67	16,94	2,67	1,59	1,47	6,22
0,54	0,91	0,50	0,29	0,41	0,04
0,12	0,38	0,01	2,20	2,18	0,83
0,10	0,08	1,02	1,71	1,72	1,00
2,60	0,45	0,74	4,39	4,36	0,86
0,75	0,47	2,38	0,18	0,00	1,97
3,02	4,20	0,83	1,00	1,00	2,26
0,76	0,84	2,35	0,94	0,87	2,20

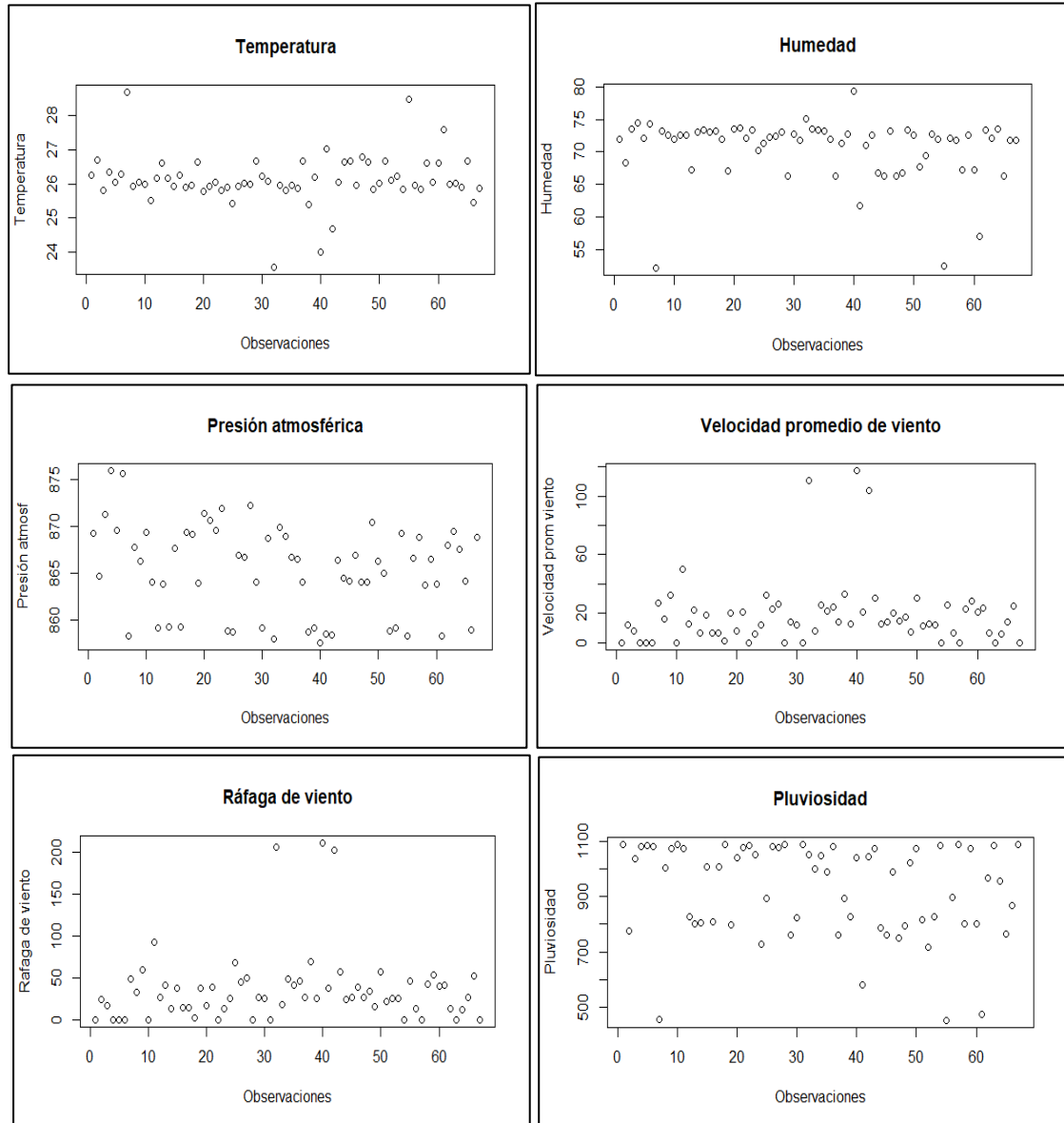
0,43	1,03	0,47	0,59	0,69	0,09
1,23	0,81	2,35	0,88	0,80	2,18
0,57	1,00	1,02	0,88	0,79	0,08
0,31	0,09	0,97	1,57	1,58	1,00
3,11	4,31	0,79	0,75	0,76	2,31
1,22	1,16	1,70	0,75	0,64	0,45
0,48	1,35	1,46	0,82	0,85	0,90
0,08	0,00	1,10	1,71	1,72	0,95
1,09	1,14	1,87	0,99	0,90	0,59
0,57	1,57	2,47	0,24	0,09	3,09
3,08	0,71	2,51	2,26	2,77	1,23
0,52	0,07	0,19	1,12	1,18	0,91
0,01	0,31	0,12	1,46	1,50	0,88
0,20	0,81	1,97	1,71	1,72	1,00
3,36	5,00	0,74	0,03	0,03	2,74
1,11	0,51	2,38	0,25	0,08	1,99
0,32	0,32	0,82	1,71	1,72	0,99
12,60	2,58	2,77	11,74	11,78	0,58
0,27	1,26	1,21	0,70	0,58	0,00
1,11	1,10	0,88	1,40	1,48	0,54
0,26	0,94	0,16	0,94	1,00	0,12
0,85	0,15	0,05	1,24	1,32	0,91
3,36	4,99	0,73	0,03	0,02	2,73
3,24	0,72	2,51	2,28	2,79	1,22
0,87	0,48	2,38	0,20	0,03	1,98
10,41	6,14	2,90	12,57	12,07	0,46
5,10	8,80	2,58	0,79	0,75	4,76
6,80	0,92	2,64	10,92	11,49	0,49
0,13	0,42	0,03	1,98	1,98	0,84
3,23	4,54	0,60	0,13	0,13	2,43
3,34	4,97	0,73	0,02	0,02	2,72
0,27	0,96	0,19	0,76	0,85	0,12
3,95	5,00	0,76	0,05	0,05	2,83
3,18	4,53	0,73	0,43	0,46	2,36
0,98	1,08	1,37	0,82	0,71	0,26
0,00	0,41	0,00	2,03	2,03	0,86
3,36	3,79	0,43	0,34	0,31	2,10
0,37	2,24	2,48	0,17	0,03	3,23
1,00	0,50	2,38	0,23	0,06	1,98
0,89	0,09	1,01	1,71	1,72	0,98
12,64	16,77	2,67	1,41	1,29	6,23
0,38	0,05	0,12	0,91	0,83	1,18
0,87	0,26	0,85	1,71	1,72	1,00
3,01	4,19	0,84	1,06	1,05	2,26
0,11	0,46	0,08	1,75	1,77	0,86

3,04	4,18	0,81	0,87	0,88	2,27
8,04	12,87	2,64	1,15	1,01	5,99
0,11	1,05	0,57	0,93	0,84	0,39
0,07	0,02	1,08	1,71	1,72	0,97
0,58	1,23	0,44	1,00	0,92	0,51
3,29	4,91	0,71	0,00	0,00	2,69
2,86	0,24	2,46	1,30	1,68	1,52
0,85	0,25	0,86	1,71	1,72	1,00

Realizado por: Abigail Chávez

Se identificaron 73 observaciones extremas dentro del dominio de las 6 variables climáticas ( $X_1$ : temperatura,  $X_2$ : humedad,  $X_3$ : presión atmosférica,  $X_4$ : velocidad del viento promedio,  $X_5$ : ráfaga de viento,  $X_6$ : pluviosidad), estas observaciones halladas se atribuyen a los acontecimientos no controlados (climatología variante) por lo que se decidió no apartar a ninguna de estas observaciones ya que se eliminaría la variabilidad de la información.

## Independencia de observaciones

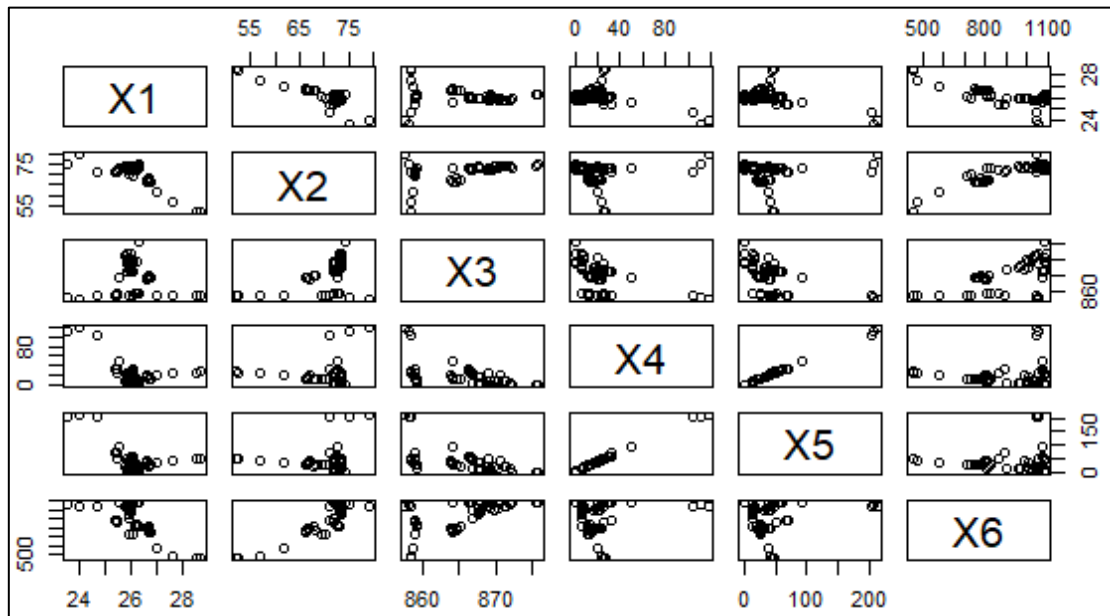


**Gráfico 13-4:** Variables climáticas.

Realizado por: Abigail Chávez

Los gráficos de dispersión asociados a cada una de las variables estadísticas mostraron que las observaciones siguieron un patrón aleatorio.





**Gráfico 14-4:** Relación entre variables climáticas.

Realizado por: Abigail Chávez

El gráfico de dispersión múltiple agrupa los diagramas de dispersión de los diferentes pares de variables en donde se puede apreciar que hay existió una relación aproximadamente lineal entre las variables de clima.

