



INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y  
APLICADAS

CENTRO DE INVESTIGACIÓN EN CIENCIAS

# Análisis estadístico de series de tiempo de bases de datos agrícolas

T E S I S

Para obtener título de

LICENCIADO EN CIENCIAS MATEMÁTICAS

Presenta

Francisco Xavier Gil Salgado

Director de tesis:

Dr. Jesús Igor Heberto Barahona Torres

Co-director de tesis:

Dr. Ronald Ernesto Ontiveros Capurata

Sinodales:

Dr. Daniel Rivera López

Dr. Raúl Salgado García

Dra. Lorena Díaz Gonzales

Dr. Paulo César Manrique Mirón

Cuernavaca, Morelos

Marzo, 2020



# Índice general

<i>Resumen</i> .....	0
<b>Capítulo 1</b> .....	1
1.1 <i>Introducción</i> .....	1
1.2 <i>Planteamiento del problema y justificación</i> .....	2
1.3 <i>Hipótesis</i> .....	2
1.4 <i>Objetivo general</i> .....	2
1.5 <i>Objetivos particulares</i> .....	2
<b>Capítulo 2</b> .....	3
2.1 <i>Principios básicos de algebra lineal y estadística</i> .....	3
2.2 <i>Análisis de componentes principales (ACP)</i> .....	5
2.3 <i>Análisis de factores (AF)</i> .....	5
2.4 <i>Análisis de correspondencia (AC)</i> .....	6
2.5 <i>Series de tiempo</i> .....	7
2.6 <i>Modelo ARIMA</i> .....	9
2.7 <i>Suavización Exponencial Simple (SES)</i> .....	11
2.8 <i>Holt-Winters</i> .....	12
<b>Capítulo 3</b> .....	14
3.1 <i>El lenguaje de programación R</i> .....	14
3.2 <i>Principales características</i> .....	15
3.3 <i>Interfaces gráficas (IG)</i> .....	16
3.3.1 <i>RStudio</i> .....	16
3.3.2 <i>Commander</i> .....	17
3.3.3 <i>Weka</i> .....	18
3.4 <i>Librerías utilizadas</i> .....	18
3.4.1 <i>FactoMineR</i> .....	19
3.4.2 <i>Factoextra</i> .....	20
3.4.3 <i>Ggplot2</i> .....	20
<b>Capítulo 4</b> .....	21
4.1 <i>Conjunto de datos inicial</i> .....	21
4.2 <i>Resultados del análisis de correspondencias</i> .....	23
4.3 <i>Análisis de series de tiempo</i> .....	37
4.4 <i>Contrastando con la realidad</i> .....	43

<b>Capítulo 5</b> .....	44
<i>5.1 Conclusiones</i> .....	44
<i>5.2 Futuras líneas de investigación</i> .....	45
<b>Referencias</b> .....	46

# Índice de figuras

<b>Figura 0.1.</b> Estructura de la tesis (Elaboración propia) . . . . .	0
<b>Figura 2.4.1.</b> Ejemplo de un diagrama de dispersión. . . . .	6
<b>Figura 3.1.1.</b> Aspecto de la página principal (Tomado de CRAN, 2018) . . . . .	14
<b>Figura 3.3.1.1.</b> Interfaz gráfica de RStudio . . . . .	16
<b>Figura 3.3.2.1.</b> Interfaz de Commander . . . . .	17
<b>Figura 3.3.3.1.</b> Interfaz de Weka . . . . .	18
<b>Figura 3.4.1.1.</b> Aspecto de su página web . . . . .	19
<b>Figura 4.2.1.</b> Gráfico de sedimentación para la producción del distrito DR-063 . . . . .	24
<b>Figura 4.2.2.</b> Nube de puntos para la producción de los principales cultivos del distrito DR-063 . . . . .	25
<b>Figura 4.2.3.</b> Gráfico de sedimentación para la producción del distrito DR-075 . . . . .	27
<b>Figura 4.2.4.</b> Nube de puntos para la producción de los principales cultivos del distrito DR-075 . . . . .	28
<b>Figura 4.2.5.</b> Gráfico de sedimentación para la producción del distrito DR-076 . . . . .	29
<b>Figura 4.2.6.</b> Nube de puntos para la producción de los principales cultivos del distrito DR-076 . . . . .	30
<b>Figura 4.2.7.</b> Gráfico de sedimentación para el rendimiento del distrito de riego DR-063 . . . . .	31
<b>Figura 4.2.8.</b> Nube de puntos para el rendimiento del distrito DR-063 . . . . .	32
<b>Figura 4.2.9.</b> Gráfico de sedimentación para el rendimiento del distrito DR-075. . . . .	33
<b>Figura 4.2.10.</b> Nube de puntos para el rendimiento de los cultivos del distrito DR-075 . . . . .	34
<b>Figura 4.2.11.</b> Gráfico de sedimentación para el rendimiento del distrito DR-076 . . . . .	35
<b>Figura 4.2.12.</b> Nube de puntos para el rendimiento del distrito DR-076 . . . . .	36
<b>Figura 4.3.1.</b> Serie de tiempo para la producción de los principales cultivos del distrito DR-063 . . . . .	37
<b>Figura 4.3.2.</b> Serie de tiempo para la producción de los principales cultivos del distrito DR-075 . . . . .	38
<b>Figura 4.3.3.</b> Serie de tiempo para la producción de los principales cultivos del distrito DR-076 . . . . .	37
<b>Figura 4.3.4.</b> Serie de tiempo para el rendimiento de los principales cultivos del distrito DR-063 . . . . .	40
<b>Figura 4.3.5.</b> Serie de tiempo para el rendimiento de los principales cultivos del distrito DR-075 . . . . .	41
<b>Figura 4.3.6.</b> Serie de tiempo para el rendimiento de los principales cultivos del distrito DR-076 . . . . .	42

# Índice de tablas

<b>Tabla 4.1.1.</b> Dimensiones de los conjuntos de datos utilizados en el análisis .....	21
<b>Tabla 4.1.2.</b> Cantidad de datos registrados y valores faltantes .....	22
<b>Tabla 4.1.3.</b> Cultivos removidos y utilizados en el análisis .....	22
<b>Tabla 5.1.2.</b> Medidas de la raíz del error cuadrático medio para cada modelo .....	45

# Claves usadas

<b>Clave</b>	<b>Cultivo</b>
ALGD	Algodón
CLBZ	Calabaza
C_AZ	Caña de Azúcar
CBLL	Cebolla
EJT	Ejote
FRJL	Frijol (Alubia)
GRB	Garbanzo
JTE	Jitomate (Tomate Rojo)
MA_GR	Maíz grano
O_HTZ	Otras Hortalizas
O_FTL	Otros Frutales
O_PST	Otros Pastos
PAP	Papa
PEP	Pepino
SND	Sandía
SRG_GRN	Sorgo Grano
TMT_CSA	Tomate de Cáscara (Tomatillo)
TRG_GRN	Trigo Grano

# Resumen

En esta tesis se trabajará con tres bases de datos. Se les realizará un análisis multivariante y de series de tiempo. Finalmente, los resultados serán contrastados con la realidad. La tesis se compone de cinco capítulos. El primero hace una presentación de los objetivos de la tesis, las preguntas e hipótesis de investigación. En el segundo capítulo se describe la metodología, es decir, la matemática que se ha utilizado en el análisis de los datos. Una descripción de las herramientas computacionales empleadas en el análisis se presenta en el capítulo tres. En el capítulo cuatro se presentan los resultados obtenidos. Y, por último, en el capítulo cinco, aparecen las conclusiones y las futuras líneas de investigación (ver figura 0.1).

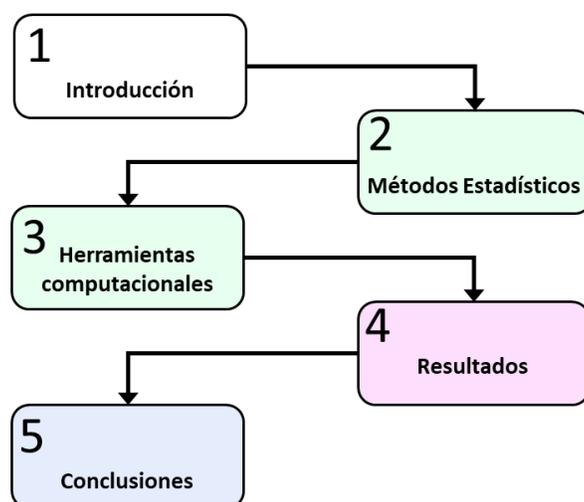


Figura 0.1. Estructura de la tesis (Elaboración propia)

# Capítulo 1

## *1.1 Introducción*

México cuenta con varias instituciones que se encargan de recopilar información estadística sobre el uso del suelo y el agua. Por ejemplo, la SEMARNAT se encarga de recopilar datos de los recursos naturales de todo el país. Por otro lado, SAGARPA se encarga de producir información estadística, geográfica y agroalimentaria. Y la CONAGUA se encarga de asignar y gestionar los recursos hídricos en las zonas agrícolas del país. Esto muestra el gran interés del país por conocer el estado de sus recursos naturales.

## *1.2 Planteamiento del problema y justificación*

El principal problema es que toda esa información se encuentra dispersa en las distintas bases de datos de las diferentes instituciones, y el acceso a ella es muy limitada, por lo que conocer información tal como: los volúmenes exactos de agua usados y las superficies exactas regadas con esa agua es difícil. Esto hace que nos veamos obligados a aprovechar al máximo cualquier base de datos que cualquiera de estas instituciones llegue a proporcionar.

El Instituto Mexicano de Tecnología del Agua (IMTA) cuenta con una gran cantidad de información relacionada con el uso del agua en distintas regiones del país de las últimas décadas, y fueron ellos quienes proporcionaron las bases de datos y recursos para poder llevar a cabo este trabajo.

### *1.3 Hipótesis*

Partiendo de una base de datos con información sobre la producción y el rendimiento de 67 tipos de cultivos en los últimos 50 años, es posible aplicar algunos métodos matemáticos de pronóstico y de series de tiempo para identificar patrones subyacentes en estos, los cuales ayudarían a mejorar el proceso de toma de decisiones.

### *1.4 Objetivo general*

Hacer una contribución en el sentido de mejorar el aprovechamiento de los datos disponibles.

### *1.5 Objetivos particulares*

- Sobre las bases de datos antes mencionadas, se realizarán procedimientos de limpieza y depuración. Lo anterior es necesario para realizar un análisis confiable.
- Se aplicarán métodos estadísticos multivariantes (como análisis de componentes principales y análisis de correspondencias) para identificar patrones subyacentes y relaciones en las variables analizadas.
- Se realizará un análisis de series de tiempo el cual incluya tres métodos de estimación de valores futuros (ARIMA, Suavización exponencial y Holt-Winters) y comparar los resultados.

# Capítulo 2

## 2.1 Principios básicos de estadística y álgebra lineal

De acuerdo con Cuadras (2014), el *análisis multivariante* es una rama de la estadística que se utiliza para estudiar e interpretar datos que provienen de registrar una o más variables de una muestra de observaciones. Debido a esto, la información que se analiza es de carácter multidimensional, por lo que el cálculo matricial juega un papel fundamental. Por esa razón se introducirán a continuación algunas definiciones, y más adelante se hablará de las técnicas de análisis multivariante que serán usadas para analizar las bases de datos recopiladas por el IMTA.

Para comenzar el análisis multivariante necesitamos la matriz que contiene las observaciones que se desea estudiar. Esta matriz  $\mathbf{X}$  mide el grado de discrepancia entre los objetos que se van a analizar. Llamemos a esos objetos  $\omega_1, \dots, \omega_m$  y  $X_1, \dots, X_n$  serán las variables que se observaron en ellos. Así  $x_{ij} = X_j(\omega_i)$  es la observación que se obtuvo de la variable  $X_j$  sobre el objeto  $\omega_i$ . De esta manera, la matriz multivariante construida a partir de los objetos antes mencionados toma la siguiente forma:

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mj} & \dots & x_{mn} \end{pmatrix} \quad (2.1.1)$$

Con esta matriz podemos construir un vector y dos matrices cuyas propiedades serán de mucha ayuda a la hora de realizar un análisis de los datos:

1.  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_j, \dots, \bar{x}_n)'$  será el vector columna de las medias de las variables, donde

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

2. La matriz simétrica  $\mathbf{S}$  será aquella que guarda las covarianzas muestrales

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix} \quad (2.1.2)$$

donde

$$s_{j_1 j_2} = \frac{1}{m} \sum_{i=1}^m (x_{ij_1} - \bar{x}_{j_1})(x_{ij_2} - \bar{x}_{j_2})$$

es la covarianza entre las variables  $j_1$  y  $j_2$ .

3. La matriz simétrica  $\mathbf{R}$  será aquella que guarda las correlaciones muestrales

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & 1 \end{pmatrix} \quad (2.1.3)$$

siendo  $r_{j_1 j_2} = \text{cor}(X_{j_1}, X_{j_2})$  el coeficiente de correlación entre las variables  $X_{j_1}$  y  $X_{j_2}$ . Este coeficiente está dado por

$$r_{j_1 j_2} = \frac{s_{j_1 j_2}}{s_{j_1} s_{j_2}}$$

donde  $s_{j_1}$  y  $s_{j_2}$  son las desviaciones típicas.

Cómo puede apreciarse, los objetos que se van a analizar están situados en espacios de altas dimensiones, de modo que no se pueden observar directamente. Sin embargo, es posible determinar subespacios de dimensión menor (el número mínimo entre filas y columnas menos uno) donde se puede aproximar la posición original de los objetos. A continuación, serán descritos algunos métodos para construir esos subespacios.

## 2.2 Análisis de componentes principales (ACP)

El análisis de las componentes principales es una técnica estadística que tiene como objetivo transformar un conjunto de variables  $X_1, \dots, X_n$  a las que se les denomina *originales*, en un nuevo conjunto de variables  $y_1, y_2, \dots, y_p$  ( $p < n$ ) denominadas *componentes principales*. Estas componentes se caracterizan por estar incorreladas entre sí. Para obtenerlas, asumimos que las variables originales están relacionadas con las componentes de la siguiente manera:

$$\begin{aligned} y_1 &= \lambda_{11}X_1 + \dots + \lambda_{1n}X_n \\ &\vdots \\ y_p &= \lambda_{p1}X_1 + \dots + \lambda_{pn}X_n \end{aligned} \quad (2.2.1)$$

cada componente es una combinación lineal de las variables originales. También puede escribirse de esta otra manera:

$$\mathbf{y} = \Lambda \mathbf{X}$$

donde:

$$\Lambda = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \dots & \lambda_{pn} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

De acuerdo con [Wolfgang \(2015\)](#), los coeficientes de la matriz  $\Lambda$  en realidad son los *vectores propios* de la matriz de correlaciones (2.1.3). Por lo que solo bastaría calcularlos (existen hoy en día softwares capaces de realizar esta tarea) para poder obtener las componentes principales.

## 2.3 Análisis de factores (AF)

El AF es una técnica de reducción de datos que pretende expresar  $n$  variables como una combinación lineal de  $p$  variables hipotéticas o latentes, denominadas *factores*. Tiene una formulación parecida a la del ACP, pero el modelo que relaciona variables y factores es diferente en AF. En AF asumimos que cada variable  $X_1, \dots, X_n$  depende de unos factores comunes  $y_1, y_2, \dots, y_p$  ( $p < n$ ) y de un factor único  $u_1, \dots, u_n$ :

$$\begin{aligned} X_1 &= \lambda_{11}y_1 + \dots + \lambda_{1p}y_p + u_1 \\ &\vdots \\ X_n &= \lambda_{n1}y_1 + \dots + \lambda_{np}y_p + u_n \end{aligned}$$

Los  $\lambda_{ij}$  son llamados *pesos factoriales*. Esta relación también puede escribirse de la siguiente manera

$$\mathbf{X} = \Lambda \mathbf{y} + \mathbf{u}$$

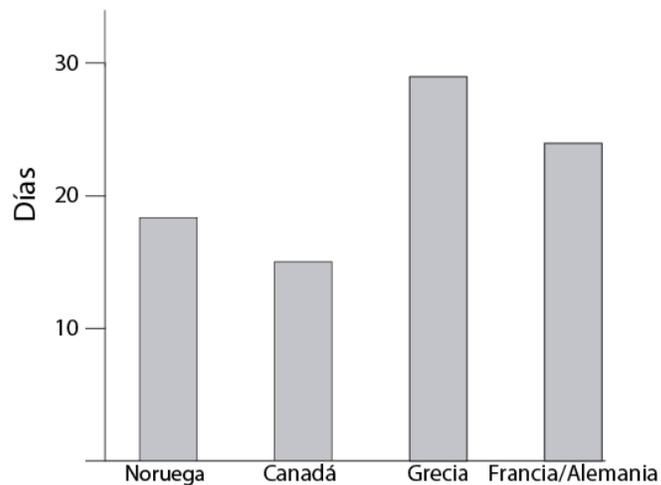
donde

$$\Lambda = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \dots & \lambda_{pn} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

Existen diversos métodos para extraer los factores, pero uno de los más utilizados es el *método de las componentes principales*. Una vez que se extraen los factores, sólo resta observar si es necesaria una *rotación* para tener una mejor interpretación de los datos.

#### 2.4 Análisis de correspondencia (AC)

El análisis de correspondencia es un método de análisis de datos que representa gráficamente tablas de datos. [Greenacre \(2008\)](#) explica que el análisis de correspondencia es una generalización de una representación gráfica con la que todos estamos especializados: *el diagrama de dispersión*. También nos proporciona un ejemplo.



