

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**POSGRADO EN CIENCIAS MATEMÁTICAS Y DE LA  
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA**

***CONSTRUCCIÓN DE UN INDICADOR DE DESEMPEÑO DE LOS  
OPERADORES DEL SERVICIO DE AGUA POTABLE EN MÉXICO***

TESINA

PARA OBTENER EL TITULO DE  
**ESPECIALISTA EN ESTADÍSTICA APLICADA**

PRESENTA

**Maribel Adriana Caballero Castrillo**

BAJO LA ASESORÍA DE

**M. en E. Leticia Eugenia Gracia-Medrano Valdelamar**

México D.F., Noviembre de 2015



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

*A Amalia*

*Porque tu llegada ha cambiado mi vida*

# CONSTRUCCIÓN DE UN INDICADOR DE DESEMPEÑO DE LOS OPERADORES DEL SERVICIO DE AGUA POTABLE EN MÉXICO

## CONTENIDO

AGRADECIMIENTOS.....	5
INTRODUCCIÓN .....	6
CAPITULO 1	
El Análisis de Componentes Principales .....	9
1.1 Componentes Principales.....	9
1.1.1 El primer componente principal.....	10
1.1.2 El segundo componente principal.....	10
1.1.3 Selección del número de componentes.....	11
CAPITULO 2	
El Análisis de Conglomerados .....	13
2.1 Medidas de proximidad .....	13
2.1.1 Disimilaridad y medidas de distancia para datos continuos .....	14
2.2 Medidas de proximidad intra-grupos.....	14
2.3 Conglomerados Jerárquicos.....	15
2.3.1 Métodos aglomerativos .....	16
2.3.2 Calculo de distancias .....	16
2.4 Conglomerados no Jerárquicos.....	17
CAPITULO 3	
El Análisis Discriminante .....	19
3.1 Análisis discriminante.....	19
3.1.1 Función lineal discriminante .....	19
3.1.2 Regla de máxima verosimilitud .....	20
CAPITULO 4	
Aplicaciones .....	21
4.1 Construcción del índice de eficiencia .....	21
4.2 Análisis Descriptivo.....	25

4.2.1 Análisis Gráfico .....	27
4.3 Análisis de las varianzas y correlación de las variables del índice.....	29
4.4 Tratamiento de los datos de las variables del índice .....	31
4.5 Relación entre variables .....	31
4.6 Análisis de componentes principales .....	35
4.6.1 El índice de desempeño .....	37
4.7 Análisis de conglomerados .....	40
4.8 Análisis discriminante .....	46
4.8.1 Análisis discriminante con clasificación de Ward.....	48
4.8.2 Discriminante con clasificación de K-Mean.....	50
4.9 Análisis de eficiencia de las empresas de agua en México .....	52
4.9.1 Modelo de eficiencia de las empresas .....	52
4.10 Resultados .....	53
4.10.1 Diagnóstico del Modelo.....	55
CONCLUSIONES .....	58
BIBLIOGRAFIA.....	61
ANEXOS .....	63

## **AGRADECIMIENTOS**

Mi más sincero agradecimiento a los profesores sinodales Dra. Silvia Ruiz-Velasco Acosta, Dra. Ruth Selene Fuentes García, M. en E. Patricia Isabel Romero Mares, Act. Miguel Ángel Chong Rodríguez y de manera especial a la M. en E. Leticia Eugenia Gracia-Medrano Valdelamar. Por la guía en la elaboración del presente trabajo y el apoyo y paciencia en este largo proceso.

Quiero agradecer a la Universidad Nacional Autónoma de México UNAM, que a través de la Facultad de Economía y del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas IIMAS, han contribuido en mi formación académica.

# INTRODUCCIÓN

---

En México el servicio de agua potable y alcantarillado se realiza a través de una concesión, ya sea estatal, municipal o local, la cual es conocida como Organismo Operador (OO). La actividad económica que realiza no está regulada por ninguna Secretaría Federal, Estatal ni por la ciudadanía, lo que incentiva a los directivos de estos organismos a trabajar con ineficiencias administrativas. Lo anterior provoca: deudas económicas con entidades bancarias, letargo municipal o local ya que se utilizan recursos económicos etiquetados para otra actividad y baja calidad del servicio de agua. Esta última condición podría ser la causante de enfermedades gastrointestinales en adultos o hasta la muerte en niños.