### Análisis de componentes principales

**Tabla 14-4:** Matriz de covarianzas.

Variables	Temperatura	Humedad	Presión atmosférica	Velocidad del viento promedio	Ráfaga de viento	Pluviosidad
Temperatura	0,530	-2,843	-0,263	-8,294	-15,885	-85,278
Humedad	-2,843	21,981	9,007	7,377	15,828	658,910
Presión atmosférica	-0,263	9,007	22,463	-54,623	-103,623	543,557
Velocidad del viento promedio	-8,294	7,377	-54,623	519,042	966,244	59,556
Ráfaga de viento	-15,885	15,828	-103,623	966,244	1803,717	134,680
Pluviosidad	-85,278	658,910	543,557	59,556	134,680	28488,918

Realizado por: Abigail Chávez

## Matriz de correlación

**Tabla 15-4:** Matriz de correlación.

Variables	Temperatura	Humedad	Presión atmosférica	Velocidad del viento promedio	Ráfaga de viento	Pluviosidad
Temperatura	1,000	-0,833	-0,076	-0,500	-0,514	-0,694
Humedad	-0,833	1,000	0,405	0,069	0,079	0,833
Presión atmosférica	-0,076	0,405	1,000	-0,506	-0,515	0,679
Velocidad del viento promedio	-0,500	0,069	-0,506	1,000	0,999	0,015
Ráfaga de viento	-0,514	0,079	-0,515	0,999	1,000	0,019
Pluviosidad	-0,694	0,833	0,679	0,015	0,019	1,000

Realizado por: Abigail Chávez

La matriz de correlaciones permitió observar que 8 de las 15 correlaciones obtenidas fueron superiores a 0,5 en valor absoluto, este hecho mostró que el análisis de componentes principales es adecuado para el conjunto de datos.

## Cálculo y Análisis de las Componentes Principales

Primero se obtuvieron los autovalores de la matriz de correlaciones mismos que coinciden con las varianzas de las diferentes componentes principales con su respectivo y acumulado.

**Tabla 16-4:** Autovalores.

Variables	Autovalores
Temperatura	2,901
Humedad	2,546
Presión atmosférica	0,425
Velocidad del viento promedio	0,068
Ráfaga de viento	0,059
Pluviosidad	0,001

Realizado por: Abigail Chávez

**Tabla 17-4:** Auto vectores.

Variables	Temperatura	Humedad	Presión atmosférica	Velocidad del viento promedio	Ráfaga de viento	Pluviosidad
Temperatura	0,559	0,081	-0,312	0,167	0,745	-0,031
Humedad	-0,517	0,206	0,454	-0,306	0,624	-0,014
Presión atmosférica	-0,161	0,535	-0,644	-0,513	-0,093	-0,021
Velocidad del viento promedio	-0,272	-0,536	-0,345	-0,099	0,169	0,694
Ráfaga de viento	-0,277	-0,537	-0,312	-0,074	0,122	-0,719
Pluviosidad	-0,493	0,298	-0,254	0,775	0,058	0,007

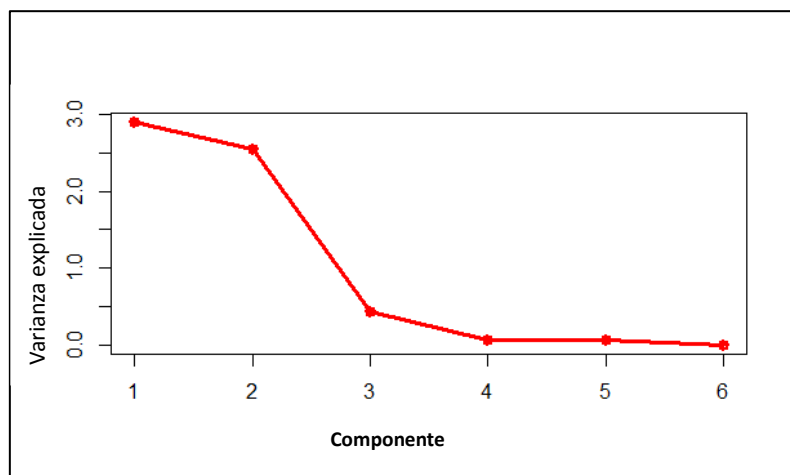
Realizado por: Abigail Chávez

**Tabla 18-4:** Varianza explicada según el orden.

Orden	Varianza explicada	Porcentaje explicado	Porcentaje acumulado
1	2,901	48%	48%
2	2,546	42%	91%
3	0,425	7%	98%
4	0,068	1%	99%
5	0,059	1%	100%
6	0,001	0%	100%

Realizado por: Abigail Chávez

Los valores de la tabla indicaron que la primera y segunda componente explicaron el 48% y 42% respectivamente de la variabilidad total de la información, en conjunto estas componentes representaron el 91% de la variabilidad total mientras que las componentes 3, 4, 5 y 6 representaron el 9% de la información.



**Gráfico 15-4:** Diagrama de sedimentación de la varianza explicada.

Realizado por: Abigail Chávez

Según el criterio de Kaiser se debe tomar como número de componentes principales los autovalores mayores que 1. Con dos componentes se puede explicar el 91% de la varianza total por lo que la elección de este número es idóneo para la formulación de los modelos asociados a cada uno de los componentes; solo contamos con dos puntos claramente elevados.

Se procedió a la construcción de las dos componentes. La siguiente tabla mostró dichos valores por columnas ordenados desde el perteneciente mayor autovalor hasta el menor.

**Tabla 19-4:** Componentes principales.

Variables	Componente 1	Componente 2
Temperatura	0,953	0,129
Humedad	-0,880	0,329
Presión atmosférica	-0,275	0,854
Velocidad del viento promedio	-0,464	-0,855
Ráfaga de viento	-0,472	-0,857
Pluviosidad	-0,840	0,476

Realizado por: Abigail Chávez

Las componentes principales vinieron dadas por las siguientes combinaciones

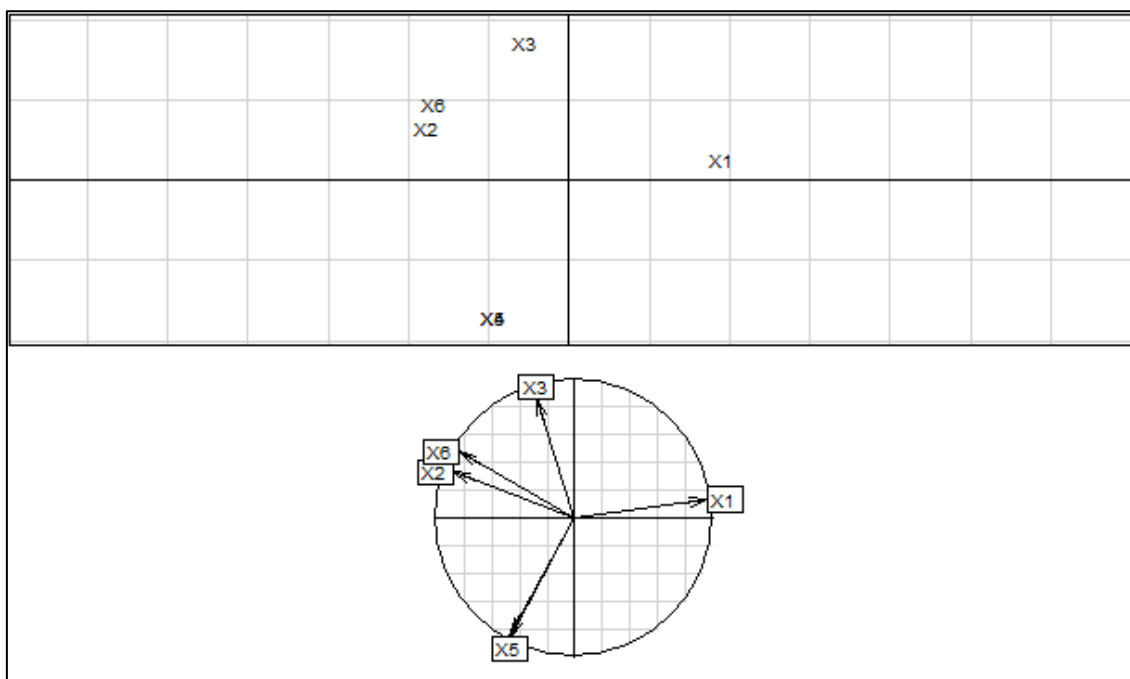
$$\begin{aligned} \text{Componente 1} = & 0,953 \text{ Temperatura} - 0,88 \text{ Humedad} - 0,275 \text{ Presion atmosférica} \\ & - 0,464 \text{ Velocidad promedio del viento} - 0,472 \text{ Ráfaga de viento} \\ & - 0,84 \text{ Pluviosidad} \end{aligned}$$

$$\begin{aligned} \text{Componente 2} = & 0,129 \text{ Temperatura} + 0,329 \text{ Humedad} + 0,854 \text{ Presión atmosférica} \\ & - 0,855 \text{ Velocidad promedio del viento} - 0,857 \text{ Ráfaga de viento} \\ & + 0,476 \text{ Pluviosidad} \end{aligned}$$

La primera componente principal reveló que cuando el clima presentó niveles altos de temperatura la presencia de *lepidoteras* fue alta mientras que cuando las variables climáticas humedad, presión atmosférica, pluviosidad, velocidad de viento promedio y ráfaga de viento presentaban valores mínimos la presencia de *lepidopteras* era menor.

La segunda componente principal mostró que cuando el clima presentó niveles altos de temperatura, humedad, presión atmosférica y pluviosidad la presencia de *lepidoteras* fue alta mientras que cuando las variables climáticas velocidad de viento promedio y ráfaga de viento tomaron valores mínimos la presencia de *lepidopteras* se reducía.

## Relación entre las variables y las componentes principales



**Gráfico 16-4:** Diagrama bidimensional de las componentes principales.

Realizado por: Abigail Chávez

El diagrama bidimensional muestra la posición de las variables respecto a las componentes. La primera componente discriminó de forma clara a la variable temperatura, humedad y pluviosidad; mientras que en la segunda componente se encontraron las variables presión atmosférica y ráfaga de viento.

## Contribución de las variables en cada componente

**Tabla 20-4:** Aportación de las variables por componente.

Variables	Componente 1	Componente 2
Temperatura	31,283%	0,06%
Humedad	26,710%	4,2%
Presión atmosférica	2,604%	28,6%
Velocidad del viento promedio	7,420%	28,7%
Ráfaga de viento	7,671%	28,8%
Pluviosidad	24,311%	8,8%

Realizado por: Abigail Chávez

La tabla 20-4 recogió la contribución de cada variable en la construcción de cada componente principal, para la primera las variables temperatura, humedad y pluviosidad son las que presentaron mayor contribución y para la segunda las variables presión atmosférica, velocidad

del viento promedio y ráfaga del viento han aportado en su mayoría.