**Figura 2.4.1.** Ejemplo de un diagrama de dispersión

Un diagrama de dispersión representa los datos en forma de puntos con relación a dos ejes de coordenadas perpendiculares. La Figura 2.4.1 representa los viajes que hizo un investigador para dar unas conferencias. Pasó 18 días en Noruega, 15 días en Canadá y 29 en Grecia. También hizo unos viajes a Francia y Alemania que tuvieron una duración de 24 días en total. Véase que mientras el eje  $y$  representa los Días, una variable *continua*, el eje  $x$  es claramente una variable no *numérica*. En este eje, los cuatro puntos sólo son posiciones en las que hemos situado las etiquetas que indican el país visitado. La escala horizontal representa una variable *categorica*. Hay dos características de este eje horizontal que no tienen significado sustantivo alguno en el gráfico: la ordenación de las etiquetas y la distancia entre ellas. Otro ejemplo de esto es que, en muchas encuestas sociales, se dan las respuestas en una escala ordinal. Un ejemplo típico de esto es la famosa escala “de acuerdo/desacuerdo”.

Con esto se quiere mostrar que, a menudo, la naturaleza de la variable  $x$  es completamente distinta a la de la variable  $y$ , de manera que solamente podemos interpretar distancia en la dirección de uno de los dos ejes, de acuerdo con una determinada escala de medida con la que hayamos calibrado el eje.

Una de las diferencias del AC con los otros dos métodos anteriores es la capacidad de poder analizar variables *categoricas*.

La calidad de representación en subespacios de dimensión menor construidos a partir de cualquiera de los métodos antes mencionados, se mide en porcentajes de inercia (la inercia es sinónimo de varianza) con respecto a la total.

## 2.5 Series de tiempo

De acuerdo con [Wei \(2006\)](#), una serie de tiempo es una secuencia ordenada de observaciones, en donde cada observación corresponde a una variable aleatoria distinta. La ordenación de la sucesión de observaciones es esencial para el análisis de la misma, no se puede alterar porque se cambiarían las características de la serie que se quiere estudiar. Además, cada serie de tiempo puede ser descompuesta en diferentes componentes, estos son:  $y_t$  el valor observado,

$T_t$  el factor de la tendencia en el periodo  $t$ ,  $E_t$  el factor de la estacionalidad en el periodo  $t$ ,  $C_t$  el factor cíclico en el periodo  $t$  y  $I_t$  el factor de irregularidad en el periodo  $t$ .

Wei (2006) también menciona que una serie temporal es en cierta forma un *proceso estocástico* con las variables aleatorias  $Z(w, t)$ , en donde los parámetros  $w$  y  $t$  representan el *espacio muestral* y el índice de sucesión respectivamente. Cuando  $t$  es fijo, entonces se tiene el caso de que  $Z(w, t)$  es una variable aleatoria. De esta forma, la distribución para la sucesión de variables aleatorias  $Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}$  puede escribirse de la siguiente forma:

$$F_{Z_{t_1}, \dots, Z_{t_n}}(x_1, \dots, x_n) = P\{w: Z_{t_1} \leq x_1, \dots, Z_{t_n} \leq x_n\} \quad (2.5.1)$$

Donde  $x_i, i = 1, \dots, n$  son números reales cualesquiera. Las condiciones que ha de cumplir el proceso son: que el proceso sea no *anticipante*, i.e. que el presente no venga determinado por el futuro, y que sea *invertible*, i.e. que el presente dependa de forma convergente de su propio pasado. En caso de que la función de distribución uno-dimensional sea invariante en el tiempo entonces diremos que el proceso es *estacionario en distribución de primer orden*. Es decir:

$$F_{Z_{t_1}}(x_1) = F_{Z_{t_1+k}}(x_1) \quad (2.5.2)$$

Para enteros cualesquiera  $t_1, k$ , y  $t_1 + k$ . Si se cumple que:

$$F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$$

Para  $t_1, t_2, k, t_1 + k$  y  $t_2 + k$ , entonces decimos que la función es *estacionaria en distribución de segundo orden*. Finalmente, y de modo general, si la función de distribución cumple que:

$$F_{Z_{t_1}, \dots, Z_{t_n}}(x_1, \dots, x_n) = F_{Z_{t_1+k}, \dots, Z_{t_n+k}}(x_1, \dots, x_n) \quad (2.5.3)$$

Para enteros cualesquiera  $(t_1, \dots, t_n)$  y  $k$ , se dice que el proceso es *estrictamente estacionario*. Por último, definiremos la función media del proceso como:

$$\mu_t = E(Z_t) \quad (2.5.4)$$

La función de varianza del proceso se define como:

$$\sigma_t^2 = E(Z_t - \mu_t)^2 \quad (2.5.5)$$

Y la función de correlación entre  $Z_{t_1}$  y  $Z_{t_2}$  es como sigue:

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}) \quad (2.5.6)$$

El cálculo de pronósticos confiables es una tarea importante en muchas áreas. Existen muchos tipos diferentes de procedimientos de pronóstico, pero no hay una respuesta simple sobre cuál es el "mejor". Existe una amplia variedad de problemas de pronóstico que requieren un tratamiento diferente y, por lo tanto, la elección del método depende de una variedad de consideraciones que incluyen los objetivos, las propiedades de los datos, el número de series a pronosticar, y así sucesivamente, consulte [Chatfield \(1988\)](#).

## 2.6 Modelo ARIMA

En el análisis de series de tiempo, existen dos representaciones muy útiles para expresar las series de tiempo. Ambas consisten en escribir el proceso  $Z_t$  como una combinación lineal de una secuencia de variables aleatorias incorrelacionadas. El primer modelo se denomina proceso *autorregresivo* y está dado por:

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \dots + \phi_p \dot{Z}_{t-p} + a_t \quad a_t \sim RB(0, \sigma^2) \quad (2.6.1)$$

Ahora, introduciendo el *operador de retardo*, el cual es  $B^j x_t = x_{t-j}$ , podemos escribir la ecuación anterior como:

$$\Phi_p(B) \dot{Z}_t = a_t \quad (2.6.2)$$

Donde  $\Phi_p(B) = 1 - \sum_{j=1}^p \phi_j B^j$  y  $\dot{Z}_t = Z_t - \mu$ . Un proceso con estas características se denomina *proceso autorregresivo de orden p*, el cual se denota por  $AR(p)$ . [Box y Jenkins \(1970\)](#) llamaron a todo proceso que puede ser escrito de esta forma *invertible*. También

explican que todo proceso estacionario es invertible, sin embargo, no todo proceso invertible es estacionario. Otra forma útil de escribir el proceso  $Z_t$  es:

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.6.3)$$

Y usando el operador de retardo se puede reescribir cómo:

$$\dot{Z}_t = \theta_q(B)a_t \quad (2.6.4)$$

Donde  $\theta_q(B) = 1 - \sum_{j=1}^q \theta_j B^j$ . Este modelo se conoce con el nombre de *medias móviles de orden  $q$*  y se denota por  $MA(q)$ . [Wold \(1938\)](#) demostró que un proceso estacionario, el cual no es determinista siempre puede ser expresado en la forma (2.6.4), por lo que a veces también es llamado un *proceso no determinístico*.

Una extensión natural de ambos procesos es la mezcla de ellos, lo que nos da un caso especial. De este modo, un proceso que es estacionario e invertible puede ser representado como:

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \dots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.6.5)$$

O también:

$$\Phi_p(B)\dot{Z}_t = \theta_q(B)a_t \quad (2.6.6)$$

Para que el proceso sea invertible, necesitaremos que las raíces de  $\theta_q(B) = 0$  se encuentren fuera del círculo unidad. Para que sea estacionario, se requiere que las raíces de  $\Phi_p(B) = 0$  estén fuera del círculo unidad. También se asumirá que ambos polinomios no tienen raíces en común.

Hasta este punto, hemos asumido que los procesos son todos estacionarios, pero en muchas ocasiones, las series que se estudian no son estacionarias. La no estacionariedad puede presentarse de diferentes maneras, ya sea con respecto a la media, la varianza, o ambas.

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \dots + \phi_p \dot{Z}_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad a_t \sim RB(0, \sigma^2) \quad (2.6.7)$$

O también:

$$\Phi_p(B)\dot{Z}_t = \theta_q(B)a_t \quad (2.6.8)$$

Donde  $\Phi_p(B)$  es el polinomio autorregresivo y  $\theta_q(B)$  es el polinomio de medias móviles. Aunque muchas series temporales no son estacionarias, debido a algunas fuerzas de equilibrio, diferentes partes de estas series se comportan de manera muy similar, excepto por su diferencia en los niveles medios locales. [Box & Jenkins \(1970\)](#) se refieren a este tipo de comportamiento no estacionario como *no estacionario y homogéneo*.

Una serie temporal no estacionaria y homogénea puede ser reducida a un tipo de serie estacionaria, a través de tomar un grado apropiado de diferenciación. Los modelos de promedios móviles autorregresivos son de utilidad para describir series temporales estacionarias. De acuerdo con [Wei \(1996\)](#), un proceso estacionario que es obtenido a partir de las raíces unitarias de una serie temporal homogénea no estacionaria, no necesariamente resulta en ruido blanco. Por ejemplo, sea  $(1 - B)^d Z_t$  una serie temporal estacionaria, autorregresiva y con promedios móviles. Entonces, un modelo ARIMA (Autorregresivo Integrado de Promedios Móviles) está dado por la siguiente formula:

$$\Phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (2.6.9)$$

En donde el operador estacionario AR  $\Phi_p(B) = 1 - \sum_{j=1}^p \phi_j B^j$  y el operador invertible MA  $\theta_q(B) = 1 - \sum_{j=1}^q \theta_j B^j$  no tienen factores en común. El parámetro  $\theta_0$  juega diferentes roles para  $d = 0$  y  $d > 0$ . Cuando  $d = 0$ , el proceso original es estacionario y  $\theta_0$  es llamado la media del proceso, i.e.,  $\theta_0 = 1 - \sum_{i=0}^p \phi_i$ , y cuando  $d \geq 1$ ,  $\theta_0$  es llamado término de tendencia determinista.

## 2.7 Suavización Exponencial Simple (SES)

De acuerdo [Wei \(2006\)](#) este método es adecuado para pronosticar datos sin tendencia clara o patrón estacional. El SES puede considerarse como una evolución del método de promedio móvil ponderado, en este caso se calcula el promedio de una serie de tiempo con un mecanismo de autocorrección que busca ajustar los pronósticos en dirección opuesta a las desviaciones del pasado. Esto mediante una corrección que se ve afectada por un coeficiente

de suavización. De acuerdo con [Trubetskoy \(2016\)](#), la clave detrás del SES consiste en dar prioridad a las observaciones más recientes, en contraste con aquellas de un pasado más distante. Los pesos disminuyen exponencialmente a medida que las observaciones son más lejanas en el pasado. De esta forma, los pesos más pequeños se asocian con las observaciones más antiguas, tal como se ilustra en la formulación (2.7.1).

$$\hat{y}_{T+1|T} = \alpha y_T + (1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots \quad (2.7.1)$$

Donde  $0 \leq \alpha \leq 1$  es el parámetro de suavizado. El pronóstico de un paso adelante para el tiempo  $T + 1$  es un promedio ponderado de todas las observaciones en la serie  $y_1, \dots, y_T$ . Se controla la velocidad a la que disminuyen los pesos por el parámetro  $\alpha$ . Para cualquier  $\alpha$  entre 0 y 1, los pesos adjuntos a las observaciones disminuyen exponencialmente a medida que retrocedemos en el tiempo, de ahí el nombre de "suavizado exponencial". Si  $\alpha$  es pequeño (es decir, cercano a 0), se da más peso a las observaciones del pasado más distante. Si  $\alpha$  es grande (es decir, cercano a 1), se da más peso a las observaciones más recientes.

Existen dos formas equivalentes de SES, la forma *promedio ponderada* y la forma *componente*, sin embargo, ambas conducen a la ecuación de pronóstico (2.7.1).

## 2.8 Holt-Winters

[Chatfield & Mohammad \(1988\)](#) explican que el método Holt-Winters, el cual es un caso particular de la suavización exponencial, permite que los datos sean modelados por una media local  $L_t$ , una tendencia local  $T_t$  y un factor estacional local  $I_t$ . Supongamos que  $\alpha$ ,  $\gamma$ , y  $\delta$  denotan los parámetros de suavizado para actualizar el nivel medio, la tendencia y el índice estacional respectivamente, tal como se muestra a continuación:

$$\begin{aligned} L_t &= \alpha(X_t / I_{t-p}) + (1 - \alpha) (L_{t-1} + T_{t-1}) \\ T_t &= \gamma(L_t - L_{t-1}) + (1 - \gamma) T_t \\ I_t &= \delta(X_t / L_t) + (1 - \delta) I_{t-p} \end{aligned}$$

Y el nuevo pronóstico hecho en el tiempo  $t$  del valor  $k$  períodos por delante viene dado por:

$$\hat{X}_t(k) = (L_t + kT_t) I_{t-p+k}$$

Para  $k = 1, 2, \dots, p$ .

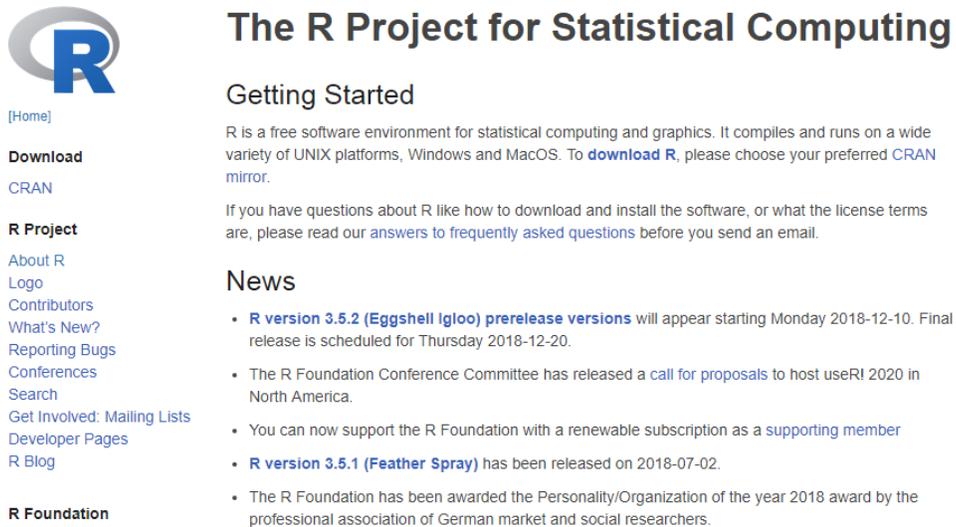
El suavizado exponencial es el caso de 1 parámetro donde sólo se actualiza  $L_t$ . En el caso de 2 parámetros,  $L_t$  y  $T_t$  se actualizan y esto a veces se denomina Modelo de Tendencia Lineal de Holt. Es adecuado para datos no estacionales. El caso completo de 3 parámetros a veces se llama Modelo Estacional de Winter. Para implementar este método, se debe:

- a) Proporcionar valores iniciales para  $L_t$ ,  $T_t$ , y  $I_t$  al comienzo de la serie
- b) Proporcionar los valores para  $\alpha$ ,  $\gamma$ , y  $\delta$
- c) Decidir si se normalizan los factores estacionales, y
- d) Elegir entre un enfoque automático o no-automático

# Capítulo 3

## 3.1 El lenguaje de programación R

Su desarrollo actual es responsabilidad del R Development Core Team. Para saber más al respecto, en el entorno del programa puede teclearse “contributors(;)” en la lista desplegada aparecen los nombres de los autores iniciales y los actuales, pertenecientes al R Development Core Team (Equipo Central de Desarrolladores R).



The screenshot shows the homepage of The R Project for Statistical Computing. On the left is a navigation menu with links for [Home], Download CRAN, R Project (About R, Logo, Contributors, What's New?, Reporting Bugs, Conferences, Search, Get Involved: Mailing Lists, Developer Pages, R Blog), and R Foundation. The main content area features the title 'The R Project for Statistical Computing', a 'Getting Started' section with introductory text and a link to download R, a 'News' section with a list of recent updates including R version 3.5.2 pre-release and R version 3.5.1 release, and a 'R Foundation' section.

### The R Project for Statistical Computing

#### Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

#### News

- [R version 3.5.2 \(Eggshell Igloo\) prerelease versions](#) will appear starting Monday 2018-12-10. Final release is scheduled for Thursday 2018-12-20.
- The R Foundation Conference Committee has released a [call for proposals](#) to host useR! 2020 in North America.
- You can now support the R Foundation with a renewable subscription as a [supporting member](#)
- [R version 3.5.1 \(Feather Spray\)](#) has been released on 2018-07-02.
- The R Foundation has been awarded the Personality/Organization of the year 2018 award by the professional association of German market and social researchers.

#### R Foundation

Figura 3.1.1. Aspecto de la página principal (Tomado de CRAN, 2018)

### *3.2 Principales características*

Entre las ventajas de R, está su gran similitud sintáctica con S-Plus, lo que ha facilitado la migración de muchos usuarios, sus frecuentes versiones y el hecho de que puede ejecutarse sobre cualquier sistema operativo convencional, e incluso en plataformas tan inverosímiles como puede ser una PlayStation 3. Algunas de sus principales características son:

- Eficiente para el manejo y almacenamiento de grandes bases de datos
- Acceso a un amplio rango de herramientas para realizar cálculos complejos en matrices y / o tensores
- Tales herramientas se encuentran interconectadas, lo cual aumenta la potencia y eficiencia en los análisis
- Interface gráfica independiente del análisis
- Como la mayoría de los lenguajes de programación de última generación, R incluye condicionales, bucles, funciones definidas por el usuario y recursivas
- En entorno de programación R está diseñando para operar con paquetes preestablecidos. Lo anterior permite extender las aplicaciones de R a un amplio rango de problemas y disciplinas.