Por otro lado, este tipo de manejos anárquicos también puede provocar el rezago de la localidad, llevando a la empresa del agua y en algunos casos a la localidad a una insostenibilidad financiera que se refleja en bajos niveles de coberturas de agua potable y/o alcantarillado, insuficiencias para atender fugas o proporcionar mantenimiento a la red del servicio.

La importancia del recurso agua como un recurso vital y estratégico para la economía y para cualquier tipo de actividades, incentivó al presente trabajo para construir un índice que evalúe el desempeño de las empresas que ofrecen el servicio de agua en México. La idea central es otorgar a los tomadores de decisiones una herramienta que les ayude a focalizar la aplicación de las políticas públicas para el uso eficiente del recurso agua; así como señalar a empresas calificadoras e inversionistas a través de un semáforo que caracterice en: eficiencia baja, eficiencia media e ineficiencia a las empresas que ofrecen el servicio de agua en el país.

Hasta ahora existe solo un trabajo para México, en este tema, y fue realizado por el Consejo Consultivo del Agua, en ese estudio se creó un indicador utilizando promedios simples de 12 variables para 26 empresas operadoras. A pesar de ser un buen acercamiento al tema, el número de empresas es menor al que presento en este trabajo, además no utilizan análisis multivariado en la construcción del indicador que proponen.

A diferencia del indicador del Consejo Consultivo del Agua, el índice que propongo se construyó con información del Programa de Indicadores de Gestión de Organismos Operadores PIGOO del Instituto Mexicano de Tecnología del Agua IMTA y de una investigación telefónica realizada por el antes llamado Instituto Nacional de Ecología INE, ahora Instituto Nacional de Ecología y Cambio Climático, INECC, así como de la recopilación de datos vía correo electrónico y a través del correo postal, en la Figura 1.A del Anexo se puede revisar el cuestionario aplicado. Cabe mencionar que la

información recopilada no proviene de un muestreo aleatorio, sino de un sondeo realizado a 46 empresas operadoras de agua. La idea de este trabajo es crear una metodología para construir un índice de desempeño; y que cuando se cuente con recursos económicos se espera aplicarlo a datos que sí provengan de un muestreo aleatorio.

El índice dimensiona la eficiencia del servicio de agua a través de siete indicadores que evalúan: 1) la puntualidad en el pago del servicio *punt\_pgo*, 2) el volumen de agua facturada *v\_fact*, 3) la rentabilidad de la empresa de agua *rentdad*, 4) la corrupción de la localidad *corrup*, 5) la cobertura de agua potable *c\_aguap*, 6) la cobertura de alcantarillado *c\_alcant* y 7) la eficiencia de cloración *c\_clora*.

Los indicadores son resultado de la construcción de razones de dos variables continuas tomadas de las cifras oficiales para los últimos cuatro indicadores *corrup*, *c\_aguap*, *c\_alcant* y *c\_clora* y de las entrevistas a las empresas de agua para los tres primeros indicadores *punt\_pgo*, *v\_fact* y *rentdad*. Para realizar el indicador se requirió homogenizar las variables utilizadas, para ello se utilizaron razones, en donde el cociente resultante oscila entre cero y uno, el cero representa la calificación más baja y el uno la más alta. Así mismo, se utilizó un método de imputación, para tres observaciones, como lo requería el tratamiento de los datos al encontrar inconsistencias.

Para lograr nuestro objetivo se utilizaron cuatro técnicas de análisis multivariado, como se muestra en el siguiente cuadro:

	MÉTODO UTILIZADO	USO
DESCRIPCIÓN DEL MODELO	Análisis de Componentes Principales	Elaboración del índice de eficiencia de las empresas de agua
	Análisis de Conglomerados	Clasificación de las empresas de agua
REGRESIÓN Y PREDICCIÓN DEL MODELO	Regresión Lineal	Búsqueda de las variables que producen la eficiencia de la empresa de agua
CLASIFICACIÓN DEL MODELO	Análisis discriminante	Búsqueda de diferencias entre los grupos realizados con el análisis de conglomerados

El informe del presente estudio se divide en cuatro capítulos, los primeros tres capítulos están dedicados a revisar la teoría, en el Capítulo 1, se revisará la técnica de Análisis de Componentes Principales utilizada en este estudio para generar el indicador que evalúe la eficiencia de las empresas que ofrecen el servicio de agua. En el Capítulo 2, se expone la técnica de Conglomerados utilizada para agrupar a las empresas de agua en tres grupos; empresas eficientes, empresas con eficiencias regulares y empresas con ineficiencias. En el Capítulo 3 se presenta la técnica de Análisis de Discriminante, con esta técnica se decidió la consolidación de los grupos dadas las características intrínsecas de su administración.