**Tabla 21-4:** Varianza explicada por las componentes.

Variables	Componente 1	Componente 2
Temperatura	90,7%	1,65%
Humedad	77,4%	10,7%
Presión atmosférica	7,5%	72,9%
Velocidad del viento promedio	2,1%	73,1%
Ráfaga de viento	2,2%	73,4%
Pluviosidad	70,53%	22,6%

Realizado por: Abigail Chávez

La tabla 21-4 situó el porcentaje de la varianza explicada de cada variable por cada una de las componentes principales construidas donde más del 90,7% fue explicada por la primera componente, en el caso de la humedad y pluviosidad explicó el 77,4% y 70,53% respectivamente. La segunda componente principal explicó el 73,1% de la varianza de la velocidad del viento promedio seguida de las variables ráfaga del viento 73,4% y presión atmosférica 72,9 %.

**Tabla 22-4:** Varianza acumulada.

Variables	Componente 1	Componente 2	Resto de variables
Temperatura	90,7%	92,4%	7,5%
Humedad	77,4%	88,29%	11,70%
Presión atmosférica	7,5%	80,50%	19,49%
Velocidad del viento promedio	21,53%	94,66%	5,33%
Ráfaga de viento	22,25%	95,88%	4,31%
Pluviosidad	70,53%	93,16%	6,83%

Realizado por: Abigail Chávez

La varianza acumulada explicada por cada componente principal se presentó de la siguiente manera: las variables temperatura y humedad se explicaron de mejor forma por la primera componente, mientras que la ráfaga del viento y pluviosidad se explicaron mejor con la segunda componente.

### **Análisis de correspondencias**

El análisis de correspondencia propuesto para los diferentes pares de mudables estadísticas tuvo como objetivo identificar en que mes, tiempo atmosférico, hospedero y hábitat se alojaron las familias de lepidópteros provenientes de los Altos Mirandinos Venezuela.

## Familia-Mes

**Tabla 23-4:** Tabla de correspondencias “Mudable Familia y Mes”.

Tabla de correspondencias											
Familia	Mes										
	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Margen activo
<i>Hesperiidae</i>	5	28	9	6	1	35	34	37	5	7	167
<i>Lycaenidae</i>	77	32	8	11	5	16	15	21	6	4	195
<i>Nymphalidae</i>	32	165	23	38	22	184	141	152	83	9	849
<i>Papilionidae</i>	2	2	0	0	0	2	0	0	2	0	8
<i>Pieridae</i>	17	98	28	13	20	75	71	81	37	4	444
<i>Riodinidae</i>	0	7	2	0	0	0	4	4	0	0	17
Margen activo	133	332	70	68	48	312	265	295	133	24	1680

Realizado por: Abigail Chávez

En la tabla de contingencia 23-4 el margen activo de las filas mostró que las familias *Nymphalidae* (849) y *Pieridae* (444) predominan ante *Hesperiidae*, *Lycaenidae*, *Papilionidae*, *Riodimidae*. El margen activo de las columnas indicó que los meses con mayor presencia de *lepidopteras* son marzo (332), julio (312), septiembre (295) y agosto (265).

**Tabla 24-4:** Perfiles de fila mudables “Familia y Mes”.

Perfiles de fila											
Familia	Mes										
	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Margen activo
<i>Hesperiidae</i>	,030	,168	,054	,036	,006	,210	,204	,222	,030	,042	1,000
<i>Lycaenidae</i>	,395	,164	,041	,056	,026	,082	,077	,108	,031	,021	1,000
<i>Nymphalidae</i>	,038	,194	,027	,045	,026	,217	,166	,179	,098	,011	1,000
<i>Papilionidae</i>	,250	,250	,000	,000	,000	,250	,000	,000	,250	,000	1,000
<i>Pieridae</i>	,038	,221	,063	,029	,045	,169	,160	,182	,083	,009	1,000
<i>Riodinidae</i>	,000	,412	,118	,000	,000	,000	,235	,235	,000	,000	1,000
Masa	,079	,198	,042	,040	,029	,186	,158	,176	,079	,014	

Realizado por: Abigail Chávez

El mes más influyente fue marzo.

**Tabla 25-4:** Perfiles de columna mudables “Familia y Mes”.

Perfiles de columna											
Familia	Mes										
	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Masa
<i>Hesperiidae</i>	,038	,084	,129	,088	,021	,112	,128	,125	,038	,292	,099
<i>Lycaenidae</i>	,579	,096	,114	,162	,104	,051	,057	,071	,045	,167	,116
<i>Nymphalidae</i>	,241	,497	,329	,559	,458	,590	,532	,515	,624	,375	,505
<i>Papilionidae</i>	,015	,006	,000	,000	,000	,006	,000	,000	,015	,000	,005
<i>Pieridae</i>	,128	,295	,400	,191	,417	,240	,268	,275	,278	,167	,264
<i>Riodinidae</i>	,000	,021	,029	,000	,000	,000	,015	,014	,000	,000	,010
Margen activo	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	

Realizado por: Abigail Chávez

La familia más influyente es *Nymphalidae*.

**Tabla 26-4:** Resumen mudables “Familia y Mes”.

Resumen								
Dimensión	Valor singular	Inercia	Chi cuadrado	Sig.	Proporción de inercia		Valor singular de confianza	
					Contabilizado para	Acumulado	Desviación estándar	Correlación 2
1	,436	,190			,826	,826	,035	-,009
2	,141	,020			,086	,912	,023	
3	,125	,016			,067	,980		
4	,054	,003			,013	,992		
5	,042	,002			,008	1,000		
Total		,230	386,974	,000 <sup>a</sup>	1,000	1,000		

a. 45 grados de libertad

Realizado por: Abigail Chávez

Con un valor p muy cercano a 0.00 se concluyó que las variables familias de *lepidopteras* y mes de aparición eran dependientes, de tal forma la inercia describió la importancia del primer y segundo eje con valores de 0,19 y 0,02 respectivamente, el primer eje aportó el 82,6% y el segundo el 8,6% de la varianza total.



**Tabla 27-4:** Puntos de fila generales mudables “Familia y Mes”.

Puntos de fila generales									
Familia	Masa	Puntuación en dimensión		Inercia	Contribución				
		1	2		Del punto en la inercia de dimensión		De la dimensión en la inercia del punto		
					1	2	1	2	Total
<i>Hesperiidae</i>	,099	,266	-,720	,017	,016	,367	,182	,431	,613
<i>Lycaenidae</i>	,116	-1,810	-,052	,166	,871	,002	1,000	,000	1,000
<i>Nymphalidae</i>	,505	,250	,255	,020	,072	,234	,705	,238	,943
<i>Papilionidae</i>	,005	-,867	1,964	,006	,008	,130	,266	,440	,707
<i>Pieridae</i>	,264	,218	-,164	,014	,029	,050	,405	,073	,478
<i>Riodinidae</i>	,010	,370	-1,734	,009	,003	,216	,070	,494	,564
Total activo	1,000			,230	1,000	1,000			

a. Normalización simétrica

Realizado por: Abigail Chávez

La familia *Lycaenidae* y *Nymphalidae* aportaron 0,166 y 0,02 respectivamente de la varianza explicada por el primer eje. Con respecto al porcentaje de cada modalidad explicado por cada eje, el primer eje explicó el 100% de la familia *Lycaenidae* y el 70,5% *Nymphalidae*; además se observó que el 49,4% de la variabilidad de *Rionidae* fue explicada por el segundo eje frente a tan solo el 7% del primer eje.

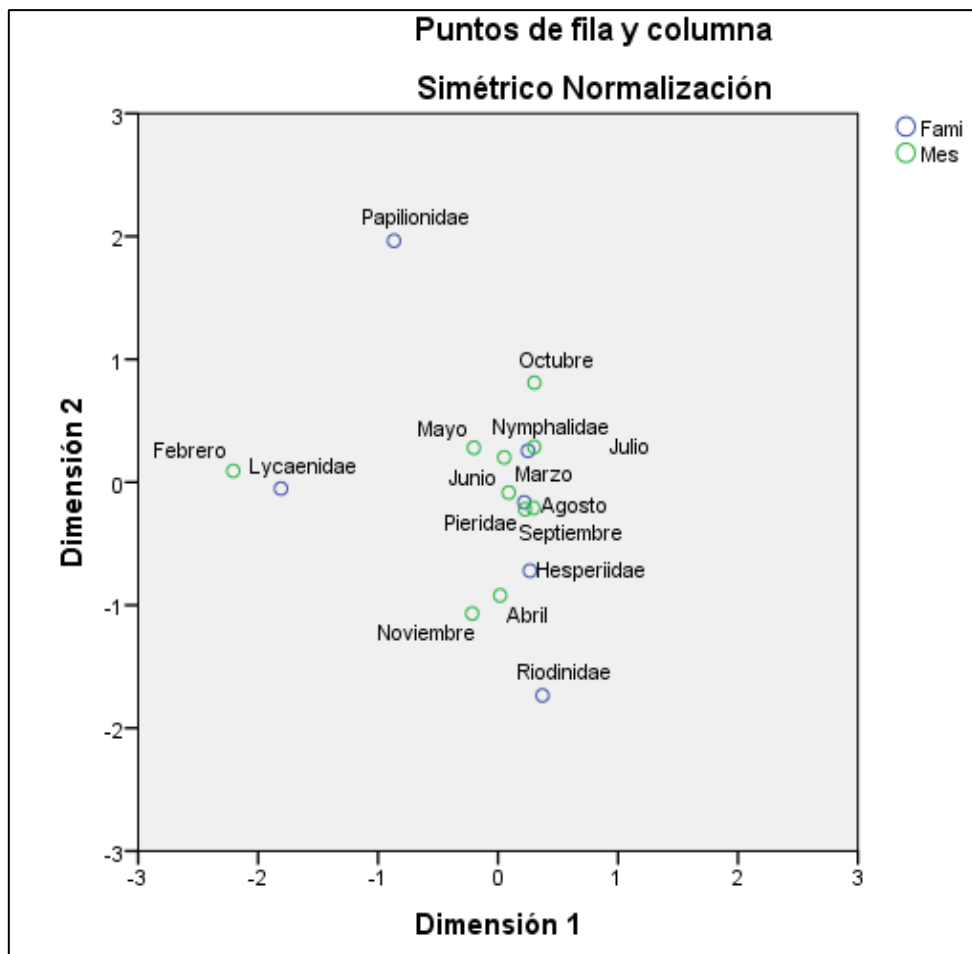
**Tabla 28-4:** Puntos de columnas generales.

Puntos de columna generales <sup>a</sup>									
Mes	Masa	Puntuación en dimensión		Inercia	Contribución				
		1	2		Del punto en la inercia de dimensión		De la dimensión en la inercia del punto		
					1	2	1	2	Total
Febrero	,079	-2,207	,092	,168	,884	,005	,999	,001	1,000
Marzo	,198	,090	-,084	,004	,004	,010	,163	,046	,210
Abril	,042	,017	-,921	,007	,000	,251	,001	,669	,669
Mayo	,040	-,201	,281	,002	,004	,023	,295	,185	,480
Junio	,029	,052	,202	,005	,000	,008	,007	,034	,041
Julio	,186	,301	,287	,012	,039	,109	,611	,179	,790
Agosto	,158	,295	-,210	,007	,031	,049	,799	,130	,929
Septiembre	,176	,225	-,220	,005	,020	,060	,721	,221	,942
Octubre	,079	,303	,810	,011	,017	,369	,280	,647	,927

Noviembre	,014	-,215	-1,068	,007	,002	,116	,042	,335	,377
Total activo	1,000			,230	1,000	1,000			
a. Normalización simétrica									

Realizado por: Abigail Chávez

El mes con mayor presencia de *lepidopteras* fue febrero, este aportó 0,168 de la varianza explicada por el primer eje. Con respecto al porcentaje de cada modalidad explicado por cada eje, el primero explicó el 99,9% de febrero y el 79,9% de agosto; mientras que el segundo explicó el 66,9% de abril y el 64,7% de octubre.