R proporciona una amplia gama de herramientas estadísticas (modelos lineales y no lineales, pruebas estadísticas, análisis de series temporales, algoritmos de clasificación y agrupamiento, etc.) y gráficas. Al igual que S, se trata de un lenguaje de programación, lo que permite que los usuarios lo extiendan definiendo sus propias funciones, además, también hereda de S su orientación a objetos. Otra de las características de R es su capacidad gráfica, que permite generar gráficos con alta calidad. R también puede usarse como herramienta de cálculo numérico, campo en el que puede ser tan eficaz como otras herramientas, tales como GNU Octave o MATLAB.

### 3.3 Interfaces Gráficas (IG)

En esta sección se explican algunas de las IG más populares que existen en R. En este sentido, una IG es un medio a través del cual, el usuario puede comunicarse con la computadora (o R, en este caso). Algunas de las características más importantes de las IGs, es que estas deben ser fáciles de entender y de usar.

#### 3.3.1 RStudio

RStudio es un Entorno de Desarrollo Integrado (EDI) construido exclusivamente para el lenguaje de programación R. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo. Entre sus características más notables encontramos la posibilidad de ejecutar código R directamente desde el editor de código fuente, salto rápido a las funciones definidas y el resaltado de sintaxis, auto completado de código y sangría inteligente, además de un depurador interactivo para diagnosticar y corregir los errores rápidamente.

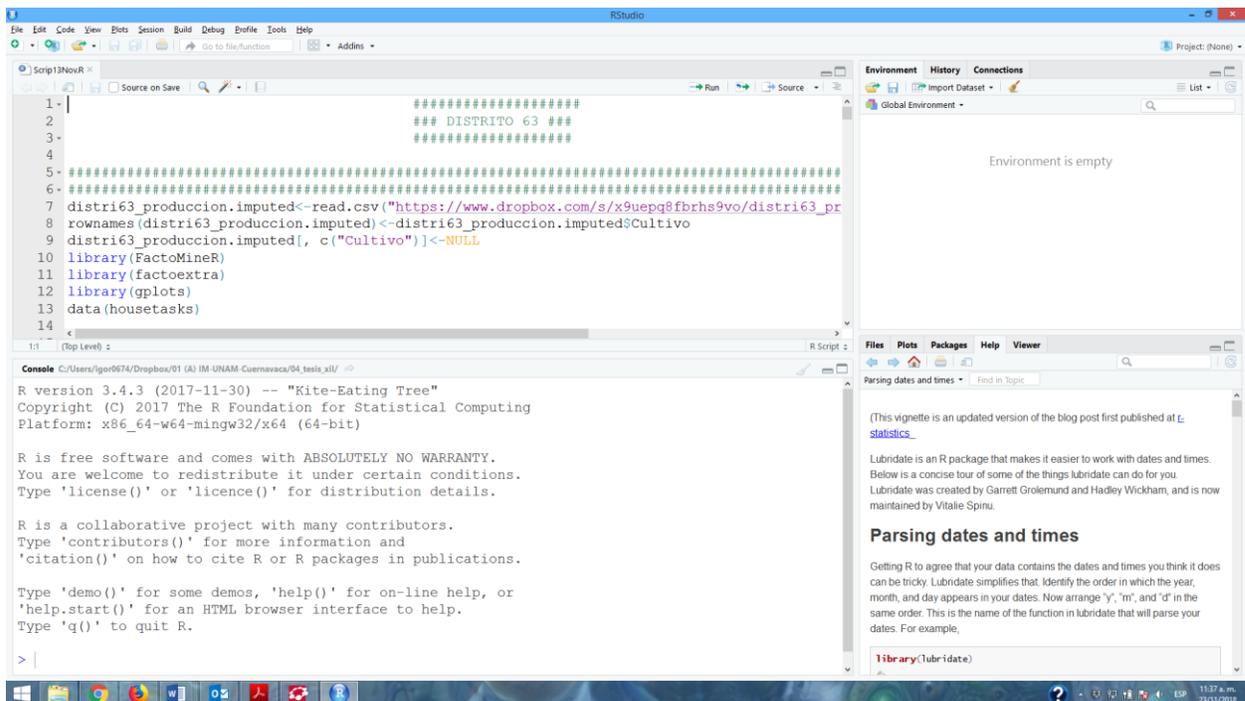


Figura 3.3.1.1. Interfaz gráfica de RStudio

### 3.3.2 Commander

También conocido como **Rcmdr**, esta aplicación es una interfaz gráfica de usuario. Desarrollada y mantenida por John Fox del Departamento de Sociología de la Universidad de McMaster, es quizá la alternativa más viable basada en R a los paquetes estadísticos como SPSS. El paquete es especialmente útil para los novatos en R, pues con cada opción de menú utilizada presenta el código R que se ha ejecutado.

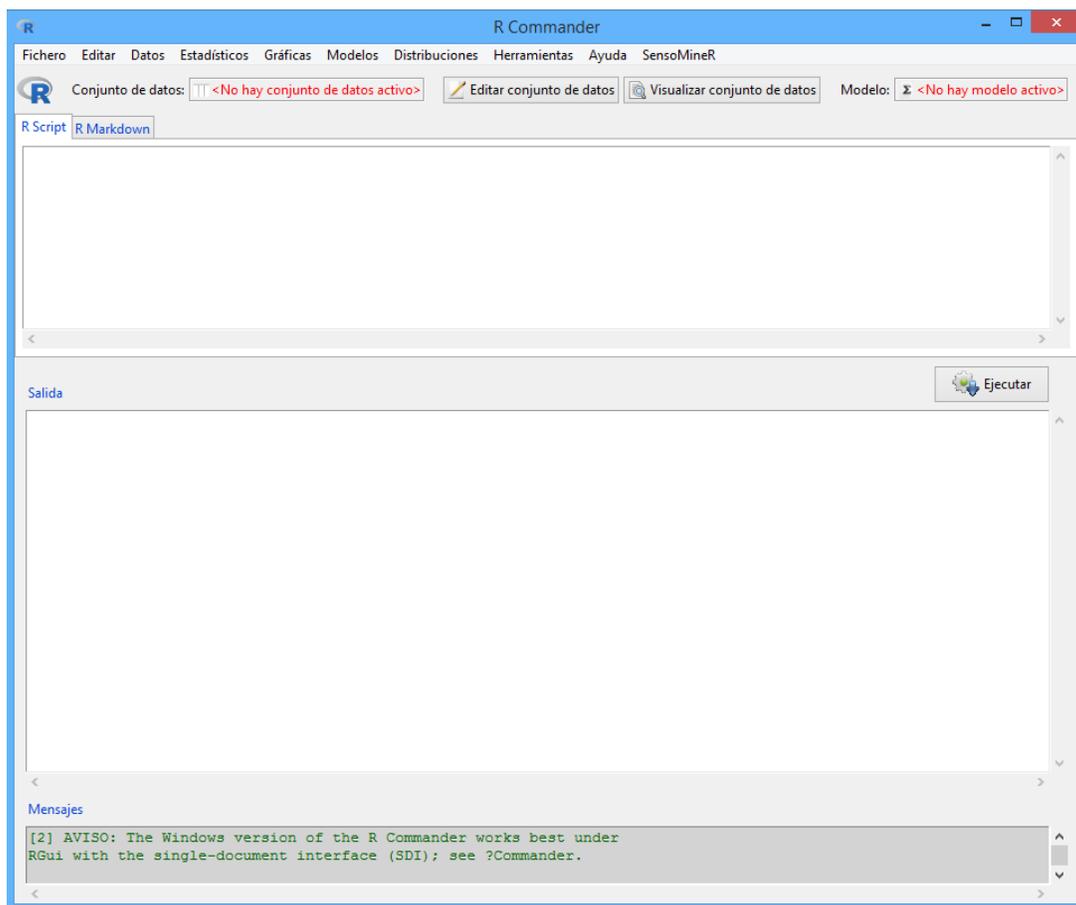


Figura 3.3.2.1. Interfaz de Commander

### 3.3.3 Weka

El paquete Weka contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, los cuales, están unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades. También soporta varias tareas estándar de minería de datos, clustering, clasificación, regresión, visualización y selección. Puede correr en casi cualquier plataforma, además contiene una extensa colección de técnicas para procesamiento de datos y modelado. Por desgracia, un área importante que actualmente no cubren los algoritmos incluidos en Weka es el modelado de secuencias.

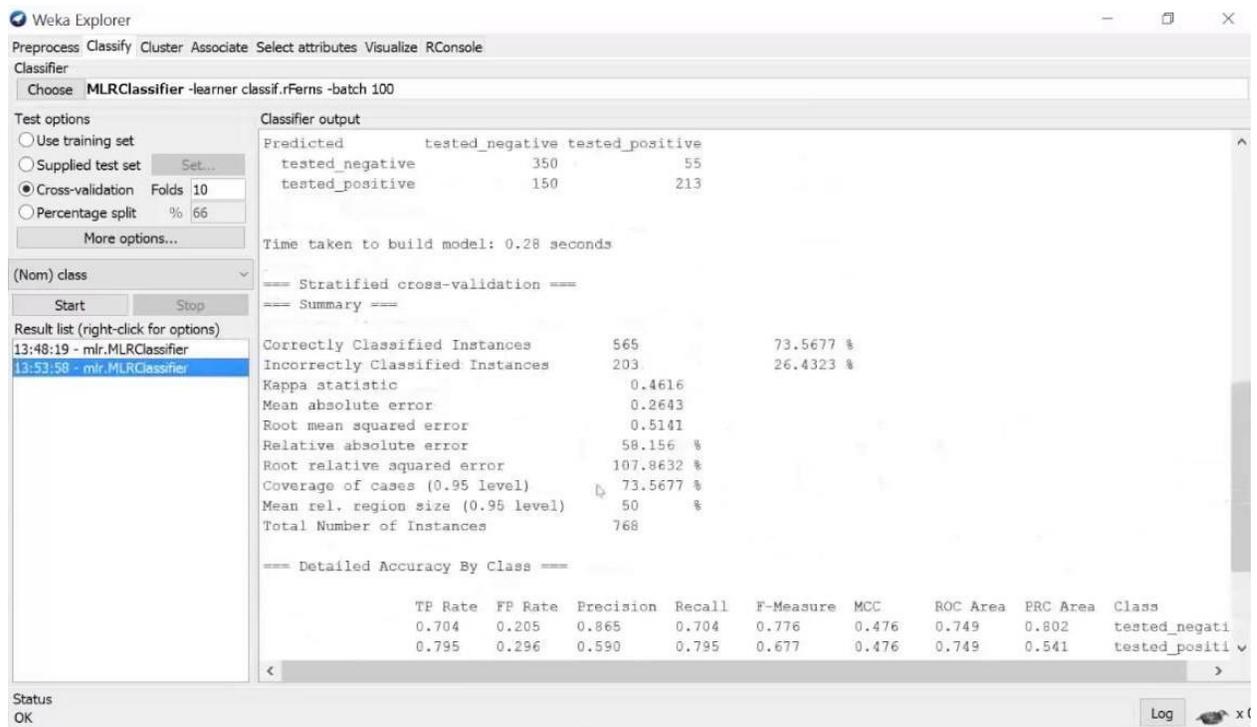


Figura 3.3.3.1. Interfaz de Weka

### 3.4. Librerías utilizadas

Una librería es un archivo o un conjunto de archivos codificados en un lenguaje de programación, los cuales facilitan la programación y ofrecen una interfaz bien definida para la funcionalidad para la que fueron programados.

### 3.4.1 FactoMineR

Métodos exploratorios de análisis de datos para resumir, visualizar y describir conjuntos de datos. Los principales métodos de componentes principales están disponibles, aquellos con el mayor potencial en términos de aplicaciones: análisis de componentes principales (PCA) cuando las variables son cuantitativas, análisis de correspondencia (CA) y análisis de correspondencia múltiple (MCA) cuando las variables son categóricas, análisis de factor múltiple cuando las variables se estructuran en grupos, etc. y en el análisis de agrupamiento jerárquico. F. Husson, S. Le y J. Pages (2017).

The screenshot shows the homepage of the FactoMineR website. At the top, there is a navigation menu with links: Home, Facto's Methods, Teaching MOOC, books, Graphic enhancements, Missing values missMDA, Automatic Reporting, Google group, and More. Below the menu is the large 'FACTOMINER' logo. The main content area is divided into two columns. The left column contains sections: 'About FactoMineR' (describing it as an R package for multivariate Exploratory Data Analysis), 'Why Use FactoMineR?' (a list of 8 features), and 'Home Menu' (a list of links: FactoMineR's description, News, Install FactoMineR, How to cite FactoMineR?, History of FactoMineR). The right column contains 'Authors' (listing François Husson, Julie Josse, and Sébastien Lê) and 'Useful Links' (listing R Project and CRAN). At the bottom, there is a footer with copyright information: © 2008 Design by: styleshout | Valid: XHTML | CSS, and links for Home and Sitemap.

Figura 3.4.1.1. Aspecto de su página web <http://factominer.free.fr/>.

### 3.4.2 *Factoextra*

Proporciona algunas funciones fáciles de usar para extraer y visualizar la salida de análisis de datos multivariados, incluyendo 'PCA' (Análisis de Componentes Principales), 'CA' (Análisis de Correspondencia), 'MCA' (Análisis de Correspondencia Múltiple), 'FAMD' (Análisis de Factor de Datos Mixtos), 'MFA' (Análisis de Factor Múltiple) y 'HMFA' (Análisis Jerárquico de Factor Múltiple). Contiene también funciones para simplificar algunos pasos de análisis de agrupamiento y proporciona una elegante visualización de datos basada en 'ggplot2'.

### 3.4.3 *Ggplot2*

Ggplot2 es una librería de visualización de datos para el lenguaje de programación estadística R. Fue presentada inicialmente por Hadley Wickham en 2005. Ggplot2 es una implementación de la Gramática de Gráficos Leland Wilkinson, un esquema general para la visualización de datos el cual divide los gráficos en componentes semánticos, como por ejemplo las escalas y otros elementos del gráfico. Ggplot puede servir como un reemplazo para los gráficos base en R y contiene una serie de valores predeterminados para la visualización web y de impresión de escalas comunes. Desde su lanzamiento en 2005, ggplot2 ha crecido en uso para convertirse en uno de los paquetes más populares de R.

# Capítulo 4

## 4.1 Conjunto de datos inicial

El análisis toma como insumo tres bases de datos, las cuales corresponden a los distritos de riego DR-063, DR-075 y DR-076, todos pertenecientes al estado de Sinaloa. La estructura de cada base de datos es tal, que los renglones hacen referencia al tipo de cultivo y las columnas corresponden a los años registrados. En cuanto a su tamaño, la dimensionalidad de cada distrito de riego aparece en la siguiente tabla:

**Tabla 4.1.1.** Dimensiones de los conjuntos de datos utilizados en el análisis

	<b>DR-063</b>	<b>DR-075</b>	<b>DR-076</b>
Número de cultivos inicial (Renglones)	65	87	66
Número de años incluidos en el estudio (Columnas)	48	48	46

De acuerdo con la Tabla 4.1.1, el distrito identificado con la clave DR-075 reportó 87 cultivos distintos, siendo éste el de mayor diversidad. En contraste, el DR-063 registró 65 tipos de cultivos diferentes. Además, podemos ver que todos los distritos incluyen de 46 a 48 años registrados. Tal como se recomienda en [Buuren & Groothuis-Oudshoorn \(2011\)](#), la primera tarea que realizamos consistió en una evaluación del número de celdas vacías en cada base de datos.

**Tabla 4.1.2.** Cantidad de datos registrados y valores faltantes

	<b>DR-063</b>	<b>DR-075</b>	<b>DR-076</b>
Cantidad total de celdas	3,055	4,089	2,970
Cantidad de celdas vacías	1,828	2,580	2,092

La Tabla 4.1.2 muestra la relación entre el total de celdas existentes en cada distrito de riego, en contraste con sus celdas vacías (aquellas que no guardan algún valor en su interior). Por ejemplo, las celdas vacías del DR-063 representan el 59% de sus datos totales, del mismo modo, las celdas vacías de los distritos DR-075 y DR-076 representan el 63% y 70% de los datos totales, respectivamente.