En el Capítulo 4 se realiza el análisis de la información, se elaboró el índice de eficiencia, el análisis de conglomerados, el análisis discriminante y el modelo de eficiencia de las empresas operadoras del servicio de agua, los resultados obtenidos, la interpretación y sus conclusiones.

El índice de eficiencia se realizó a través del análisis de componentes principales usando la matriz de covarianza. Para ello se utilizó solo el primer componente principal que incluye los tres indicadores con mayor variabilidad; *punt\_pgo*, *v\_fact* y *rentdad*, que dicho está de paso, reflejan el comportamiento financiero de cualquier empresa. Este último es el aporte principal del presente trabajo, ya que hasta ahora no existe un semáforo de eficiencia financiera de las empresas que ofrecen el servicio de agua en México. Los trabajos que ha realizado la Comisión de Nacional del Agua se han quedado a un nivel de eficiencia hidráulica, sin evaluar la rentabilidad de la empresa.

En cuanto a la metodología de la construcción del índice, se realizó un análisis de conglomerado con cinco técnicas diferentes, de las cuales los mejores resultados se obtuvieron por los métodos de Ward y de K-mean, para ambos métodos se realizó a través de los tres primeros componentes principales ya que explican el 83% de la variabilidad.

También se usó el método de análisis discriminante lineal de Fisher, y para ello, se utilizaron los todos los indicadores de eficiencia propuestos.

Finalmente, el modelo de eficiencia se realiza bajo una relación lineal entre la variable eficiencia, índice construido en la sección anterior y ocho variables explicativas subdivididas en tres temas: variables legales, variables administrativas y variables exógenas.

# CAPÍTULO 1

---

## El análisis de componentes principales<sup>1</sup>

### *Introducción*

El objetivo de este trabajo es realizar un índice que evalúe la eficiencia de las empresas que ofrecen el servicio de agua en México y una manera de hacerlo es utilizando algún método de Análisis Multivariante como puede ser el Análisis de Componentes Principales, ya que, éste es capaz de transformar ortogonalmente al conjunto de variables que evaluará la eficiencia de las empresas en un nuevo conjunto no correlacionado. La idea central es representar la información con el menor número de variables construidas como combinaciones lineales de las variables originales y, en donde el primer componente explica la mayor variación posible de los datos originales.

Como lo explica Peña, D (2002), uno de los problemas centrales en el análisis de datos multivariantes es la reducción de la dimensionalidad: si es posible describir con precisión los valores de  $p$  variables por un pequeño subconjunto  $r < p$  de ellas, se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información.

### *1.1 COMPONENTES PRINCIPALES*

La idea de esta metodología es encontrar un subespacio de dimensión menor que  $p$  tal que al proyectar sobre éste, los puntos conserven su estructura con la menor distorsión posible.

Para ello se supondrá que  $X' = [X_1, \dots, X_p]$  es un vector aleatorio  $p$ -dimensional con media cero  $\mu = 0$ , y con varianza  $S$ .

Entonces se buscará un conjunto de variables  $Z_1, \dots, Z_p$  de manera que no estén correlacionadas y en donde las varianzas decrezcan de la primera a la última. Cada  $Z_j$  es una combinación lineal de  $X' = [X_1, \dots, X_p]$ , así;

$$Z_i = a_{1i}X_1 + \dots + a_{pi}X_p = a'_i X \quad \text{en donde } i=1, \dots, p$$

---

<sup>1</sup> La técnica de componentes principales es debida a Hotelling (1933), aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson (1901).

Con  $\acute{a}_l = [a_{1i}, \dots, a_{pi}]$  que es un vector de constantes.