**Gráfico 17-4:** Mudables familia y mes.

Realizado por: Abigail Chávez

Dentro del gráfico de dispersión se observaron marcados grupos correlacionados de categorías, el primer grupo indicó que las familias *Nymphalidae*, *Peridae* *Esperidae* y *Riodinidae* aparecieron en su mayoría en los meses de marzo a noviembre, el segundo grupo mostró que *Lycaenidae* apareció mayoritariamente en febrero, finalmente *Papilionidae* en cualquier mes del año.

## Subfamilia-Hospedero

**Tabla 29-4:** Matriz de correspondencias mudables “Subfamilia y Hospedero”.

Tabla de correspondencias								
Subfamilia	Hospedero							
	al vuelo	libando en compositae	libando sobre compositae	posada en hierva	posada en pasto	posada en piperaceae	posada en vegetación	Margen activo
<i>Biblidinae</i>	4	0	0	0	0	0	1	5
<i>Charaxinae</i>	2	0	0	0	0	0	0	2
<i>Coliadinae</i>	4	0	0	0	0	1	0	5
<i>Danainae</i>	2	0	2	0	0	0	0	4
<i>Dismorphiinae</i>	1	0	0	0	0	0	0	1
<i>Eudaminae</i>	3	0	0	0	0	0	0	3
<i>Heliconiinae</i>	4	0	0	0	1	0	0	5
<i>Hesperiinae</i>	4	0	0	0	1	0	0	5
<i>Heteropterinae</i>	1	0	0	0	0	0	0	1
<i>Limenitidinae</i>	2	0	0	0	0	0	0	2
<i>Nymphalinae</i>	4	0	0	1	0	1	0	6
<i>Papilioninae</i>	2	0	0	0	0	0	0	2
<i>Pierinae</i>	4	0	0	0	0	0	0	4
<i>Polyommatinae</i>	3	0	0	0	0	0	0	3
<i>Pyrginae</i>	4	0	0	0	0	1	0	5
<i>Riodininae</i>	3	0	0	0	0	0	1	4
<i>Satyrinae</i>	4	0	0	0	0	0	0	4
<i>Theclinae</i>	4	1	0	0	0	0	1	6
Margen activo	55	1	2	1	2	3	3	67

Realizado por: Abigail Chávez

La distribución conjunta de las variables subfamilia y hospedero presentó observaciones menores a 5 por lo que se sugirió agrupar con alguna clase adyacente con el fin de corregir la variabilidad de los datos, motivo por el cual la variable subfamilia se agrupó según la familia de lepidópteros (*Nymphalidae*, *Pieridae*, *Hesperiidae*, *Lycaenidae*, *Papilionidae*, *Riodimidae*) y la variable hospedero fusionó sus seis categorías en tan solo tres modalidades (al vuelo, posada en vegetación, libando en compositae).

**Tabla 30-4:** Matriz de correspondencias mudables “Familia y Hospedero”.

Tabla de correspondencias				
Familia	Hospedero			
	al vuelo	libando en compositae	posada en vegetación	Margen activo
<i>Hesperiidae</i>	22	2	4	28
<i>Lycaenidae</i>	12	0	2	14
<i>Nymphalidae</i>	7	1	1	9
<i>Papilionidae</i>	2	0	0	2
<i>Pieridae</i>	9	0	1	10
<i>Riodinidae</i>	3	0	1	4
Margen activo	55	3	9	67

Realizado por: Abigail Chávez

En la tabla de contingencia 30-4 el margen activo de las filas mostró que las familias *Hesperiidae* (28) y *Lycaenidae* (14) predominaron ante *Nymphalidae*, *Papilionidae*, *Pieridae* y *Riodinidae*. El margen activo de las columnas indicó que el hospedero con mayor presencia de lepidópteros fue al vuelo (55).

**Tabla 31-4:** Perfiles fila “Familia y Hospedero”.

Perfiles de fila				
Familia	Hospedero			
	al vuelo	libando en compositae	posada en vegetación	Margen activo
<i>Hesperiidae</i>	,786	,071	,143	1,000
<i>Lycaenidae</i>	,857	,000	,143	1,000
<i>Nymphalidae</i>	,778	,111	,111	1,000
<i>Papilionidae</i>	1,000	,000	,000	1,000
<i>Pieridae</i>	,900	,000	,100	1,000
<i>Riodinidae</i>	,750	,000	,250	1,000
Masa	,821	,045	,134	

Realizado por: Abigail Chávez

El hospedero más influyente fue al vuelo

**Tabla 32-4:** Perfiles de columna “Familia y Hospedero”.

Perfiles de columna				
Familia	Hospedero			
	al vuelo	libando en compositae	posada en vegetación	Masa
<i>Hesperiidae</i>	,400	,667	,444	,418
<i>Lycaenidae</i>	,218	,000	,222	,209
<i>Nymphalidae</i>	,127	,333	,111	,134
<i>Papilionidae</i>	,036	,000	,000	,030
<i>Pieridae</i>	,164	,000	,111	,149
<i>Riodinidae</i>	,055	,000	,111	,060
Margen activo	1,000	1,000	1,000	

Realizado por: Abigail Chávez

La familia más influyente fue *Hesperiidae*.

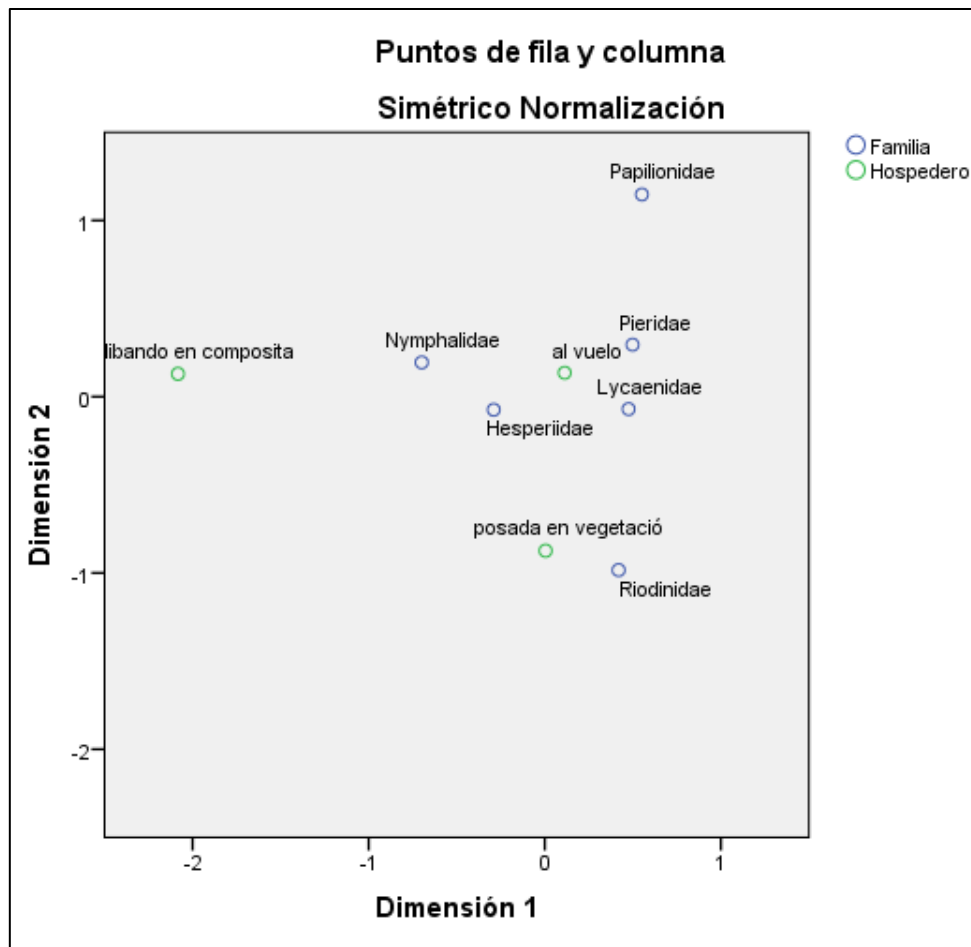
**Tabla 33-4:** Resumen “Familia y Hospedero”.

Resumen								
Dimensión	Valor singular	Inercia	Chi cuadrado	Sig.	Proporción de inercia		Valor singular de confianza	
					Contabilizado para	Acumulado	Desviación estándar	Correlación 2
1	,205	,042			,749	,749	,078	,006
2	,118	,014			,251	1,000	,117	
Total		,056	3,750	,958 <sub>a</sub>	1,000	1,000		

a. 10 grados de libertad

Realizado por: Abigail Chávez

El valor p (0,958) fue mayor que el 5% de significancia, por lo que se concluye que las variables familia de lepidópteras y hospedero eran independientes, de tal forma la inercia describió la importancia del primer y segundo eje con valores de 0,42 y 0,14 respectivamente, el primer eje aportó el 74,9% y el segundo el 25,1% de la varianza total.



**Gráfico 18-4:** Mudables Familia y Hospedero.  
Realizado por: Abigail Chávez

A pesar de que las variables presentaron independencia la gráfica indicó un patrón de agrupación de *lepidóteras* entre el hospedero y la familia, el primer grupo indicó que la familia *Rionidae* se hospedó en la vegetación, el segundo grupo mostró que *Hesperidae*, *Lycaenidae*, *Pieridae* y *Nymphalidae* se las encontró al vuelo finalmente la familia *Papilionidae* apareció en cualquier hospedero.

Finalmente el análisis de correspondencia propuesto para los pares de variables subfamilia-tiempo atmosférico, subfamilia-hábitat, familia-tiempo atmosférico, familia-hábitat, mostraron como resultado la independencia de variables sin presentar ningún patrón de clasificación por lo que se concluye que la única variable que discriminó la presencia de lepidópteros en los Altos Mirandinos de Venezuela fue la mudable estadística mes.