**Tabla 4.1.3.** Cantidad de cultivos removidos y cultivos utilizados en el análisis

	<b>DR-063</b>	<b>DR-075</b>	<b>DR-076</b>
Cultivos iniciales	65	87	66
Cultivos removidos en el estudio	49	71	57
Cultivos usados en el análisis	16	16	9
Cantidad de celdas vacías	67	33	36

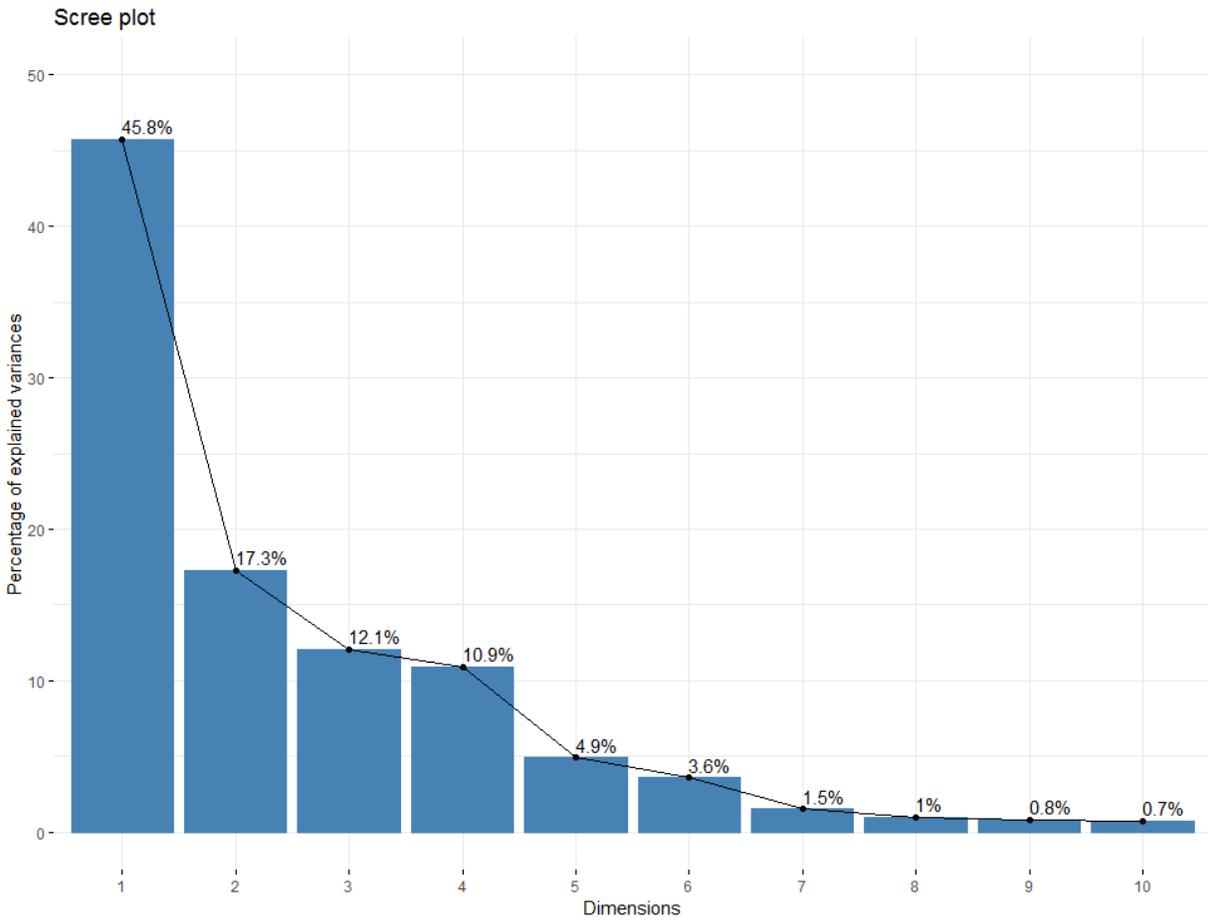
Esto representaba un problema, porque un requisito indispensable para la correcta aplicación de los métodos estadísticos multivariados es la inexistencia de celdas vacías. Por esta razón, se decidió conservar únicamente aquellos cultivos que tuvieran menos de quince celdas vacías. De esta forma, 49, 71 y 57 fueron la cantidad de cultivos removidos a los distritos DR-063, DR-075 y DR-076 respectivamente, tal y cómo se muestra en la Tabla 4.1.3. Nótese que una vez removidos tales cultivos (renglones) de los conjuntos de datos, todavía se tenía un cierto número de celdas vacías, aunque esta vez menor. Además, los conjuntos de datos se componen por dos tipos de celdas: la primera, es aquella que guarda en su interior el valor 0, que representa la pérdida total de la cosecha (la producción fue cero), mientras que la segunda, es una celda vacía.

La producción de cualquier cultivo está regida por la ley de la oferta y la demanda, por lo que una celda vacía representa la ausencia de demanda de ese cultivo. De esta forma, asignaremos a cada celda vacía un cero, ya que estas también representan la producción nula, al igual que las celdas con un cero en su interior, aunque por motivos distintos. Con lo anterior, los datos estaban listos para el análisis multivariado. Los resultados obtenidos aparecen a continuación.

#### *4.2 Resultados del análisis de correspondencias*

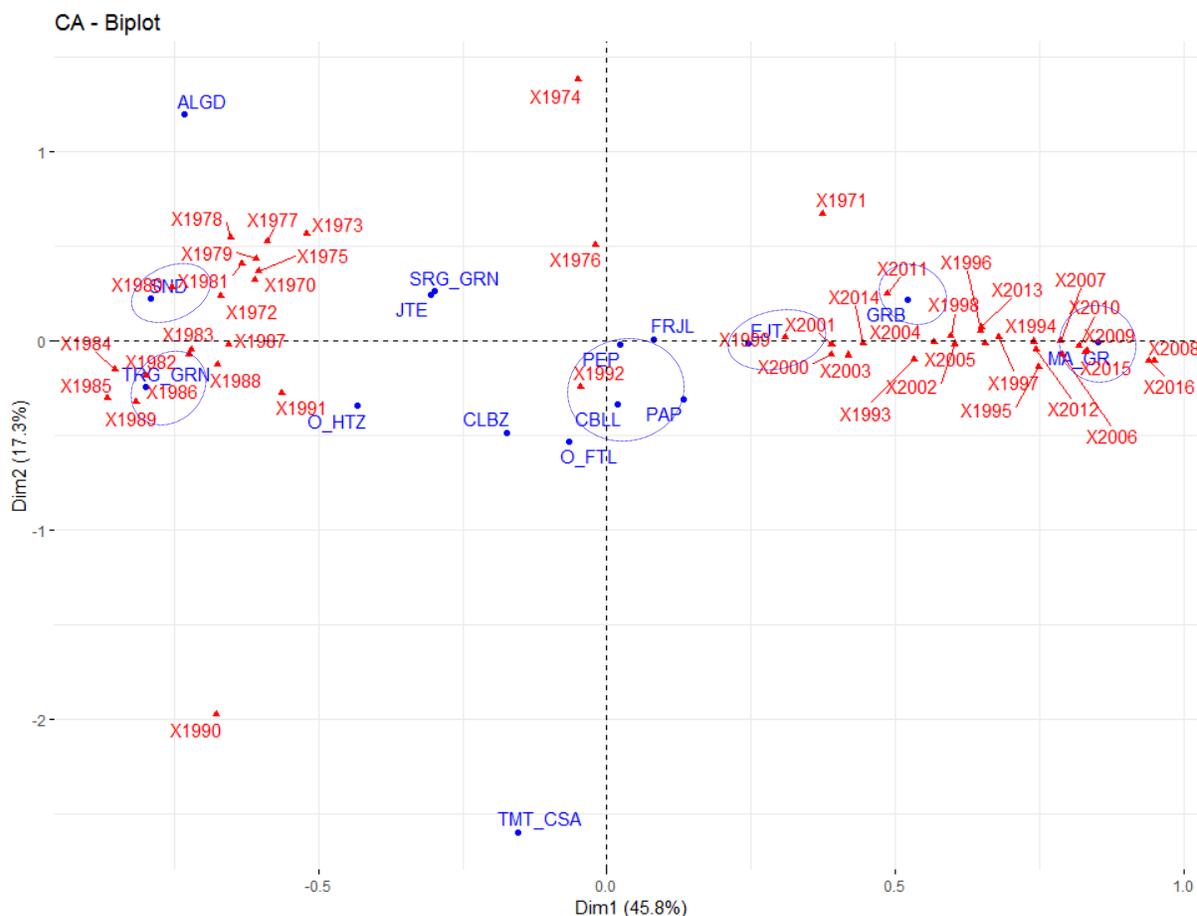
Se realizó un análisis de correspondencias a cada uno de los distritos de riego de acuerdo con el criterio sugerido por Cuadras (2014), además, se estudió la varianza contenida en cada uno de los componentes retenidos. De acuerdo con Kaiser (1974), dada una matriz de correlaciones  $\mathbf{S}$ , se deberán retener aquellos  $m$  primeros componentes, tal que  $\lambda_m \geq 1$  y que además  $v = tr(\mathbf{S})/p$  donde  $v$  es la medida de las varianzas. De esta manera, el punto de corte aconsejable es  $0.7 \times v$ . Es decir, las primeras  $m$  componentes, las cuales explican el 70% de la varianza.

Ahora se mostrarán los gráficos de sedimentación, los cuales muestran las componentes que más varianza retienen, así como las nubes de puntos obtenidos gracias a estas componentes. Cabe señalar que los términos que se encuentran cerca de los ejes coordenados (Dim1 y Dim2) en las nubes de puntos, hacen referencia a la primera y segunda componentes respectivamente, y los porcentajes que los acompañan, son el porcentaje de varianza que contienen.



**Figura 4.2.1.** Gráfico de sedimentación para la producción del distrito DR-063

En el gráfico mostrado por la Figura 4.2.1 podemos apreciar que las primera y segunda componentes contienen el 63% de la variabilidad de los datos analizados. Después de este resultado precedemos analizar la nube de puntos (bitplot) obtenida a través del análisis de correspondencias y que representa la producción en este distrito de riego.



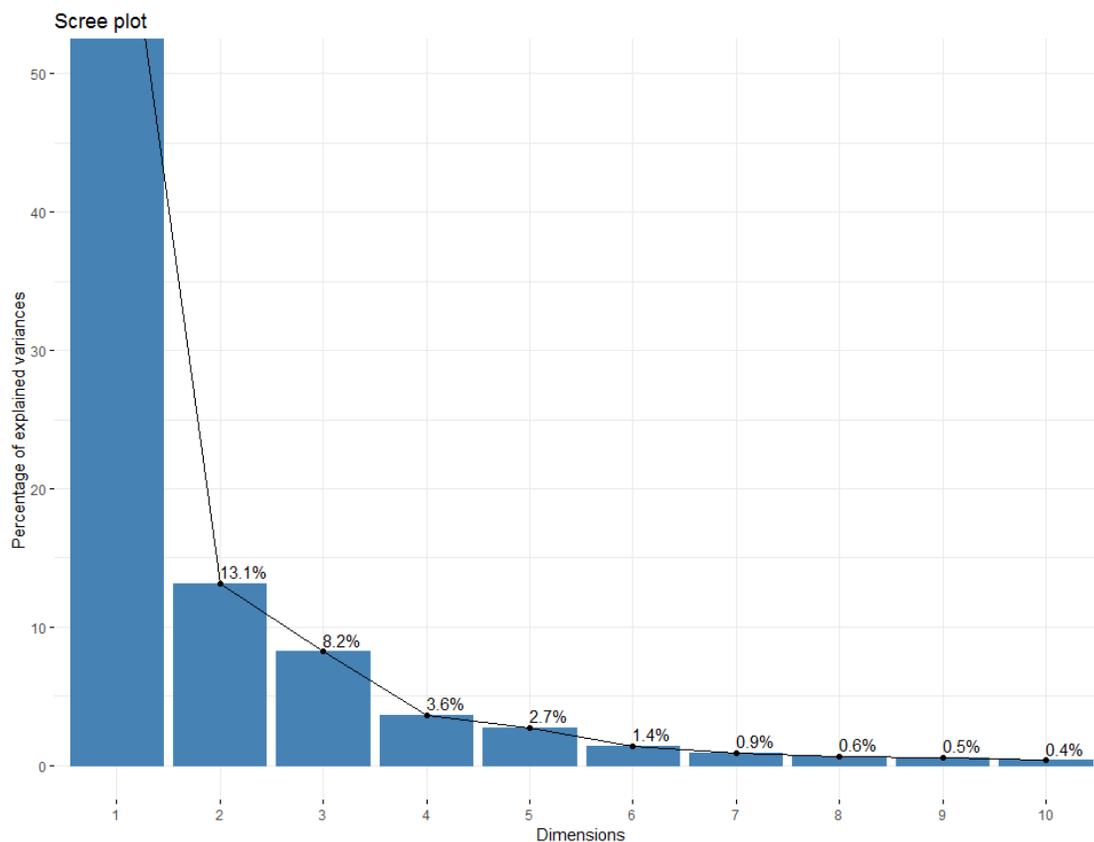
**Figura 4.2.2.** Nube de puntos para la producción de los principales cultivos del distrito DR-063

En la Figura 4.2.2 se tiene una aproximación de los puntos en un subespacio menor que en el que se encuentran, en este caso de orden dos. En esta figura podemos apreciar la relación que tiene la producción de ciertos cultivos con algunos años pertenecientes al registro. Por ejemplo, la producción de SND (Sandía) tiene una relación cercana con el año 1980. La producción de TRG\_GRN (Trigo grano) se encuentra relacionado con los años 1982, 1986 y 1989, mientras que la de EJT (Ejote) se encuentra vinculada con el año 1999. El PEP (Pepino) y CBL (Cebolla), ambos se encuentran ligeramente correlacionados con el año 1992. Por otra parte, el GRB (Garbanzo) tiene un vínculo cercano con el año 2011 y el MA\_GR (Maíz grano) se relaciona fuertemente con los años 2009, 2010 y 2015. Entre más cerca estén el cultivo y el año, la relación entre ellos será más fuerte, lo que significa que la producción de

ese cultivo en ese año fue predominante en comparación con la producción del resto de los cultivos.

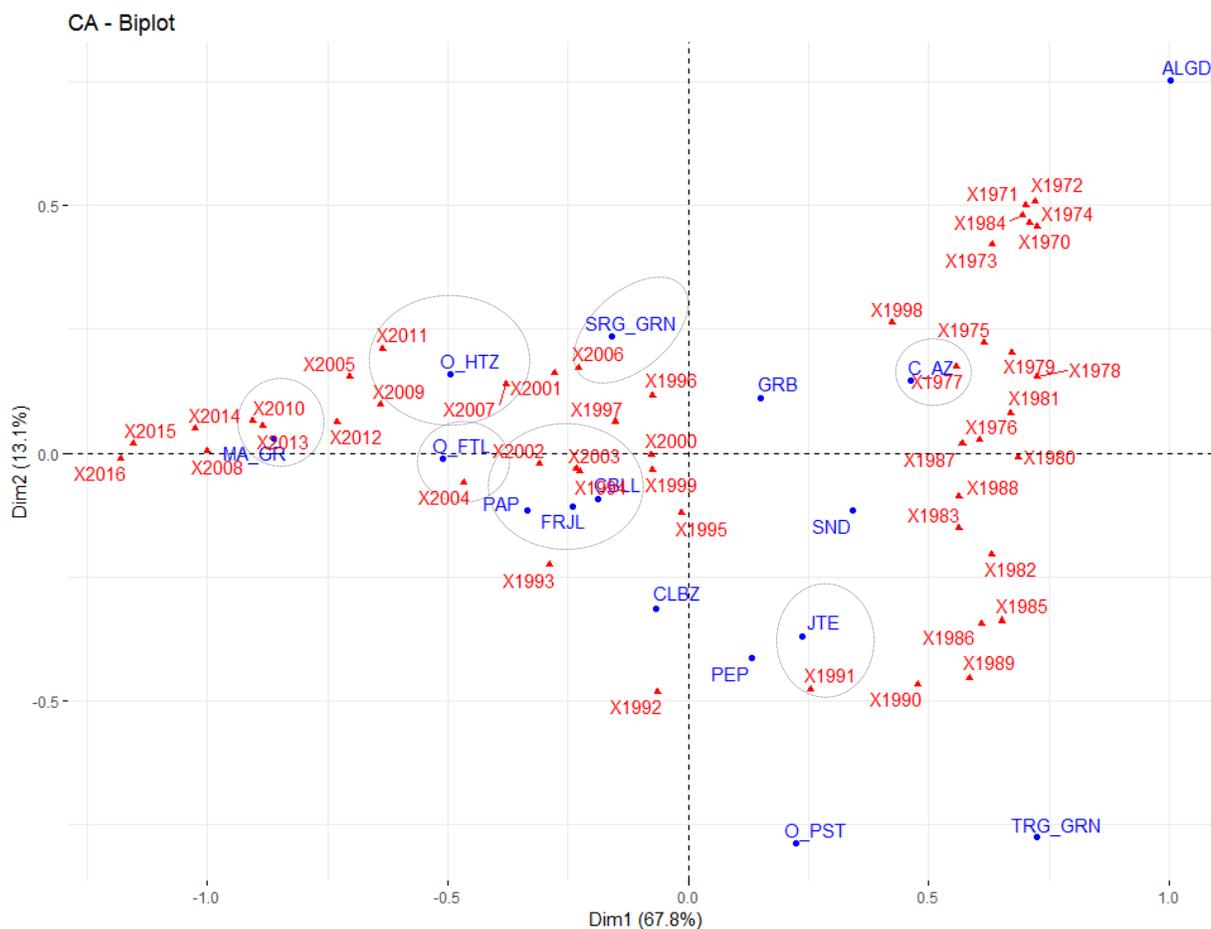
Es interesante hacer notar el siguiente comportamiento. Los cultivos de SND y TRG\_GRN están rodeados por una nube de años, los cuales van de 1970 a 1991. Mientras que los cultivos GRB y MA\_GR los rodea una nube de años que van desde 1993 hasta 2016. Esto nos dice que de 1970 a 1991 los cultivos que más solían producirse en este distrito eran la Sandía y el Trigo en grano, pero a partir de 1993 esto cambió, y los cultivos de Maíz y Garbanzo pasaron a tomar el primero y segundo lugar.

Por último, podemos apreciar cómo algunos cultivos no tienen una relación estrecha con algún año en específico, tal es el caso de TMT\_CSA (Tomate cáscara), ALGD (Algodón), SRG\_GRN (Sorgo grano), JTE (Jitomate), CLBZ (Calabaza), O\_FTL (Otros Frutales), PAP (Papa). Esto nos da a entender que estos cultivos no son representativos en este distrito de riego.



**Figura 4.2.3.** Gráfico de sedimentación para la producción del distrito DR-075

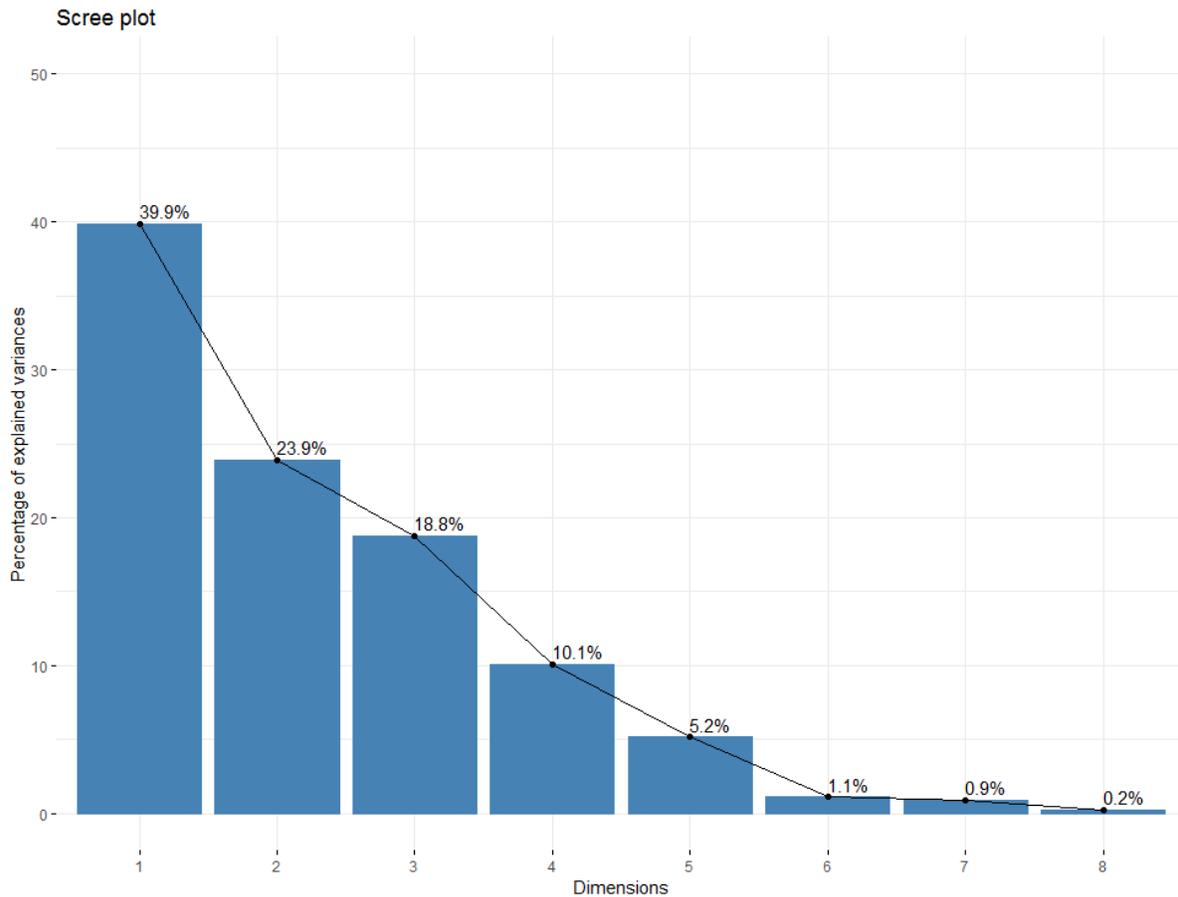
La Figura 4.2.3 muestra el gráfico de sedimentación para el análisis de componentes principales realizado a la producción del distrito DR-075. En este caso, las primeras dos componentes explican del 81% de la variabilidad y, por lo tanto, se cumple el criterio de [Kaiser \(1974\)](#). De forma similar, las primeras dos componentes serán utilizadas para generar la nube de puntos (Biplot) que se presenta a continuación.



**Figura 4.2.4.** Nube de puntos para la producción de los principales cultivos del distrito DR-075

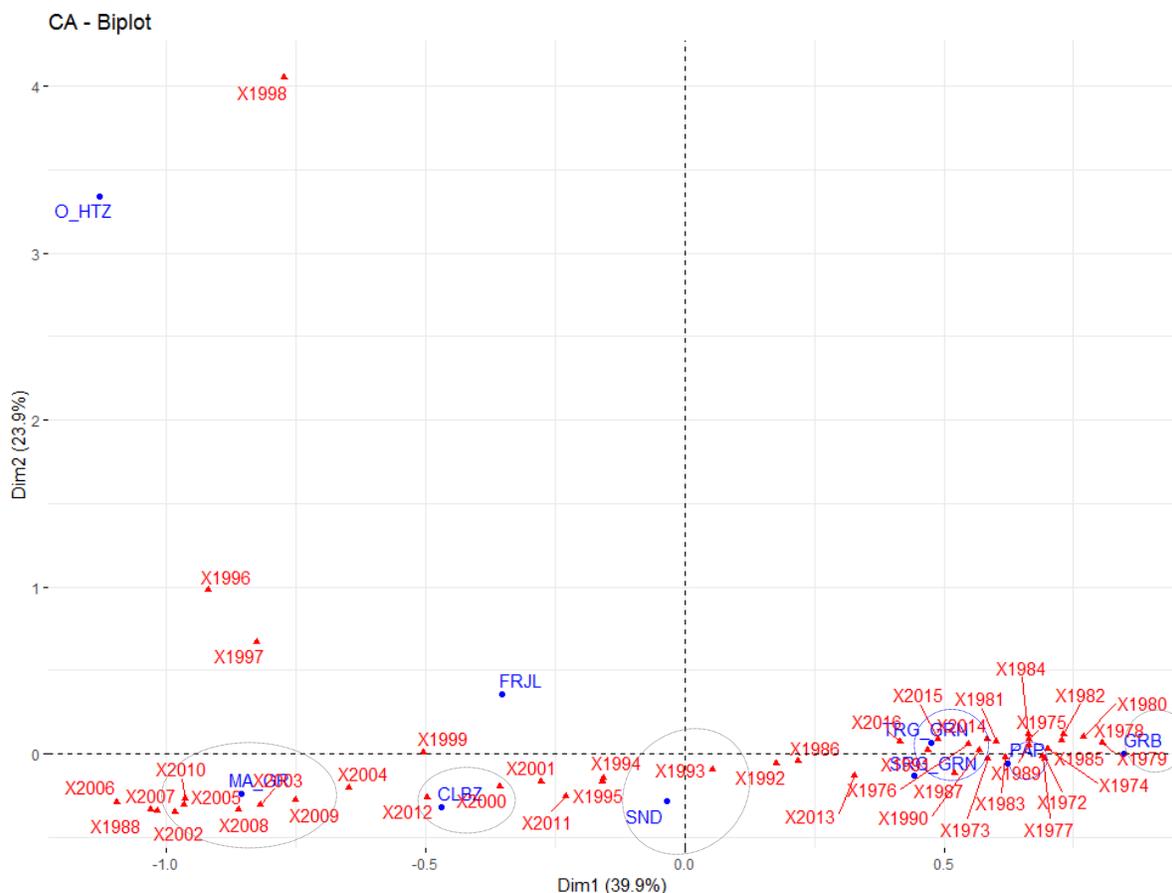
En la Figura 4.2.4 podemos ver como saltan a la vista algunas relaciones entre los cultivos y los años analizados. Entre estos, los cultivos de MA\_GR (Maíz grano) con los años 2010 y 2013. Las O\_HTZ (Otras hortalizas) tuvieron una producción ligeramente sobresaliente en los años 2007, 2009 y 2011. También podemos ver que O\_FTL (Otros frutales) tienen una pequeña correlación con el año 2004. Los cultivos de PAP (Papa), FRJL (Frijol) y CBLL (Cebolla) están prácticamente rodeados por los años 1993, 1994, 2002, 2003 y 2004. El SRG\_GRN (Sorgo grano) se encuentra vinculado con el año 2006. El JTE (Jitomate) se encuentra relacionado con el año 1991. Por último, podemos ver cómo la C\_AZ (Caña de azúcar) se encuentra cerca de diversos años, pero más cerca del año 1997.

Por último, veamos aquellos cultivos que no se relacionan con ningún año. ALGD (Algodón), GRB (Garbanzo), CLBZ (Calabaza), PEP (Pepino), SND (Sandía). Esto nos sugiere que estos cultivos no son importantes en este distrito de riego.



**Figura 4.2.5.** Gráfico de sedimentación para la producción del distrito DR-076

En la Figura 4.2.5 se muestra el gráfico de sedimentación para el distrito DR-076. Nótese que las primeras dos componentes concentran el 63% de la variabilidad. Con lo anterior se satisface el criterio de [Kaiser \(1974\)](#) y se procede a la nube de puntos en el plano factorial, tal como se explica a continuación.

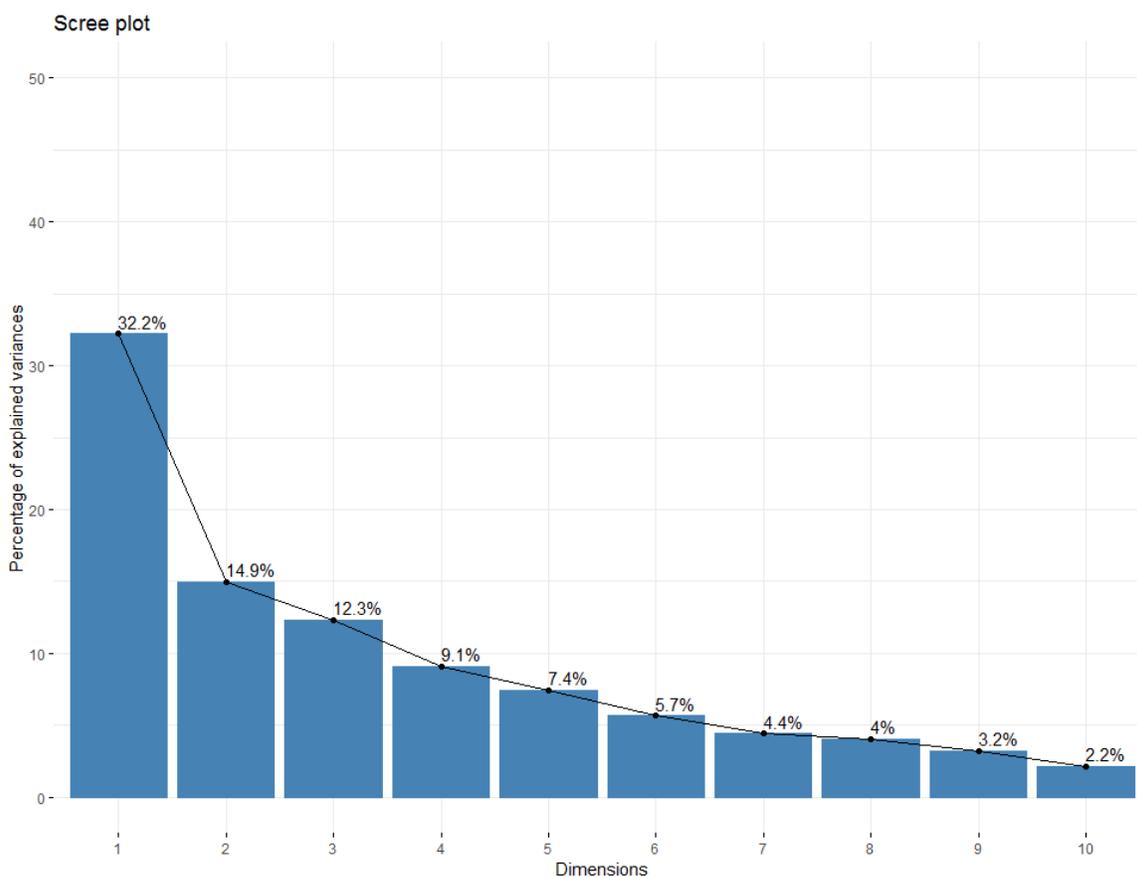


**Figura 4.2.6.** Nube de puntos para la producción de los principales cultivos del distrito DR-076

De acuerdo con la Figura 4.2.6 observamos que la producción de MA\_GR (Maíz grano) fue remarcable en los años 2002, 2003, 2005, 2007, 2008, 2009 y 2010. Así también, se resalta la conexión entre el cultivo de CLBZ (Calabaza) y el año 2012. La SND (Sandía) tuvo una pequeña predominancia en los años 1993, 1994 y 1995. Otra relación que llama la atención es la nube de puntos del lado derecho de la figura. En ella podemos encontrar al TR\_GRN (Trigo grano) con una relación muy estrecha con los años 1991 y 2015, el SRG\_GRN (Sorgo grano) con el año 1991, la PAP (Papa) tuvo una producción sobresaliente en 1983. La producción del GRB (Garbanzo) tiene una pequeña relación con el año 1979. Cabe resaltar que estos últimos cuatro cultivos están rodeados por una nube de años que van desde 1972 hasta 2016, dejándonos en claro que estos cultivos siempre han tenido presencia en este distrito. Por último, revisemos aquellos cultivos que no tienen una cercanía con algún año.

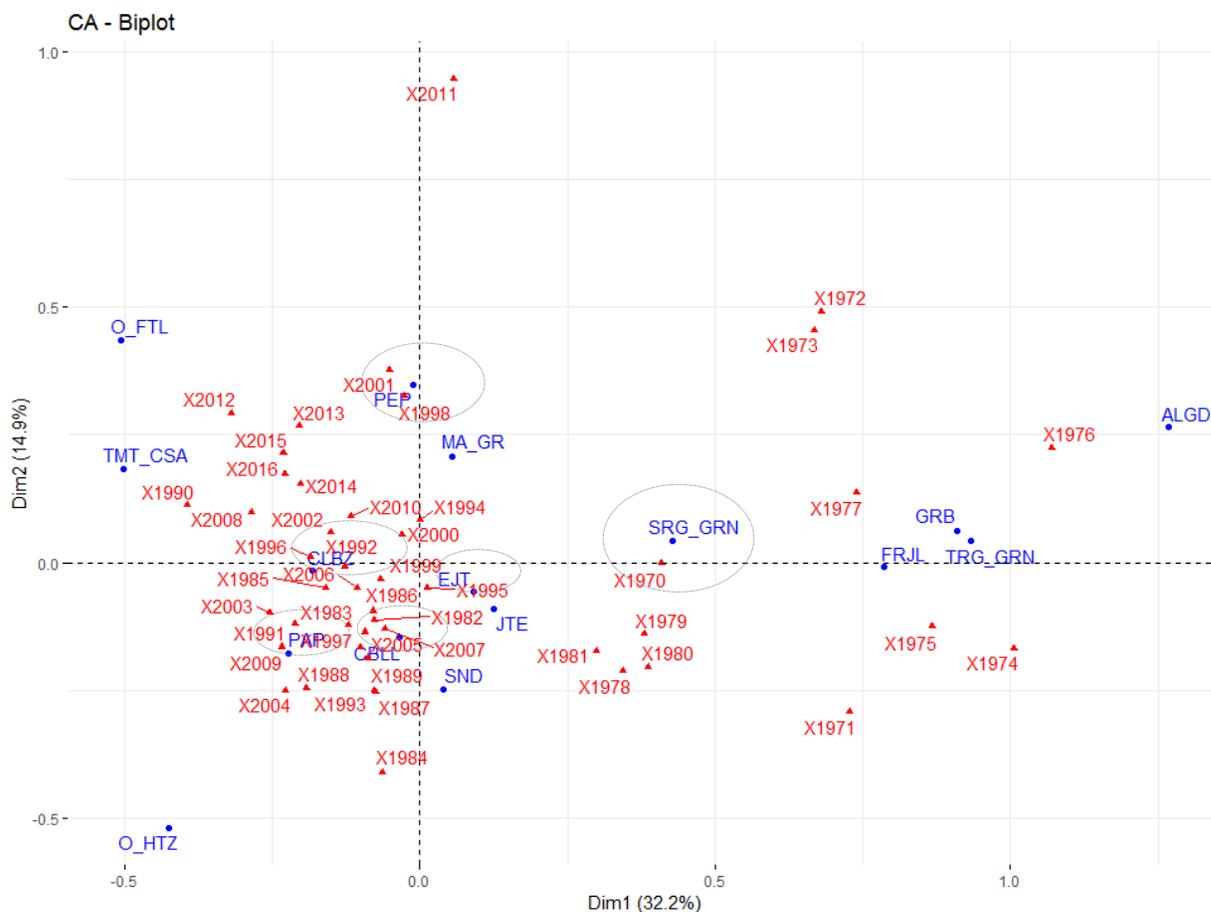
FRJL (Frijol) y O\_HLZ (Otras Hortalizas). Esto nos deja claro que la producción de estos cultivos no es sobresaliente en este distrito de riego.

En las siguientes páginas presentaremos los resultados obtenidos al analizar el rendimiento de los cultivos. Para fines prácticos de este trabajo, el rendimiento es calculado al hacer el cociente de la producción entre la cosecha. Así obtendremos una medida en términos de toneladas sobre hectárea. De este modo podremos saber en qué proporción aumentó o disminuyó la producción.