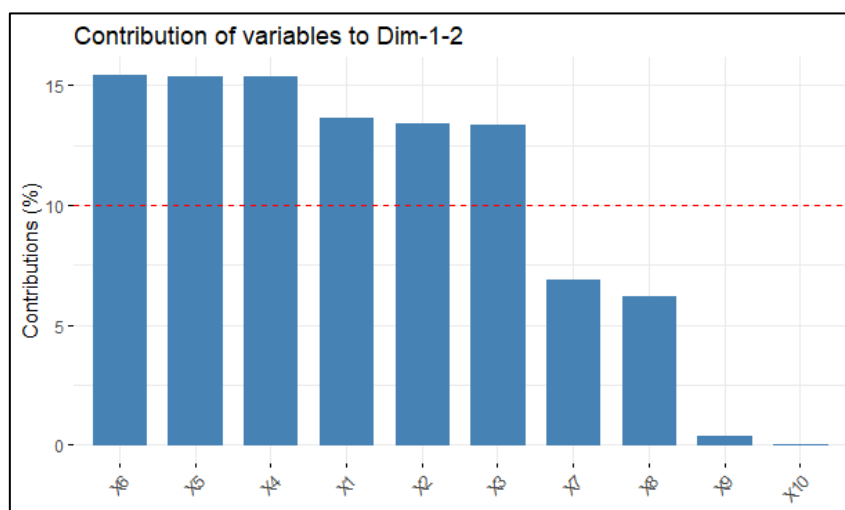
## Análisis Factorial Mixto

**Tabla 34-4:** Varianza explicada según la dimensión.

Dimensiones	Varianza explicada	Porcentaje explicado	Porcentaje acumulado
Dim.1	3,27	21,8	21,8
Dim.2	2,79	18,58	40,37
Dim.3	1,46	9,76	50,13
Dim.4	1,22	8,16	58,29
Dim.5	1,11	7,4	65,69
Dim.6	1,02	6,77	72,46
Dim.7	1	6,66	79,12
Dim.8	0,93	6,18	85,3
Dim.9	0,82	5,47	90,77
Dim.10	0,65	4,33	95,1

Realizado por: Abigail Chávez

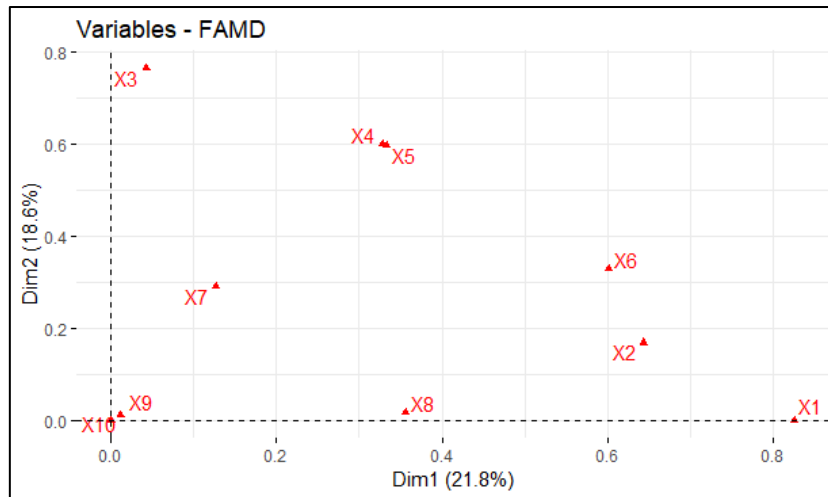
Los valores de la tabla indicaron los porcentajes que explicaron cada una de las 10 dimensiones visibilizando que las dos primeras dimensiones no superaron el 50% de la variabilidad total y se requirió de hasta la dimensión 6 para explicar el 72,46% de la variabilidad total por lo que se concluyó que el análisis factorial mixto no fue una técnica de interdependencia adecuada para medir las correlaciones entre las variables climáticas y de hábitat.



**Gráfico 19-4:** Contribución de las mudables y variables estadísticas.

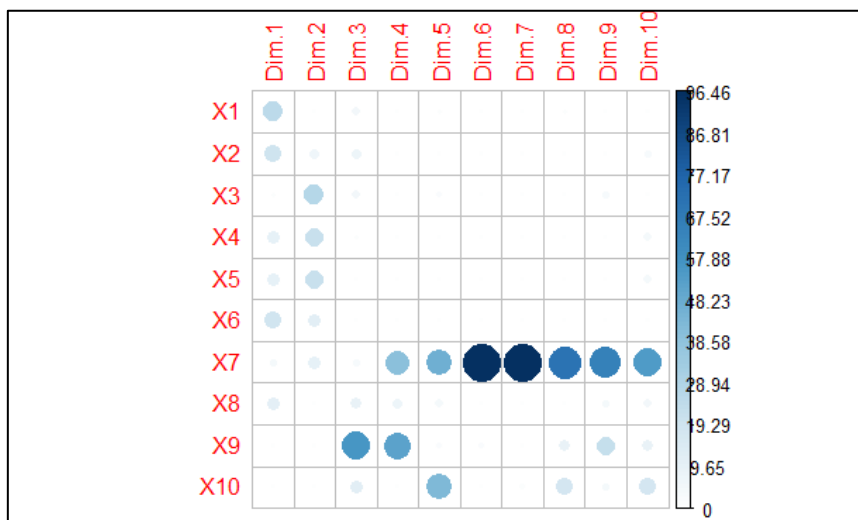
Realizado por: Abigail Chávez

El gráfico 19-4 nos indicó que dentro de la dimensión 1 y 2 las variables influyentes fueron  $X_6$ : pluviosidad,  $X_5$ : ráfaga de viento,  $X_4$ : velocidad del viento promedio,  $X_1$ : temperatura,  $X_2$ : humedad y  $X_3$ : presión atmosférica.



**Gráfico 20-4:** Gráfico de variables.  
Realizado por: Abigail Chávez

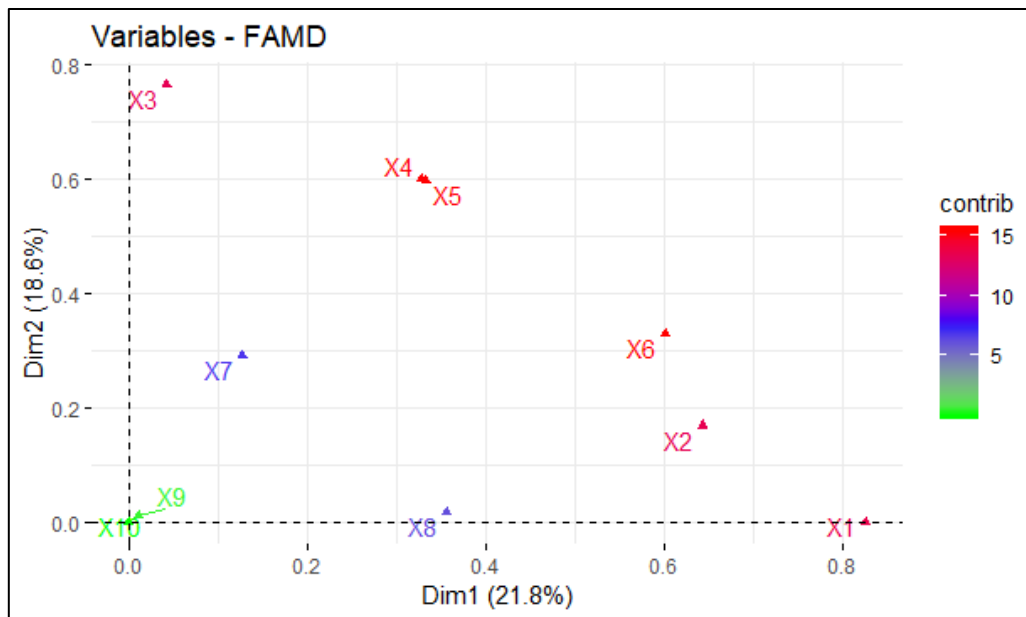
La gráfica 20-4 indicó que la variable que mayor aporte presentó a la dimensión 1 es la variable climática  $X_1$ : Temperatura con un 83%; sin embargo la dispersión de las variables mostraron la falta de correlación entre ellas.



**Gráfico 21-4:** Variables de acuerdo a la tonalidad.  
Realizado por: Abigail Chávez

Para efectos de identificación de factores se consideraron relevantes aquellas variables que asociaron tonalidades intensas de color en cada una de las dimensiones, al buscar construir únicamente dos dimensiones se presenció que las variables influyentes en la dimensión 1 fueron las variables climáticas  $X_1$ : temperatura,  $X_2$ : humedad y  $X_6$ : pluviosidad mientras que en la dimensión 2 se encontraron  $X_3$ : presión atmosférica,  $X_4$ : velocidad del viento promedio y  $X_5$ : ráfaga de viento.





**Gráfico 22-4:** Variables resaltadas.  
Realizado por: Abigail Chávez

La gráfica 22-4 mostró que las variables influyentes en la presencia de *lepidopteras* según sus comunidades fueron  $X_6$ : pluviosidad,  $X_4$ : velocidad del viento promedio y  $X_5$ : ráfaga de viento.

**Tabla 35-4:** Contribuciones de la dimensión 1.

variables	correlation	p.value
X2	0,80	0,00
X6	0,78	0,00
X5	0,58	0,32
X4	0,57	0,40
X1	-0,91	0,00
categoría	estimate	p.value
nublado_X8	1.107.396	9,78E-03
soleado_X8	-1.107.396	9,78E-03

Realizado por: Abigail Chávez

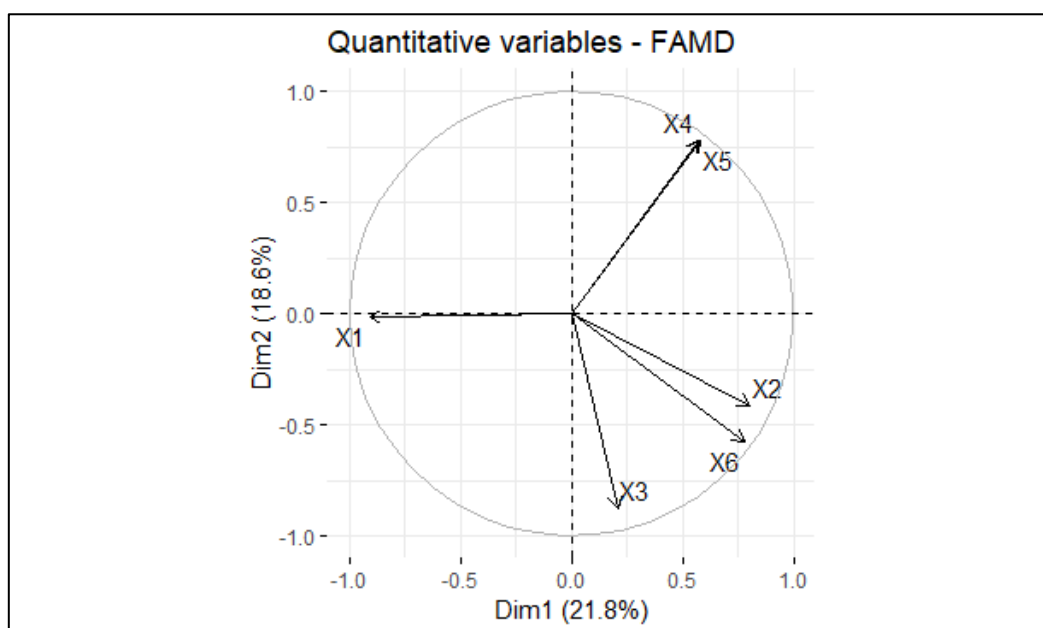
Analizando el conjunto de variables en dos grupos (variables climáticas y de hábitat) asociados a la dimensión 1 se mostró que dentro de las variables estadísticas, las variables climáticas  $X_2$ : humedad,  $X_6$ : pluviosidad,  $X_5$ : ráfaga de viento,  $X_4$ : velocidad del viento promedio y  $X_1$ : temperatura fueron significativas y dentro de las mudables únicamente  $X_8$ : tiempo atmosférico con sus categorías nublado y soleado.