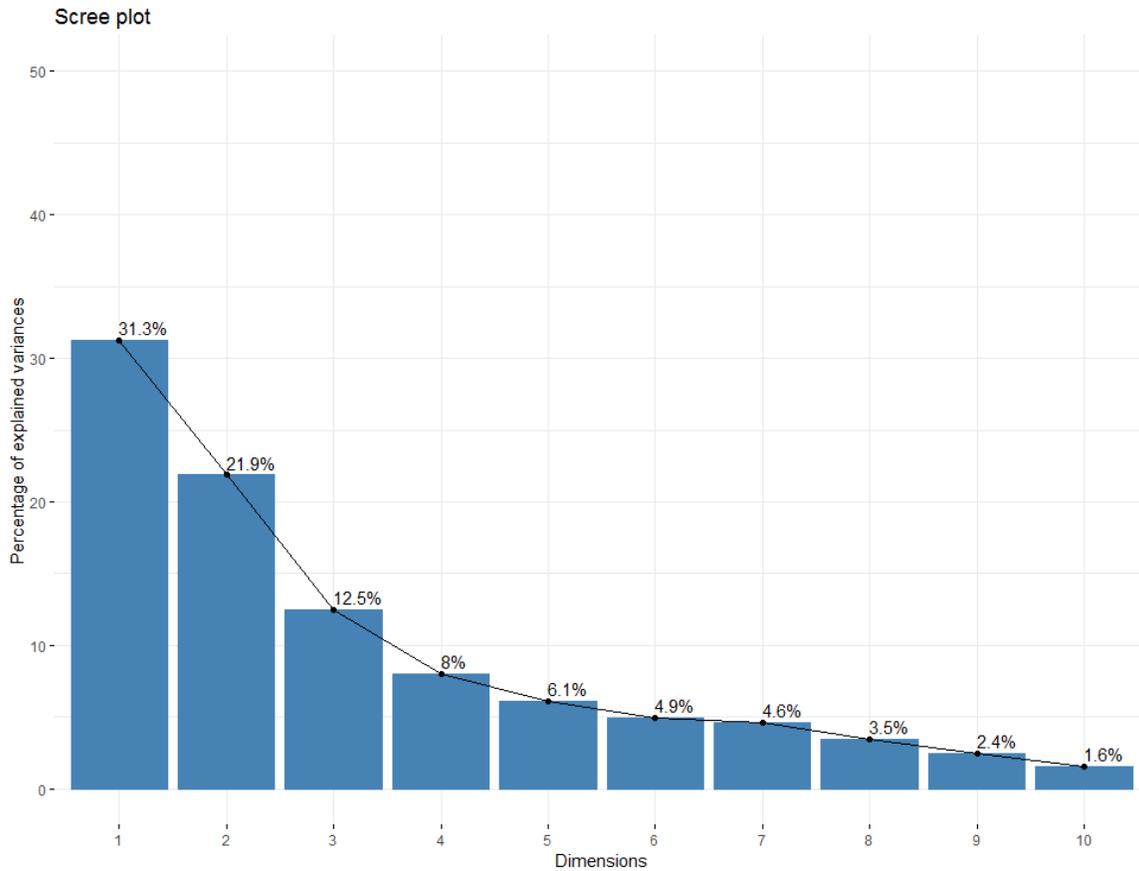
**Figura 4.2.7.** Gráfico de sedimentación para el rendimiento del distrito de riego DR-063

De acuerdo con la Figura 4.2.7, las primeras dos componentes concentran el 47% de la variabilidad. De manera que estaremos utilizando estas dos componentes para representar visualmente las relaciones entre los rendimientos de cada cultivo estudiado, por una parte, y los años de las cosechas, por la otra.



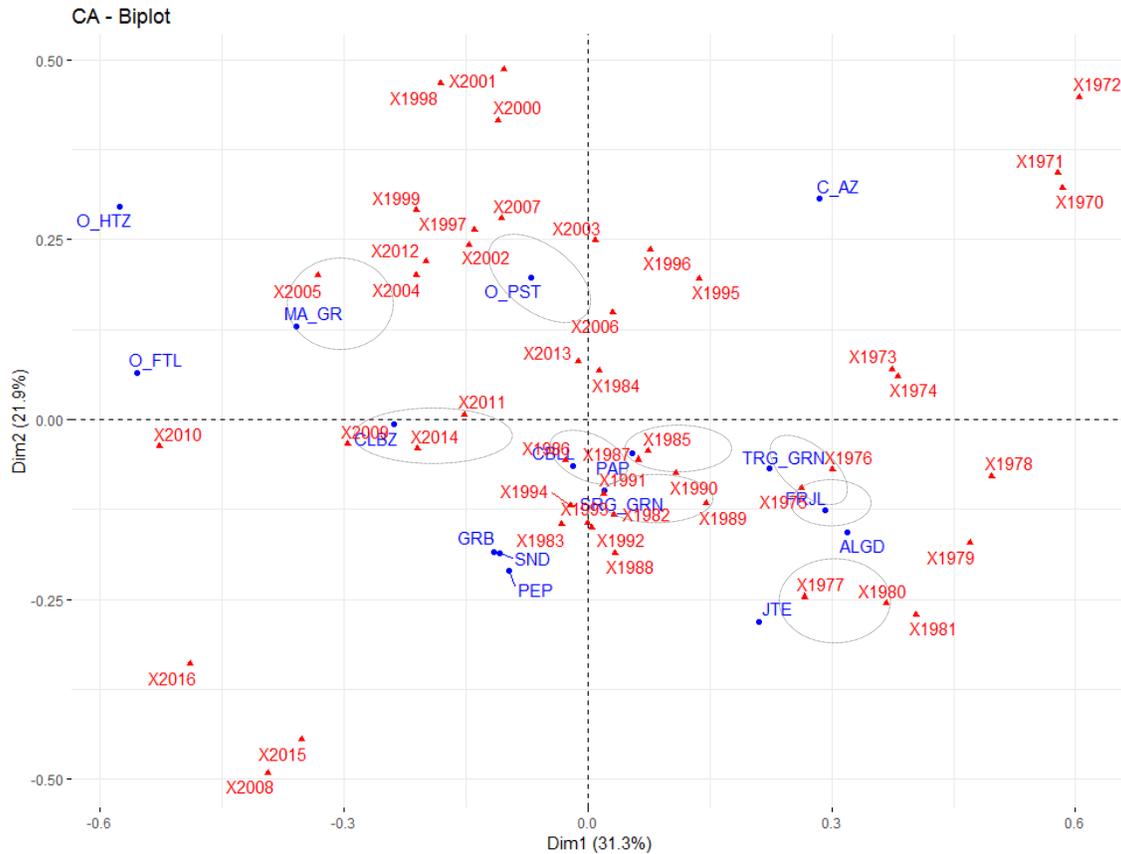
**Figura 4.2.8.** Nube de puntos para el rendimiento del distrito DR-063

Cómo podemos ver en la Figura 4.2.8 el rendimiento del PEP (Pepino) fue muy sobresaliente en los años 1998 y 2001. La CLBZ (Calabaza) se encuentra rodeado por los años 1985, 1992 y 1996. El cultivo de PAP (Papa) está muy próximo a los años 1991 y 2009. El rendimiento de la CBL (Cebolla) fue preponderante en el año 2007, mientras que el EJT (Ejote) presentó un ligero lazo con el año 1995. Por último, el SRG\_GRN (Sorgo grano) tiene una pequeña relación con el año 1970. Se resalta que el GRB, FRJL, TRG\_GRN no muestran relación preponderante con algún año en particular. De lo anterior se puede deducir que tales cultivos no son representativos para este distrito.



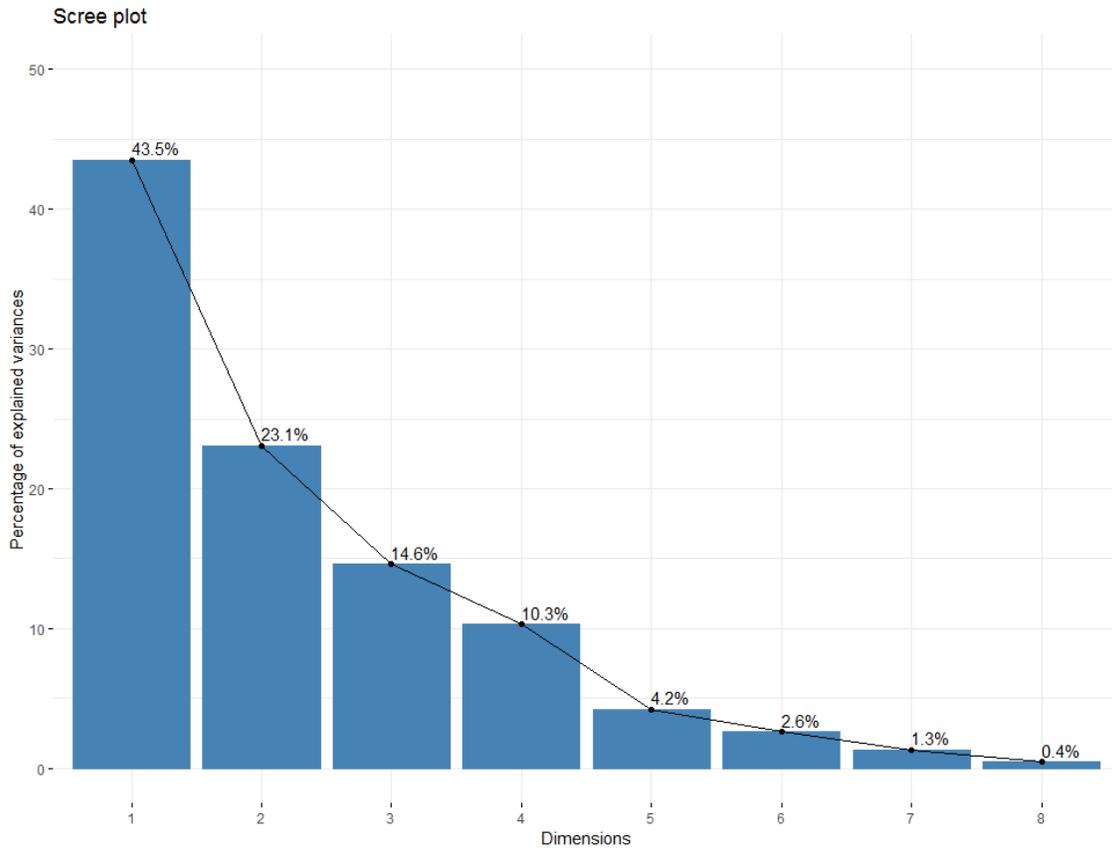
**Figura 4.2.9.** Gráfico de sedimentación para el rendimiento del distrito DR-075

De acuerdo con la Figura 4.2.9, los primeros dos vectores propios capturan el 53% de la variabilidad. De este modo podremos utilizarlos para ver de manera visual las relaciones que entre los rendimientos de los cultivos y los años registrados.



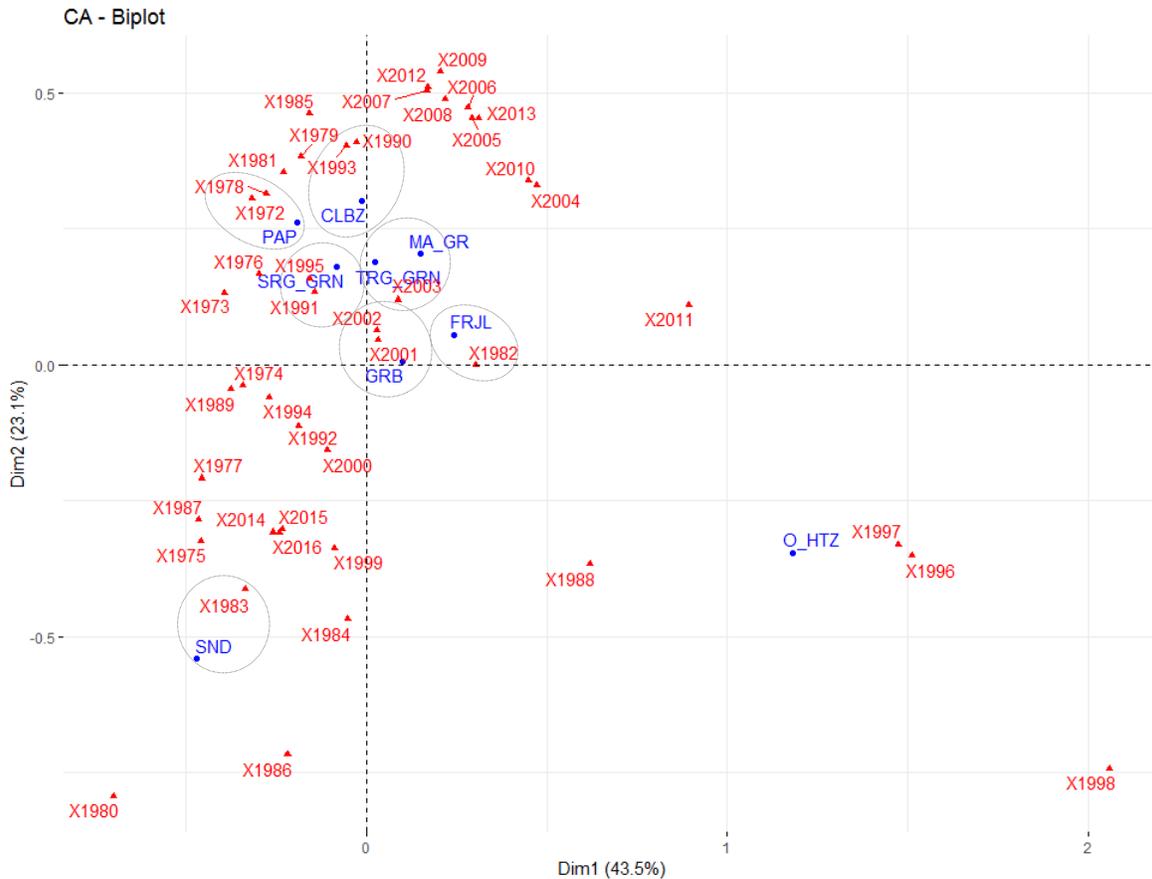
**Figura 4.2.10.** Nube de puntos para el rendimiento de los cultivos del distrito DR-075

La Figura 4.2.10 nos muestra la proximidad que tiene el rendimiento de los cultivos seleccionados con los años registrados. El MA\_GR (Maíz grano) se encuentra ligeramente cerca del año 2005. La CLBZ (Calabaza) está muy próxima a los años 2009 y 2014. En la parte central se encuentran tres cultivos, la CBLZ (Cebolla) la cual está muy ligada al año 1986, la PAP (Papa) que se encuentra muy cerca de los años 1985 y 1987, así como el SRG\_GRN (Sorgo grano), el cual está prácticamente encima del año 1991. Finalmente, del lado derecho tenemos el rendimiento de los cultivos de TRG\_GRN (Trigo grano) y FRJL (Frijol), los cuales están ligeramente enlazados al año 1975.



**Figura 4.2.11.** Gráfico de sedimentación para el rendimiento del distrito DR-076

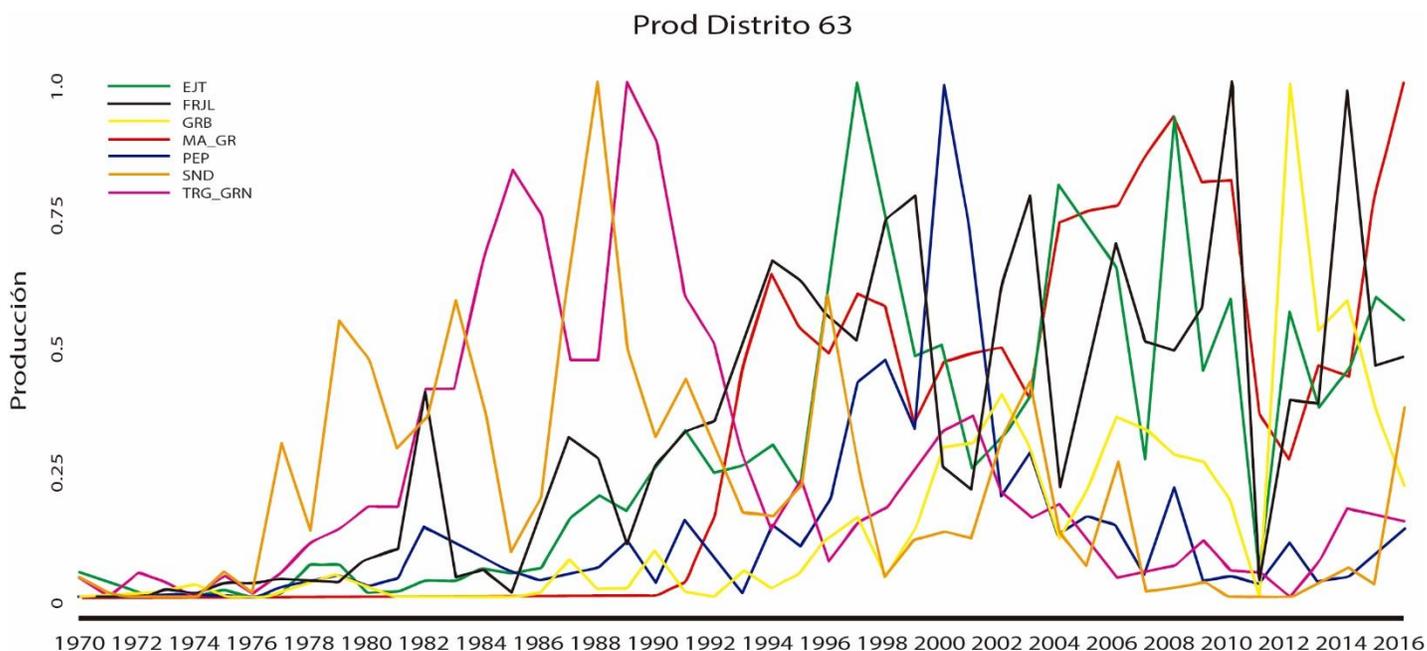
De acuerdo con la Figura 4.2.11, los primeros dos vectores propios capturan el 66% de la variabilidad. De modo que, si utilizamos esas dos primeras componentes para representar visualmente las relaciones entre el rendimiento de cada cultivo con los años del registro, podremos tener una muy buena aproximación del fenómeno real.



**Figura 4.2.12.** Nube de puntos para el rendimiento del distrito DR-076

La Figura 4.2.12 muestra qué tan estrecha es la relación entre los rendimientos de los cultivos analizados y los años registrados. Podemos apreciar la ligera relación entre la CLBZ (Calabaza) con los años 1990 y 1993, así como la PAP (Papa) y los años 1972 y 1978. El rendimiento del SRG\_GRN muestra una cercanía con los años 1995 y 1991. El GRB (Garbanzo) está próximo a los años 2001 y 2002. Por último, podemos ver la pequeña relación que entre el rendimiento del FRJL (Frijol) y el año 1982.

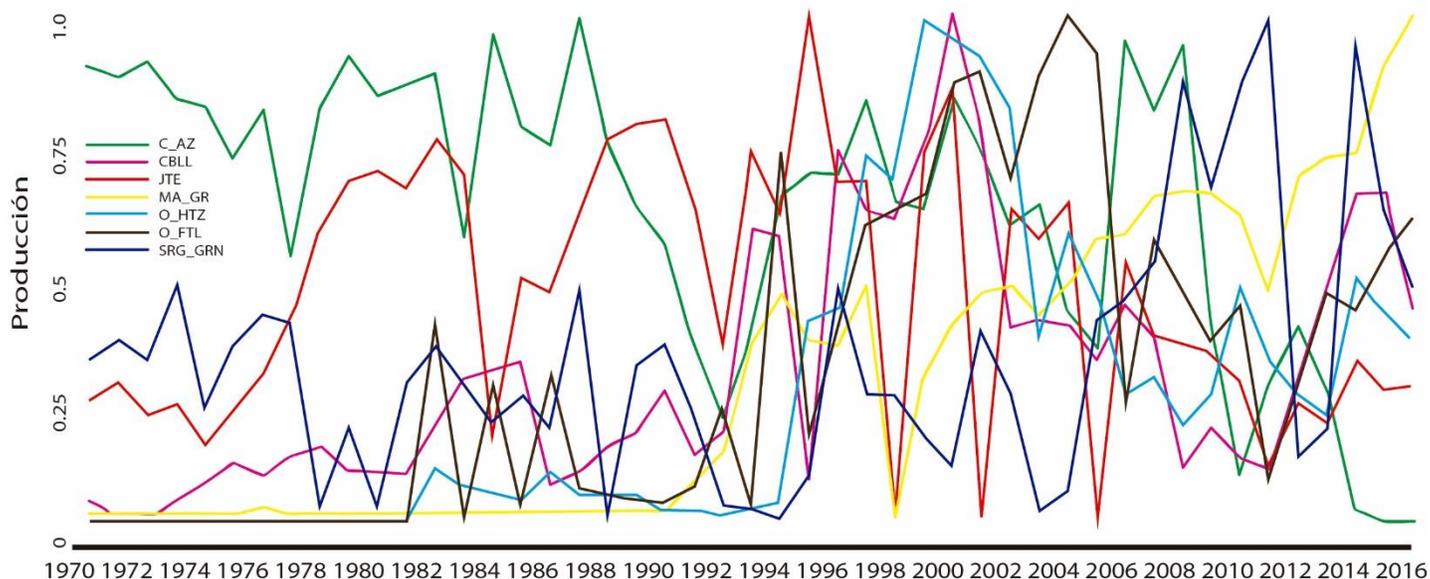
### 4.3 Análisis de series de tiempo



**Figura 4.3.1.** Serie de tiempo para la producción de los principales cultivos del distrito DR-063

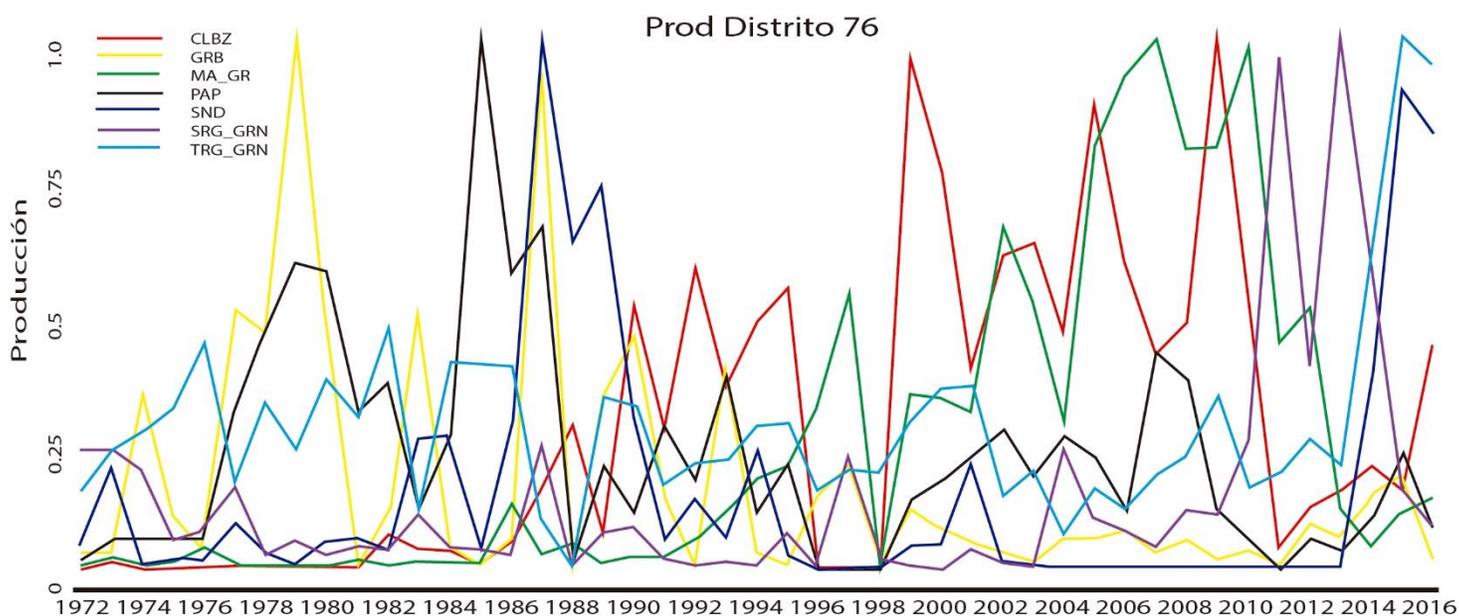
En la Figura 4.3.1 se muestra la serie de tiempo para la producción del distrito 63. En total se incluye la producción de siete tipos de cultivos a lo largo del periodo comprendido de 1970 a 2016. El comportamiento de la producción de maíz de grano (MA\_GRN) es interesante. Durante los años 1970 a 1991, la producción era casi igual a cero y después del año 1992 creció de tal forma que en el 2016 tuvo su máximo histórico con más de 850 mil toneladas producidas. En el año 1988, la sandía (SND) registró una producción récord la cual no se ha repetido desde entonces. En 1989 el trigo en grano (TRG\_GRN) tuvo su máxima producción y desde entonces ha venido disminuyendo. El caso del frijol (FRJL) llama mucho la atención. En el 2010 y 2014 las producciones fueron las más altas que hayan tenido, con 49 mil y 48 mil toneladas respectivamente. El ejote (EJT), el garbanzo (GRB) y el pepino (PEP) tuvieron sus máximos en los años 1997, 2000 y 2012 respectivamente. Cabe también señalar que en el año 2011 los cultivos de GRB, EJT y FRJL tuvieron su mínimo histórico. De esta forma, a través de las series de tiempo podemos identificar como ha sido la evolución de la producción de los cultivos más importantes de este distrito.

### Prod Distrito 75



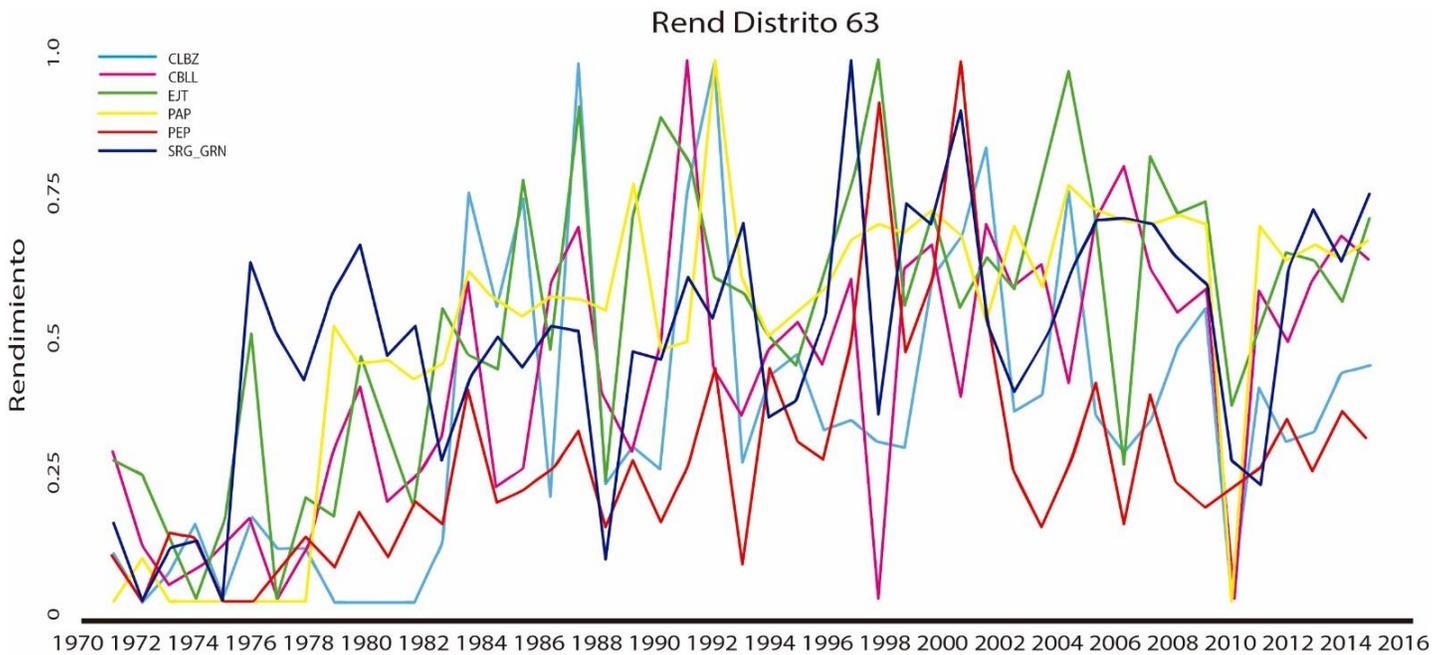
**Figura 4.3.2.** Serie de tiempo para la producción de los principales cultivos del distrito DR-075

La Figura 4.3.2 nos muestra cómo los cultivos de caña de azúcar (C\_AZ), sorgo en grano (SRG\_GRN), jitomate (JTE) y cebolla (C BLL) siempre han estado presentes en la producción de este distrito. La C\_AZ es el cultivo de mayor notoriedad, ya que en 8 ocasiones registró márgenes muy grandes de producción, mucho más que el resto de los cultivos, dando a entender que este distrito es azucarero. Sin embargo, en los años 2015 y 2016 su producción se redujo a cero. El JTE también ha tenido una buena producción hasta que en el año 1995 registró la mayor producción que ha tenido desde entonces, aun así, en los años 1998, 2001 y 2005 la producción fue cero. Por otro lado, el maíz de grano (MA\_GR) tuvo un periodo (de 1970 a 1990) en el que su producción era muy baja, pero a partir de 1991 fue aumentando hasta conseguir su máximo histórico en 2016 teniendo un solo un año de producción nula (1998). Otras hortalizas (O\_HTZ), C BLL, otros frutales (O\_FTL) y el sorgo de grano (SRG\_GRN) registraron sus producciones más grandes en 1999, 2000, 2004 y 2011 respectivamente.



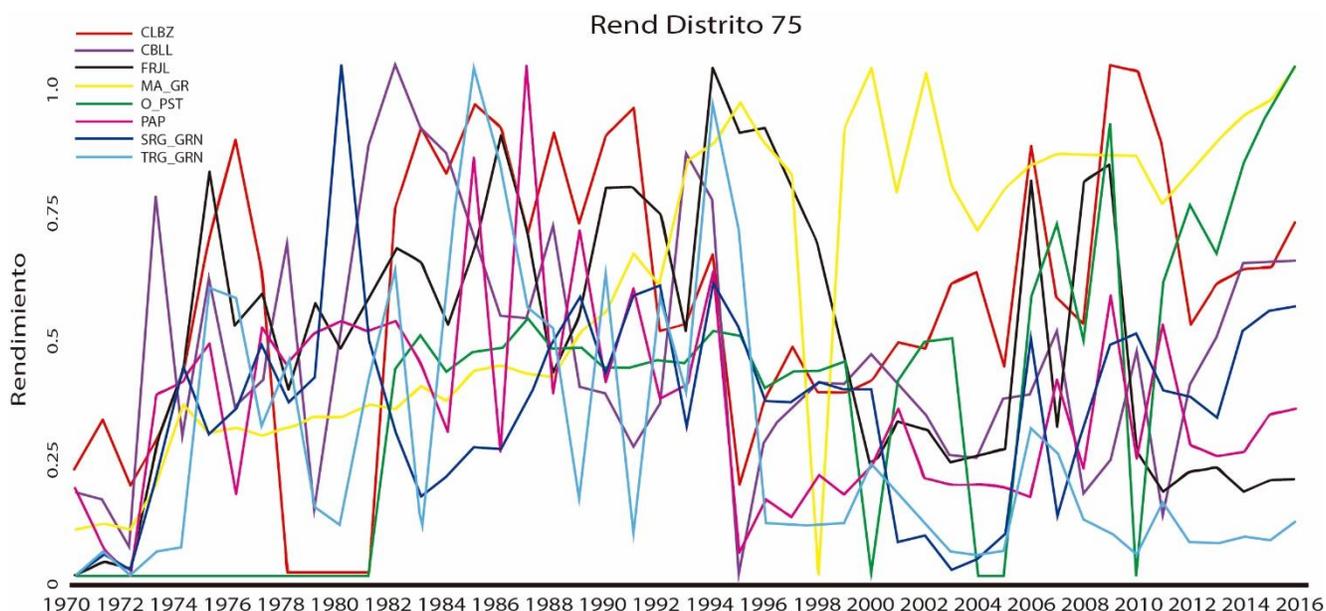
**Figura 4.3.3.** Serie de tiempo para la producción de los principales cultivos del distrito DR-076

La Figura 4.3.3 nos muestra el comportamiento de la producción de los cultivos más destacados del distrito 75 en los últimos cuarenta años. En ella se puede apreciar como algunos cultivos consiguieron una enorme producción en más de una ocasión. Por ejemplo, el garbanzo (GRB) tuvo dos años maravillosos: 1979 y 1987. La calabaza (CBLZ) tuvo tres años de gran producción: 1999, 2005 y 2009. El ascenso del maíz de grano (MA\_GR) comenzó en 1991, antes de eso, su producción era casi cero, hasta que en los años 2007 y 2010 consiguió las dos más grandes producciones de las que se tiene registro. La sandía tiene dos años de enorme producción: 1987 y 2015, sin embargo, de 2004 a 2013 su producción fue cero. En los años 2011 y 2013, el sorgo en grano (SRG\_GRN) consiguió cifras históricas. El trigo en grano (TRG\_GRN) es el único cultivo que siempre ha estado presente, a excepción de 1988, donde su producción fue nula, sin embargo, consiguió una producción histórica en el 2015.