**Tabla 36-4:** Contribuciones de la dimensión 2.

variables	correlation	p.value
X4	0,78	0,00
X5	0,77	0,00
X2	-0,41	542,28
X6	-0,58	0,35
X3	-0,87	0,00
categoría	Estimate	p.value
<i>Papilionidae_X7</i>	4.099.074	1,95E+01

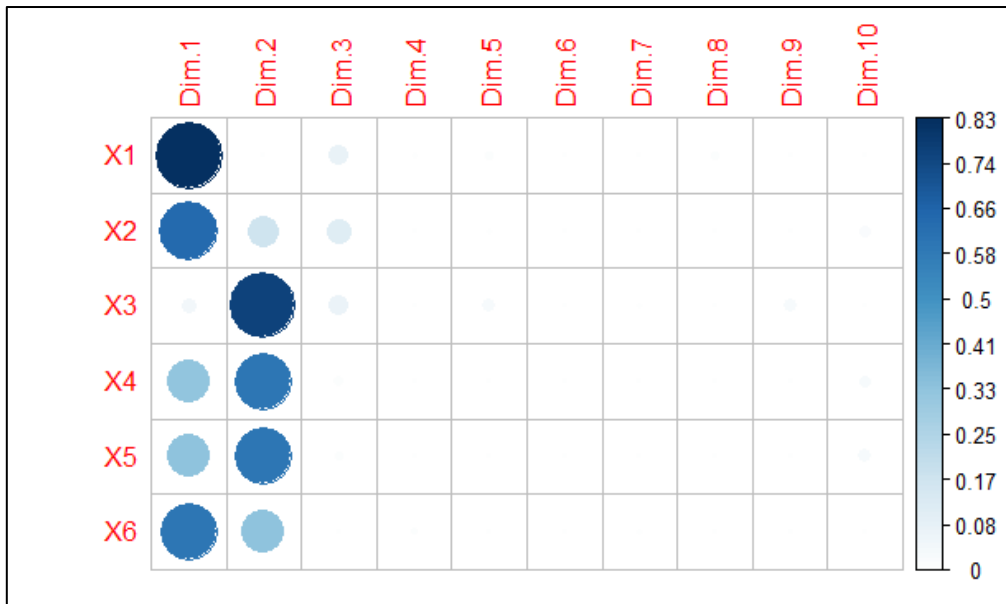
Realizado por: Abigail Chávez

Analizando el conjunto de variables en dos grupos (variables climáticas y de hábitat) asociados a la dimensión 2 se mostró que dentro de las variables estadísticas, las variables climáticas  $X_4$ : Velocidad del viento promedio,  $X_5$ : Ráfaga de viento,  $X_2$ : Humedad,  $X_6$ : Pluviosidad y  $X_3$ : Presión atmosférica fueron significativas y dentro de las mudables únicamente la familia *Papilionidae*.

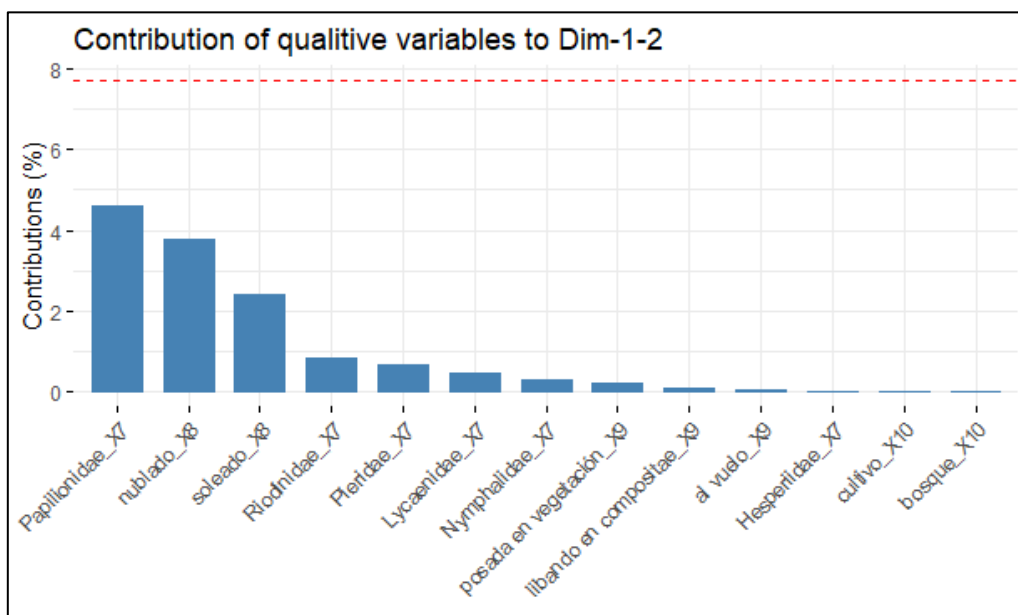


**Gráfico 23-4:** Círculo de correlaciones.

Realizado por: Abigail Chávez

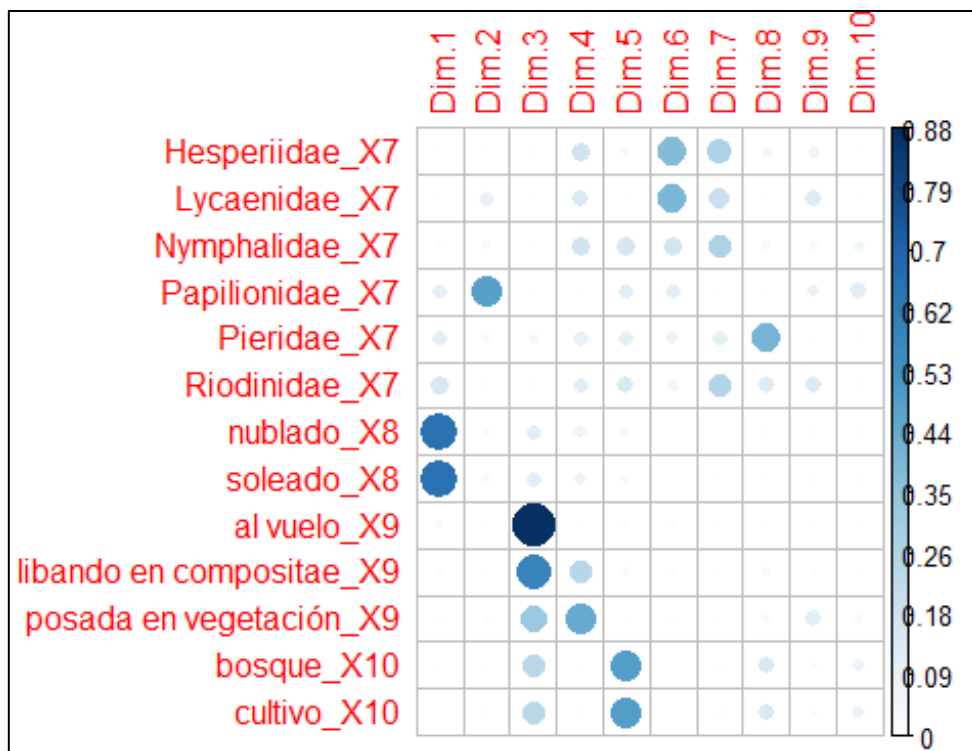


**Gráfico 24-4:** Variables en todas las dimensiones.  
Realizado por: Abigail Chávez



**Gráfico 25-4:** Calidad en el mapa de factores.  
Realizado por: Abigail Chávez

La gráfica radial 25-4 mostró que la primera dimensión condensó tres variables relevantes  $X_1$ : temperatura,  $X_2$ : humedad y  $X_6$ : pluviosidad mismas que conformaron el primer factor denominado elementos meteorológicos, la segunda dimensión en cambio englobó a las variables  $X_3$ : presión atmosférica,  $X_4$ : velocidad del viento promedio y  $X_5$ : ráfaga de viento mismas que constituyeron el segundo factor denominado fenómenos atmosféricos. Es importante hacer notar que los dos factores construidos solo constituyeron parte de la información de variables climáticas excluyendo de los factores de relevancia a las variables de hábitat.



**Gráfico 26-4:** Variables contribuyentes.  
Realizado por: Abigail Chávez

Tomando en cuenta que las variables de hábitat no fueron significativas, es importante rescatar que dentro de la dimensión 1 la única variable influyente fue  $X_8$ : tiempo atmosférico con sus categorías nublado y soleado, y en la dimensión 2 fue más influyente el  $X_9$ : hospedero con sus categorías al vuelo, libando en compositae y posada en vegetación; de entre ellas la categoría más representativa fue al vuelo lo que indicó que hubo mayor presencia de *lepidopteras* cuando el tiempo atmosférico se encontraba nublado o soleado y se las halló al vuelo.

### Comprobación de la hipótesis planteada

#### Planteamiento de hipótesis

$H_0$ : El análisis de componentes principales requieren mayor o igual numero de dimensiones que el análisis factorial, para explicar el mayor porcentaje de variabilidad de las variables estudiandas de las *lepidopteras* de los Altos Mirandinos Venezuela, 2013.

$H_1$ : El análisis de componentes principales requieren menor numero de dimensiones que el análisis factorial, para explicar el mayor porcentaje de variabilidad de las variables estudiandas de las *lepidopteras* de los Altos Mirandinos Venezuela, 2013.

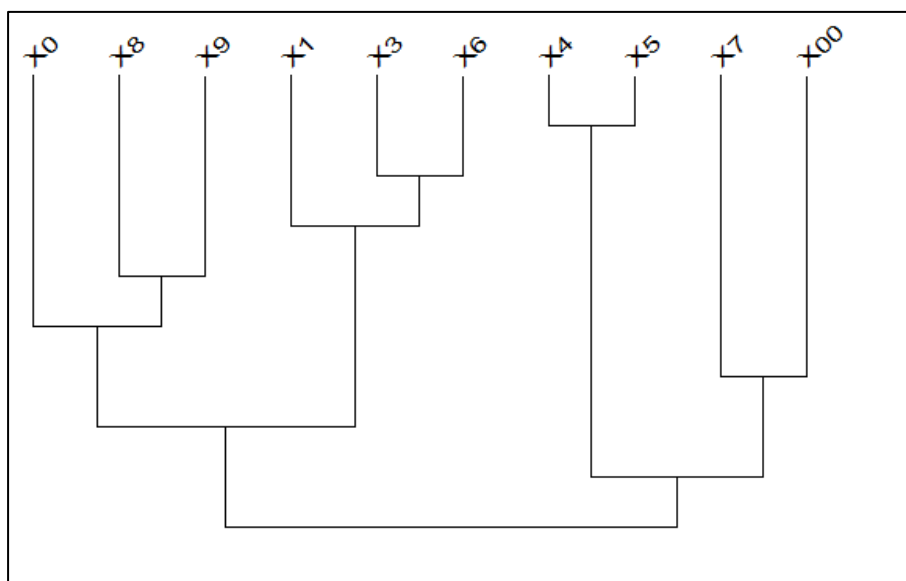
$$H_0: N^{\circ}_{dACP} \geq N^{\circ}_{dAF}$$

$$H_1: N^{\circ}_{dACP} < N^{\circ}_{dAF}$$

El análisis de componentes principales necesitó tan solo dos dimensiones para explicar el 91% de la variabilidad total de los datos (revisar Tabla 18-4), mientras que el análisis factorial necesitó 6 para explicar el 72,46% de la variabilidad (revisar Tabla 34-4); mostrando una diferencia de 4 dimensiones, la misma que no requiere ningún método estadístico para comprobar la existencia de diferencia significativa.

Con lo expuesto se puede concluir que el análisis de componentes principales requieren menor numero de dimensiones que el análisis factorial, para explicar el mayor porcentaje de variabilidad de las variables estudiadas de las *lepidopteras* de los Altos Mirandinos Venezuela, 2013.

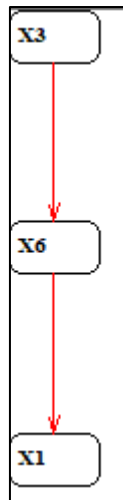
#### 4.4 Análisis Implicativo



**Gráfico 27-4:** Dendrograma de similaridad.

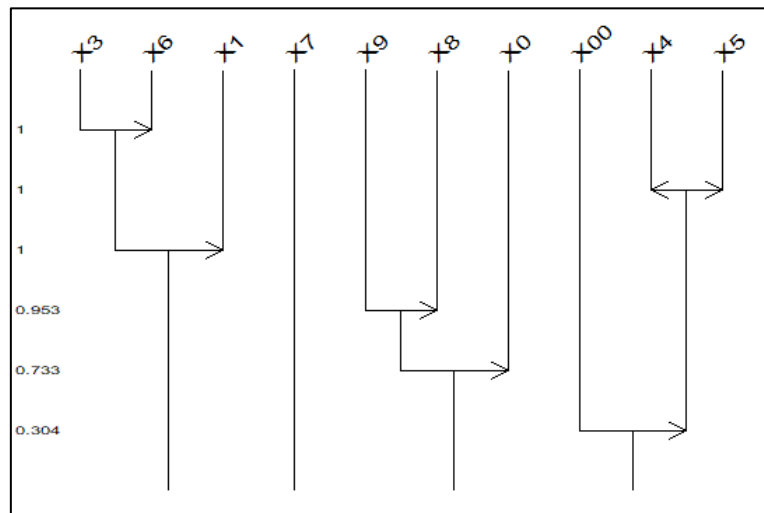
Realizado por: Abigail Chávez

El dendrograma mostró que las variables similares fueron las variables  $X_4$ : velocidad del viento promedio y  $X_5$ : ráfaga de viento seguidas por las variables  $X_3$ : presión atmosférica y  $X_6$ : pluviosidad, además se visualizó la presencia de dos factores, el primer factor agrupó a las variables  $X_2$ : humedad,  $X_8$ : tiempo atmosférico,  $X_9$ : hospedero,  $X_1$ : temperatura,  $X_3$ : presión atmosférica y  $X_6$ : pluviosidad y el segundo factor agrupó a las variables  $X_4$ : velocidad del viento promedio,  $X_5$ : ráfaga de viento,  $X_7$ : subfamilia y  $X_{10}$ : hábitat, la comparación de estos dos factores indica que son completamente diferentes.



**Gráfico 28-4:** Grafo implicativo.  
Realizado por: Abigail Chávez

El grafo implicativo mostró que la presencia de las *lepidopteras* dependió de los niveles de  $X_3$ : presión atmosférica,  $X_1$ : temperatura, y  $X_6$ : pluviosidad, es decir la presencia de estos insectos estuvo condicionada a variables climáticas.



**Gráfico 29-4:** Árbol cohesivo.  
Realizado por: Abigail Chávez

El árbol de cohesión indicó que las variables altamente correlacionadas fueron  $X_3$ : presión atmosférica y  $X_6$ : pluviosidad seguidos por  $X_4$ : velocidad del viento promedio,  $X_5$ : ráfaga de viento.

#### 4.5 Índice de Shannon

**Tabla 37-4:** Cálculo del Índice de Shannon

	Especies	N° de especies	$p_i$	$p_i \cdot \log_2 p_i$
1	<i>Achlyodes pallida</i>	3	0,002	-0,016
2	<i>Actinote anteas bytia</i>	5	0,003	-0,025
3	<i>Actinote hylonome basilia</i>	2	0,001	-0,012
:	:	:	:	:
:	:	:	:	:
50	<i>Eurema salome limoneus</i>	15	0,009	-0,061
51	<i>Eurema xantochlora xantochlora</i>	43	0,026	-0,135
52	<i>Euselasia venezolana</i>	1	0,001	-0,006
	:	:	:	:
	:	:	:	:
125	<i>Xenophanes tryxus</i>	3	0,002	-0,016
126	<i>Ypthimoides renata</i>	4	0,002	-0,021

Realizado por: Abigail Chávez

$$H' = - \sum_{i=1}^{126} p_i \log_2 p_i$$

$$H' = -\{[0,002(-0,016)] + [0,003(-0,025)] + [0,001(-0,012)] + \dots + [0,002(-0,021)]\}$$

$$H' = 5,60$$

como el índice de Shannon es superior a 3 ( $5,60 > 3$ ) se concluye que la biodiversidad de las *lepidopteras* en la zona de Altos Mirandinos Venezuela es alta.

## CONCLUSIONES

El mes con mayor presencia de *lepidopteras* en la zona de los Altos Mirandinos Venezuela fue marzo con un porcentaje de aparición del 19,8% seguido por julio con un 18,6%. Durante el estudio en esta zona se encontraron 6 familias de *lepidopteras* de entre las cuales predominó *Nymphalidae* con un valor de 50,5%; a su vez estas se dividieron en 18 Subfamilias siendo *Nymphalinae* y *Theclinae* las que presentaron mayor presencia de *lepidopteras* con un porcentaje del 6%. En cuanto al Tiempo atmosférico el estado soleado tuvo un 61,2% de afluencia de *lepidopteras* frente a un 38,8% del estado nublado; por otra parte, el 82,1% de las *lepidopteras* consideradas en el estudio se las recolectó al vuelo, adicionalmente el hábitat que asoció mayor cantidad de *lepidopteras* fue el bosque con un porcentaje del 64,2% mientras que la zona de cultivo se mantuvo con 35,8%. En cuanto a las variables climáticas, la humedad promedio de la zona de los Altos Mirandinos fue de 70,6% con una temperatura media de 26,13 °C y una presión atmosférica de 865,26 Hpa mientras que la velocidad del viento promedio fue de 18,66 cm/s, la ráfaga de viento 35,73 cm/s y la pluviosidad de 926,77 Hpa.

A partir del método robusto de detección de valores extremos se identificaron 73 observaciones extremas, estas observaciones halladas se atribuyeron a los acontecimientos no controlados (climatología variante) por lo que se decidió no apartar a ninguna de estas observaciones ya que se eliminaría la variabilidad de la información.

De acuerdo con el criterio de Kaiser y con el apoyo de la gráfica de sedimentación se determinó la construcción de dos componentes principales, con la primera componente principal se determinó que existió mayor presencia de *lepidoteras* cuando la variable temperatura se mantuvo en sus valores máximos y las variables humedad, presión atmosférica, pluviosidad, velocidad y ráfaga de viento en sus valores mínimos, en cambio, para la segunda componente existió mayor presencia de *lepidoteras* cuando los niveles de temperatura, humedad, presión atmosférica y pluviosidad fueron altos y los niveles de las variables velocidad y ráfaga de viento fueron mínimos.

El análisis de correspondencia de la mudable familia-mes identificó grupos correlacionados de categorías, el primero reveló que las familias *Nymphalidae*, *Peridae* *Esperidae* y *Riodinidae* aparecieron en su mayoría en los meses de marzo a noviembre, mientras que *Lycaenidae* apareció mayoritariamente en el mes de febrero y finalmente la familia *Papilionidae* apareció en cualquier mes del año. A pesar de que en el AC para las mudables Familia-Hospedero las variables presentaron independencia, existió un patrón de agrupación de categorías, el primer



grupo indicó que la familia *Rionidae* se alojó en la vegetación, el segundo grupo albergó a las familias *Hesperiidae*, *Lycaenidae*, *Pieridae* y *Nymphalidae* al vuelo y la familia *Papilionidae* apareció en cualquier hospedero.

Por otro lado, el análisis factorial mixto corroboró la información obtenida con el ACP y el AC evidenciando dos factores representativos, el primero consolidado únicamente por variables climáticas denominado factor de elementos meteorológicos el cual abarcó a las variables temperatura, humedad y pluviosidad, el segundo denominado fenómenos atmosféricos agrupó a las variables presión atmosférica, ráfaga y velocidad promedio de viento. Sin embargo, las mudables de hábitat no aportaron mayor información acerca de la concurrencia de estos insectos, no obstante, las mudables tiempo atmosférico, familia (*Papilionidae*) y el hospedero (al vuelo) reflejaron la variabilidad de su presencia.

Al comparar los resultados de cada una de las técnicas de interdependencia se evidenció de manera directa que el análisis de componentes principales explicó el mayor porcentaje (91%) de variabilidad total de los datos con tan solo 2 dimensiones, mientras que el análisis factorial necesitó de hasta 6 dimensiones para explicar el 72,46% de la variabilidad total de la información de las *lepidopteras*.

El análisis implicativo ayudó a contrastar los resultados de técnicas paramétricas y no paramétricas determinando que las variables influyentes seleccionadas por las dos técnicas fueron temperatura y pluviosidad, mientras que la variable climática humedad resultado de las técnicas de interdependencia fue reemplazada por la variable presión atmosférica en el análisis implicativo.

Por último, el cálculo del índice de diversidad de Shannon con un valor superior a 3 reflejó la diversidad de especies que existen en la zona de Altos Mirandinos Venezuela.