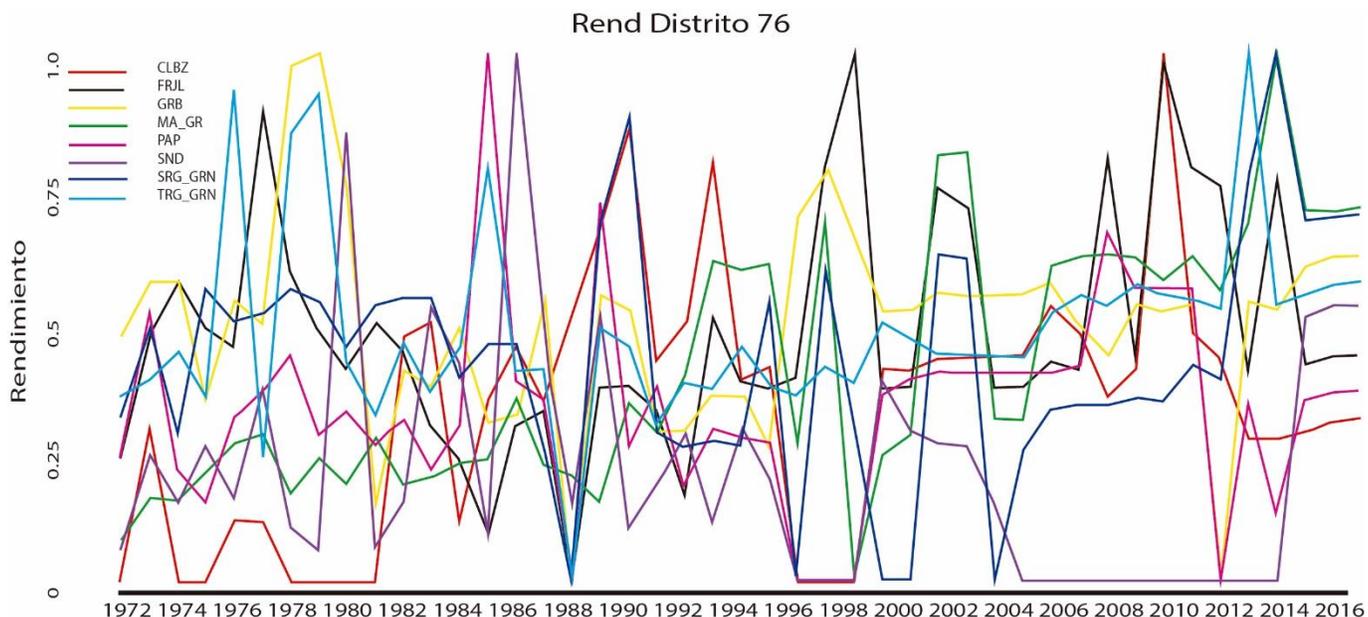
**Figura 4.3.4.** Serie de tiempo para el rendimiento de los principales cultivos del distrito DR-063

La Figura 4.3.4 nos muestra cómo el rendimiento de cada cultivo ha ido cambiando a través de los años. De 1978 a 1981, la calabaza (CLBZ) tuvo un rendimiento nulo, desde entonces se ha mantenido positivo, llegando en dos ocasiones (1987 y 1992) a un estupendo rendimiento. La cebolla (CBL) sólo ha conseguido un gran margen de rendimiento en una ocasión (1991). El ejote (EJT) muestra dos años de estupendo rendimiento (1998 y 2004). Mientras que la papa (PAP), el pepino (PEP) y el sorgo en grano (SRG\_GRN) tuvieron su mejor rendimiento en los años 1992, 1995 y 2002 respectivamente. Además, algunos años en particular son interesantes, por ejemplo: en el 2011 el rendimiento de tres cultivos fue nulo (CLBZ, CBL y PAP) mientras que los rendimientos del EJT y el SRG\_GRN disminuyeron. En 1988 los rendimientos de todos los cultivos disminuyeron.



**Figura 4.3.5.** Serie de tiempo para el rendimiento de los principales cultivos del distrito DR-075

Si miramos la Figura 4.3.5 podemos ver cómo ha evolucionado el rendimiento de los cultivos en el distrito 75 en los últimos cuarenta y cinco años. La calabaza (CLBZ) ha tenido un rendimiento positivo en todo momento, con excepción de un periodo de cuatro años (de 1978 a 1981). La cebolla tuvo su máximo rendimiento en 1982. 1994 fue el año de mayor rendimiento para el frijol (FRJL). El rendimiento del maíz ha ido en aumento en los últimos 45 años, salvo por el año 1998, donde su rendimiento fue cero. De 1970 a 1981, el rendimiento de otros pastos (O\_PST) fue nula, sin embargo, desde entonces ha ido en aumento, salvo por cuatro ocasiones en las que el rendimiento volvió a caer a cero (2000, 2004, 2005 y 2010). La papa (PAP) alcanza su máximo rendimiento en el año 1987 y el único rendimiento nulo es en 1972. En el resto de los años se mantiene positiva. El sorgo en grano (SRG\_GRN) consiguió un rendimiento máximo en 1980. El trigo en grano (TRG\_GRN) alcanzó su máximo en el año 1985.



**Figura 4.3.6.** Serie de tiempo para el rendimiento de los principales cultivos del distrito DR-076

La Figura 4.3.6 nos muestra cómo la calabaza (CLBZ) alcanza su rendimiento más grande en el 2009 y en al menos diez años tuvo un rendimiento nulo. El frijol (FRJL) tuvo su mayor rendimiento en 1998, mientras que el garbanzo (GRB) lo tuvo en 1979. El maíz en grano registró el mayor rendimiento que ha tenido en 2013 y la papa (PAP) lo hizo en 1985, además de que en cinco años tuvo un rendimiento de cero. La sandía (SND) en consigue su mayor rendimiento en 1986, además de reportar en trece ocasiones un rendimiento de cero, y el sorgo en grano (SRG\_GRN) consiguió su rendimiento más grande en 2013, sin embargo, reportó cinco años con rendimiento nulo. Mientras tanto, el trigo en grano registra su mayor rendimiento en 2012. El año 1988 es de especial interés, debido a que cinco de los cultivos registraron un crecimiento nulo.

#### *4.4 Contrastando con la realidad*

En las figuras 4.3.1 y 4.3.4, las cuales presentan las líneas del tiempo de producción y rendimiento (respectivamente) del distrito de riego 63, el año 2011 es muy relevante, debido a que, en ese año, el rendimiento y la producción de algunos cultivos cayó estrepitosamente. De acuerdo con [Domínguez \(2016\)](#) la razón de esto fue que en ese año hubo una gran sequía en el norte del país, la cual llegó a afectar este distrito. Además, [Ortega \(2013\)](#) nos comenta que el área en donde se encuentra este distrito de riego tiene una precipitación media anual menor al promedio nacional.

# Capítulo 5

## 5.1 Conclusiones

El análisis multivariante fue el protagonista de este trabajo, y entre las ventajas que tiene es la posibilidad de tener de entrada información de múltiples variables, aunque estas no sean linealmente dependientes. También permite trabajar con matrices que tengan más variables que observaciones y con matrices incompletas (siempre que los valores faltantes estén aleatoriamente distribuidos y no superen un 10% de la base de datos original) [Cuadras \(2014\)](#).

Para concluir este trabajo, se compararán los valores arrojados por los tres modelos (Holt-Winters, SES y ARIMA) con ayuda de la raíz del error cuadrático medio (RMSE). El RMSE es una medida de las diferencias entre los valores predichos por un modelo o un estimador y los valores observados, y la forma de calcular este valor es mediante la siguiente fórmula:

$$\sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}} \quad (5.1.1)$$

Donde  $\hat{y}_t$  es el valor estimado,  $y_t$  el valor observado y  $T$  es el número de observaciones. Con ayuda de (5.1.1) se pudo calcular el valor del RMSE para cada uno de los métodos de pronóstico que se utilizaron para analizar los cultivos.

**Tabla 5.1.2.** Medidas de la raíz del error cuadrático medio para cada modelo

Cultivo	Distrito	Holt-Winters	Suavizado Exponencial	ARIMA
		RMSE	RMSE	RMSE
MA_GR	63p	0.1346104	0.1317338	0.1301478
TR_GR	63p	0.1399581	0.1328533	0.1264592
CBLL	63r	0.246532	0.1842603	0.2006857
CLBZ	63r	0.275522	0.2450059	0.2451957
CBLL	75p	0.1831958	0.1726099	0.1657858
C_AZ	75p	0.2005172	0.19822	0.1968042
CLBZ	75r	0.2466918	0.2320917	0.2247952
SR_GR	75r	0.1785466	0.171336	0.1700442
TR_GR	76p	0.1609187	0.1531656	0.1484613
GRBZ	76p	0.2357749	0.2671406	0.2278659
FRJL	76r	0.2512798	0.2117819	0.2018669
TR_GR	76r	0.1999288	0.1817188	0.1904361

La Tabla 5.1.2 contiene dichos valores, y con ayuda de ella, podemos notar que, en la mayoría de los casos, los valores más bajos (color rojo) del RMSE los tiene el modelo ARIMA, mostrando que el mejor ajuste lo tiene este modelo, por lo que se concluye lo siguiente: se le recomienda a cualquier investigador usar el modelo ARIMA si lo que desea es calcular valores futuros de una base de datos que contenga información histórica sobre la producción y el rendimiento de cultivos.

### 5.2 Futuras líneas de investigación

Siendo más minuciosos, pueden encontrarse más patrones de los que se han hablado hasta ahora, los cuales podrían llevar a nuevas investigaciones, por ejemplo: las seis gráficas que muestran el rendimiento y la producción de los cultivos más importantes de cada distrito de riego están escaladas de 0 a 1, donde 0 representará al valor más pequeño y 1 al valor más grande que se hayan registrado. De este modo cada serie de tiempo alcanza el 1 en algún año, lo que significaría que, en ese año, la producción o el rendimiento de ese cultivo alcanzó su máximo histórico. Por lo que sería interesante saber las razones que llevaron a ese máximo.

## Referencias

Badii, M. H., and J. Castillo (2017). *Análisis de correlación canónica (ACC) e investigación científica*. Innovaciones de Negocios 8.

Betancourt, D. F. (15 de febrero de 2016). *Cómo usar la suavización exponencial simple para pronosticar la demanda*. Recuperado el 21 de octubre de 2019, de Ingenio Empresa.

Box, George & Jenkins, Gwilym (1970) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). mice: *Multivariate imputation by chained equations in R*. Journal of statistical software, 1-68.

Chatfield, Chris, and Mohammad Yar (1988). *Holt-Winters forecasting: some practical issues*. Journal of the Royal Statistical Society: Series D (The Statistician) 37.2 (1988): 129-140.

CRAN (2018), Página principal, Proyecto R. Tomado de: <https://cran.r-project.org/>, 02 diciembre 2018

Cuadras, Carles M. (2014) *Nuevos métodos de análisis multivariante*. Barcelona: CMC Editions, 2007.

Domínguez, J. (septiembre-octubre, 2016). *Revisión histórica de las sequías en México: de la explicación divina a la incorporación de la ciencia*. Tecnología y Ciencias del Agua, 7(5), 77-93.

Ferrando, P. J., & Anguiano-Carrasco, C. (2010). *El análisis factorial como técnica de investigación en psicología*. Papeles del psicólogo, 31(1), 18-33.

Fuente Fernández, Santiago (2016). *Series Temporales: Modelo Arima*. UAM Facultad Ciencias Económicas y Empresariales, Madrid-España.

Ggplot2 (2018), Visión general. Tomado de <https://ggplot2.tidyverse.org/>, 20 de noviembre 2018

González, J. A., Alonso, P. B., & Gallego, A. G. (2008). *Análisis de correspondencias y estudio de historias de vida: una aplicación a la encuesta de transición educativo-formativa e inserción laboral*. Pecvnia: Revista de la Facultad de Ciencias Económicas y Empresariales, Universidad de León, (6), 1-27.

Greenacre, M. (2008). *La práctica del análisis de correspondencias*. Fundacion BBVA.

Hyndman, Rob J., and George Athanasopoulos (2018). *Forecasting: principles and practice*. OTexts.

Kaiser, H. F. (1974). *An index of factorial simplicity*. Psychometrika, 39(1), 31-36.

Leyva Morales, José Belisario, et al (2014). *Uso de plaguicidas en un valle agrícola tecnificado en el noroeste de México*. Revista internacional de contaminación ambiental 30.3: 247-261.

Mira, L. L., Martínez, A.T., & Cruz, D. L. (2018) *Aplicación de Holt-Winters para pronósticos de inventarios*. CIENCIA UANL / AÑO 21, No.90 julio-agosto 2018

Nau, Robert (2017). "ARIMA models for time series forecasting." Statistical Forecasting: Notes on Regression and Time Series Analysis. Duke University, Durham.

Ortega-Gaucin, David, and Velasco, Israel. (2013) *Aspectos socioeconómicos y ambientales de las sequías en México*. Aqua-LAC - Vol. 5 - Nº 2 - Set. 2013. pp. 78 - 90

Rivero, Omar Maguiña (2016) *El Método de Pronóstico Holt-Winters*

Trubetskoy, Gregory (2016), *Holt-Winters Forecasting for Dummies*, Notes to self

Wei, William (2006) *Time series analysis univariate and multivariate methods*, Pearson Education

Wikipedia (2018), RStudio. Tomado de: <https://es.wikipedia.org/wiki/RStudio>,  
21 noviembre 2018

Wikipedia (2018), Weka. Tomado de:

[https://es.wikipedia.org/wiki/Weka\\_\(aprendizaje\\_autom%C3%A1tico\)](https://es.wikipedia.org/wiki/Weka_(aprendizaje_autom%C3%A1tico)). 21 noviembre 2018

Wikipedia (2018), Rcmdr. Tomado de: [https://es.wikipedia.org/wiki/R\\_Commander](https://es.wikipedia.org/wiki/R_Commander).

21 noviembre 2018

Wold, Herman (1938) *A study in the analysis of stationary time series*. Diss. Almqvist & Wiksell

Wolfgang, Karl H. & Simar, Léopold (2015) *Applied Multivariate Statistic Analysis*, Springer



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS



## INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS



Control Escolar de Licenciatura

### VOTOS DE APROBATORIOS

Presidente del Consejo Directivo del Instituto de Investigación en Ciencias Básicas Aplicadas de la Universidad Autónoma del Estado de Morelos.  
P r e s e n t e .

Por medio de la presente le informamos que después de revisar la versión escrita de la tesis que realizó el C. **GIL SALGADO FRANCISCO XAVIER** con número de matrícula **20134003938** cuyo título es:

#### “Análisis estadístico de series de tiempo de bases de datos agrícolas”

Consideramos que **SI** reúne los méritos que son necesarios para continuar los trámites para obtener el título de **Licenciado Ciencias Área Terminal en Matemáticas**.

Cuernavaca, Mor a 21 de septiembre del 2020

Atentamente  
Por una universidad culta

Se adiciona página con la e-firma UAEM de los siguientes:

DR. ANTONIO DANIEL RIVERA LÓPEZ  
DR. RAÚL SALGADO GARCÍA  
DR. JESUS IGOR HEBERTO BARAHONA TORRES  
DRA LORENA DÍAZ GONZÁLEZ  
DR. PAULO CESAR MANRIQUE MIRON

PRESIDENTE  
SECRETARIO  
VOCAL  
PRIMER SUPLENTE  
SEGUNDO SUPLENTE (NO PARTICIPA)



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

### Sello electrónico

**LORENA DIAZ GONZALEZ** | Fecha:2020-10-12 13:50:32 | Firmante

bvMv9CymC4S81T9IbznhlzfO2pFxcj4mKbGu2WHWJ38NQDxU+fdGvysL93IFtut7cKwvuYtpZRboq6Z0j9IMFSVOxP7nonXeAOu3mbmX6plk5r+6/B5arh59CUGM4wTm71Zj33VXoDSZwvGfs5nCUmA1Fm1iwozM9p7PCzt6eXjeVvryWTgu4xo6hCOfgAklqruy34ZK0bV0DII/+m2TBwhSkgid3cFOHJYmZFLzQazYo3adZgZMH7AVUtPpe7C9e7cXf5ahRpGfMF1FJYQy1o5wF8EkkUs7SdbE3zBp49Ib72NUR5AfvoSdqv1BplZ64glfQ2IZ+VkULOKWxJQ==

**RAUL SALGADO GARCIA** | Fecha:2020-10-13 08:22:22 | Firmante

hOQduATbQxu4DS8rlazlzHU0+Ns+3sNHNIftQtzUbZPKhnoVfpJDD9+NqF7AexO/6qTdnA24MPOj606ejmduQOG6P4iu0vxfE1SYPIVcmEZIACm2HdeTmkkmFDdetCWD8qTUuGcFDJ7yUdl9uYokZt3IPtqcRmo26AoyjD7c0IDFGg2agPV7B2K4C6PA9oGd9VdMgRCxxPGxsVIGYk1mwPs39UzBHP9rWKPp+YVT/ISO0WrtRp1IdQnNd2WbsMX2Q1wE4370hJ2bxAMyapeFi5zya0dJqMQQat3vrOhmlmIHGqcM+axwGtZdum6UkzthjtizUYkFMmVSPWB7N8EqlvA==

**JESUS IGOR HEBERTO BARAHONA TORRES** | Fecha:2020-10-14 22:31:09 | Firmante

hlx5T9cW7d5PjPewLH7aZfV+UQ7gWCqACgGBayjcwndOYkKJBh2rf1z4/Vg4obDyrniFXWyiUohMedEb0I+iYLGVoXUdPyn8H01n6p2eEhFusmQA7fbMf1uiJQUUilAGRmEVYPYiCJ6ZzvLfscO6BPLGR93HDGNYtTHHe0/ucjrwX5vNVvvlmFhnUmLLOoKhMssCW2xhTnbTc4ovHK3HPj+W11k/OyJK5mJBM3BjGWTG6gJucZNXlmFVDQ/QiLwHP0dYWhFf3MVgOzh8qNKg0u0AjqptnZq4oYJ3zippLE1n0H7xELMCPDPpWaHcPm3dwlVtWVLSqHQwbuUr2Q==

**ANTONIO DANIEL RIVERA LOPEZ** | Fecha:2020-10-15 11:34:19 | Firmante

IcuZhl4XCOjPKTNkejMD3/xR3qfNPSGaCwFwQvFK2KOovkY1cX0OdFvpzeveCARQAIeAhb7LksoKEmDjNHry8h0Y37IYCanCDjwZ1dD/+FevK7hDI1aTphvkifDgLSZSe37/UxokoSl79sIUQ4WcWsb1JXswFKUoUwnApKX+5GDfQZ0t1tURDhp53KK0p3zeE2OnRohyC/OxKiGai4EOpzeGf3HgsSxhdjc7/tpc6x4uqtW6F9I/vt+nQac+8Q4XFk1OYKYGP+cMnrulZk/Cg4DsW9HvpL/AphijYkOoPywv7mxjPgTu0wQT5EECbsffBIGtB3SNgA8kOQPh5nTyA==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



sLJ6Xg

<https://efirma.uaem.mx/noRepudio/nEKDORGFdMNvZD2U5Tih0esMuN8gqagR>