## **RECOMENDACIONES**

Socializar los resultados de la investigación con la comunidad ecológica de los Altos Mirandinos de manera que se puedan enriquecer los conocimientos acerca de la dinámica poblacional de este grupo de insectos y de esta forma incentivar la creación de planes de acción que permitan preservar la especie de *lepidopteras*.

Continuar con la presente investigación de manera que se busquen técnicas alternativas como la simulación para perfeccionar las investigaciones en el campo ecológico en Ecuador.

Extender el estudio de técnicas estadísticas multivariadas en la malla curricular, debido a su aplicabilidad en diferentes campos de la ciencias, con el fin de incrementar las líneas de investigación de la carrera de Estadística.

## BIBLIOGRAFÍA

**Báez, M. y Oromi, P.**, Orden Lepidoptera. *Lista de especies silvestres de Canarias. Hongos, Plantas y Animales Terrestres* [en línea], 2010. p. 302-318. Disponible en: [http://www.gobcan.es/cmayerot/medioambiente/medionatural/biodiversidad/especies/bancodatos/Lista\\_Especies\\_Silvestres.pdf](http://www.gobcan.es/cmayerot/medioambiente/medionatural/biodiversidad/especies/bancodatos/Lista_Especies_Silvestres.pdf).

**Botánico, J. et al.**, Estudios preliminares de Mariposas Diurnas ( Lepidoptera : Rhopalocera ) en el , 2007. p. 185-189.

**Butterfly Conservation Europe**, Why butterflies and moths are important - Why butterflies and moths are important. [en línea], 2008. p. 2. [Consulta: 3 junio 2018]. Disponible en: [http://bc-europe.eu/upload/Why\\_butterflies\\_and\\_moths\\_are\\_important\\_3.pdf](http://bc-europe.eu/upload/Why_butterflies_and_moths_are_important_3.pdf).

**Calabuig, L.**, *Aplicación del análisis de componentes principales en ecología*. 1980. S.l.: s.n. 1980.

**Camero, E. y Calderón, C.**, Comunidad de mariposas (Lepidoptera: Rhopalocera) en un gradiente altitudinal del cañón del río Combeima-Tolima, Colombia. *Acta biológica Colombiana*, 2007. p. 95-110. ISSN 0120-548X.

**Carvalho, O.**, Especies exóticas e invasiones biológicas. *Journal of Biogeography*, 2009. p. 15-21.

**Closas, A.H. et al.**, Análisis multivariante , conceptos y aplicaciones en Psicología Educativa y Psicometría. , 2013. p. 65-92.

**Cuadras, C.M.**, *Nuevos Métodos de Análisis Multivariante* [en línea]. S.l.: s.n. 2007. [Consulta: 6 julio 2018]. ISBN 9788578110796. Disponible en: [http://www.est.uc3m.es/esp/nueva\\_docencia/getafe/estadistica/analisis\\_multivariante/doc\\_generica/archivos/metodos.pdf](http://www.est.uc3m.es/esp/nueva_docencia/getafe/estadistica/analisis_multivariante/doc_generica/archivos/metodos.pdf).

**D'elía, A.**, *Los Teques de la ciudad imprevista a la capital del cambio* [en línea]. S.l.: s.n. 2010. [Consulta: 10 julio 2018]. ISBN 9789807316125. Disponible en: <http://www.libreroonline.com/venezuela/libros/558/sin-autor/los-teques-de-la-ciudad-imprevista-a-la-capital-del-cambio.html>.

**Instituto Nacional de Estadística**, XIV Censo Nacional de Población y Vivienda. [en línea], 2014. p. 1-97. Disponible en: <http://www.ine.gov.ve/documentos/Demografia/CensodePoblacionyVivienda/pdf/miranda.pdf>.

**Gauch JR., H.G.**, *Multivariate analysis in community ecology* [en línea]. 1982. S.l.: Cambridge University Press. [Consulta: 18 junio 2018]. 1982. ISBN 9780521282406. Disponible en: [https://books.google.es/books?hl=es&lr=&id=mLH\\_15sv734C&oi=fnd&pg=PR9&dq=Importance+of+multivariate+analysis&ots=4VAe1ZOQd2&sig=ueZ8sbYoISeWRhRY\\_ucnrGZhLKg#v=onepage&q=Importance+of+multivariate+analysis&f=false](https://books.google.es/books?hl=es&lr=&id=mLH_15sv734C&oi=fnd&pg=PR9&dq=Importance+of+multivariate+analysis&ots=4VAe1ZOQd2&sig=ueZ8sbYoISeWRhRY_ucnrGZhLKg#v=onepage&q=Importance+of+multivariate+analysis&f=false).

**Grimaldi, D. y Engel, M.S.**, *Evolution of the Insects*. S.l.: s.n. 2005. ISBN 9788578110796.

**Gutierrez, R. y Gutierrez-Sanchez, R.**, *Análisis de Componentes Principales*. , 2005. p. 1-37.

**Gutierrez, R. y Gutierrez-Sanchez, R.**, *Análisis Factorial*. , 2005. p. 1-62.

**Gutierrez, R. y Gutierrez-Sanchez, R.**, *Análisis de Correspondencias*. , 2005. p. 1-60.

**Díaz I, S.B.S.**, Estructura y composición de la comunidad de insectos en una especie nativa (*Lupinus bogotensis*) y una invasora (*Ulex europaeus*) a la luz de Variables ecológicas: hora y temperatura en ambientes de borde de bosque y de camino. [en línea], 2011. p. 1-23. Disponible en: <https://repository.javeriana.edu.co/bitstream/handle/10554/12436/DiazConchaIvanAndres2011.pdf?sequence=1>.

**Jiménez, E.U. y Manzano, J.A.**, *Análisis multivariante aplicado* [en línea]. 2005. S.l.: s.n. 2005. ISBN 9788497323727. Disponible en: <http://dspace.ucbscz.edu.bo/dspace/bitstream/123456789/4268/1/4542.pdf%0Ahttps://books.google.com/nl/books?id=EJ1hIRLmuiwC>.

**Kauffmann, B.G.**, *Life Cycle*. , 2002. p. 1-3.

**Khan, F.K.P. and A.**, *Introductory Chapter: Lepidoptera*. *Intech open* [en línea], 2017. p. 1-17. DOI 10.5772/32009. Disponible en: <https://www.intechopen.com/books/lepidoptera/introductory-chapter-lepidoptera>.

**Lévy Mangin, et al.**, *Modelización con estructuras de covarianzas en ciencias sociales : temas*

*esenciales, avanzados y aportaciones especiales* [en línea]. S.l.: Netbiblo. 2006. [Consulta: 25 julio 2018]. ISBN 8497451368. Disponible en: [https://books.google.com.ec/books/about/Modelización\\_con\\_estructuras\\_de\\_covaria.html?id=WEfC1TGVJBgC&redir\\_esc=y](https://books.google.com.ec/books/about/Modelización_con_estructuras_de_covaria.html?id=WEfC1TGVJBgC&redir_esc=y).

**Moreno, C.E.**, Métodos para medir la biodiversidad. *M&T - Manuales y Tesis SEA*, 2001. p. 84. ISSN 19326203. DOI 10.1371/journal.pone.0103709.

**Moreno, G. et al.**, Caracterización De Lepidópteros Diurnos En Dos Sectores Del Santuario De Flora Y Fauna Los Flamencos (San Lorenzo De Camarones, La Guajira). *Boletín Científico. Centro de Museos. Museo de Historia Natural* [en línea], 2015. p. 221-234. ISSN 01233068. DOI 10.17151/bccm.2015.19.1.16. Disponible en: [http://200.21.104.25/boletincientifico/downloads/Boletin%2819%291\\_16.pdf](http://200.21.104.25/boletincientifico/downloads/Boletin%2819%291_16.pdf).

**Munn, L.Y.**, Using Daisy ( digital automated identification system ) for automated identification of moths of the Superfamily Bombycoidea of Borneo, 2013. p. 73.

**Naoki, K.**, El pasado , presente y futuro en el uso de la estadística en estudios ecológicos en Bolivia. , 2012. p. 79-87.

**Ortega, M.T.**, Mariposas diurnas: bioindicadoras de eventos actuales e históricos Butterflies as bioindicators of present and historical events. [en línea], 2015. p. 320-347. Disponible en: [https://www.researchgate.net/profile/Marysol\\_Ortega/publication/281897722\\_Mariposas\\_diurnas\\_bioindicadoras\\_de\\_eventos\\_actuales\\_e\\_historicos\\_Butterflies\\_as\\_bioindicators\\_of\\_present\\_and\\_historical\\_events/links/55fcf3d708aec948c4c02488/Mariposas-diurnas-bioi](https://www.researchgate.net/profile/Marysol_Ortega/publication/281897722_Mariposas_diurnas_bioindicadoras_de_eventos_actuales_e_historicos_Butterflies_as_bioindicators_of_present_and_historical_events/links/55fcf3d708aec948c4c02488/Mariposas-diurnas-bioi).

**Ospina, L., et al.**, Diversidad de mariposas y su relación con el paisaje en la cuenca del río Lagunillas , Tolima , Colombia. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, 2015. p. 455-474.

**Peña, D.**, Análisis de datos multivariantes. *Book* [en línea], 2002. p. 515. ISSN 1098-6596. DOI 8448136101. Disponible en: [http://www.mhe.es/universidad/ciencias\\_matematicas/pena/index.html](http://www.mhe.es/universidad/ciencias_matematicas/pena/index.html).

**Rencher, A.C.**, *Methods of Multivariate Analysis, Second Edition*. S.l.: s.n. 2012. ISBN 0471418897.

**Sampieri H., et al.,** *Metodología de la investigación*. S.l.: s.n. [sin fecha]. ISBN 9786071502919.

**Silva Ayçaguer, et al.,** *Regresión logística* [en línea]. S.l.: La Muralla, Hespérides. 2004. [Consulta: 6 agosto 2018]. ISBN 9788471337382. Disponible en: <https://latam.casadellibro.com/libro-regresion-logistica/9788471337382/948990>.

**Solomon, J.D.,** *Guide to insect borers in North American broadleaf trees and shrubs - Order:Lepidoptera*. 1995. S.l.: s.n. 1995.

**Urretabizkaya, et al.,** Insectos perjudiciales de importancia agronómica: Lepidópteros. *Instituto Nacional de Tecnología Agropecuaria* [en línea], 2010. p. 1-77. Disponible en: [https://inta.gob.ar/sites/default/files/script-tmp-inta\\_lepidopteros.pdf](https://inta.gob.ar/sites/default/files/script-tmp-inta_lepidopteros.pdf).

**Zamora, R. y Esnaola, J.,** Análisis factorial y de componentes principales. *Análisis multivariado: un manual para investigadores* [en línea]. S.l.: s.n., 2007. pp. 121-159. Disponible en: [https://www.u-cursos.cl/facso/2015/1/SO01023/1/material\\_docente/bajar?id\\_material=1049217](https://www.u-cursos.cl/facso/2015/1/SO01023/1/material_docente/bajar?id_material=1049217).