### 1.1.1 EL PRIMER COMPONENTE PRINCIPAL

Se define como la combinación lineal de las variables originales que tiene varianza máxima. Los valores en este primer componente de los  $n$  individuos se representarán por el vector  $Z_1$ ;

$$Z_1 = X'a_1 = a_1'X$$

Y la varianza de  $Z_1$  está dada por:

$$\begin{aligned} \text{var}(Z_1) &= \text{var}(a_1'X) \\ \text{var}(Z_1) &= \acute{a}_1'Sa_1 \end{aligned}$$

En donde  $S$  es la matriz de varianzas y covarianzas de las observaciones y para maximizar  $Z_1$  se considera la siguiente restricción  $a_1'a_1 = 1$ . Introduciendo la restricción mediante el uso del multiplicador de Lagrange se obtiene la varianza<sup>2</sup>;

$$\text{var}(Z_1) = \lambda$$

### 1.1.2 EL SEGUNDO COMPONENTE PRINCIPAL

Se define como la combinación lineal de las variables originales con la segunda mayor varianza y es ortogonal a la primera. Los valores del segundo componente de los  $n$  individuos se representan por la extensión del primer componente, así que se establece como función objetivo, el resultado de la adición de las varianzas de  $Z_1$  y  $Z_2$  es decir:

$$Z_1 = X'a_1 \text{ y } Z_2 = X'a_2$$

En donde:  $Z_1$  y  $Z_2$  no deben estar correlacionadas;  $a_1$  y  $a_2$  son vectores propios ortogonales y que definen un plano; y en donde  $S$  es la matriz de varianzas y covarianzas de las observaciones.

Como se quiere maximizar  $Z_2$  se deben considerar las siguientes restricciones  $a_2'a_2 = 1$  y  $a_1'a_2 = 0$ .

---

<sup>2</sup>  $M = a_1' Sa_1 - \lambda(a_1'a_1 - 1)$

$\frac{\partial M}{\partial a_1} = 2Sa_1 - 2\lambda a_1 = 0, Sa_1 = \lambda a_1$

Entonces:  $\text{var}(Z_1) = \lambda$  en donde  $\lambda$  será el mayor valor propio de la matriz  $S$  y  $a_1$  es el vector propio correspondiente.

Y la varianza<sup>3</sup> de esta segunda componente es:

$$var(Z_2) = a'_2 S a_2$$

De lo anterior se deriva la función objetivo:

$$\phi = a'_1 S a_1 + a'_2 S a_2 - \lambda_1(a'_1 a_1 - 1) - \lambda_2(a'_2 a_2 - 1)$$

De donde después de derivar se llega a la solución:

$$\phi = \lambda_1 + \lambda_2$$

En donde  $\lambda_1$  y  $\lambda_2$  son los dos autovalores mayores de la matriz S y  $a_1$  y  $a_2$  sus correspondientes autovectores. Observamos que la covarianza entre  $z_1$  y  $z_2$ , dada por  $a'_1 S a_2$  es cero ya que  $a'_1 a_2 = 0$ , y las variables  $z_1$  y  $z_2$  estarán no correlacionadas. La construcción de las sucesivas componentes principales se construye de manera semejante.

### 1.1.3 SELECCIÓN DEL NÚMERO DE COMPONENTES

Peña, D (2002) enumera las siguientes reglas para seleccionar el número de componentes

1. Realizar un gráfico de  $\lambda_i$  frente a  $i$ . Comenzar seleccionando componentes hasta que los restantes tengan aproximadamente el mismo valor de  $\lambda_i$ . La idea es buscar un "codo" en el gráfico, es decir un punto a partir del cual los valores propios son aproximadamente iguales.

Dicha gráfica es llamada de eigenvalores ordenados y proceden de una matriz de correlación o de covarianza, es usada para identificar el número apropiado de factores o componentes principales. La característica clave buscada en la gráfica es un "codo" de modo que el número de componentes/factores tomados son el número de eigenvalores que se encuentren a la izquierda de la

<sup>3</sup> Varianza de la segunda componente  $var(Z_2) = var(a'_2 X)$

Derivadas de la función objetivo del segundo componente principal

$$\frac{\partial \phi}{\partial a_1} = 2S a_1 - 2\lambda_1 a_1 = 0$$

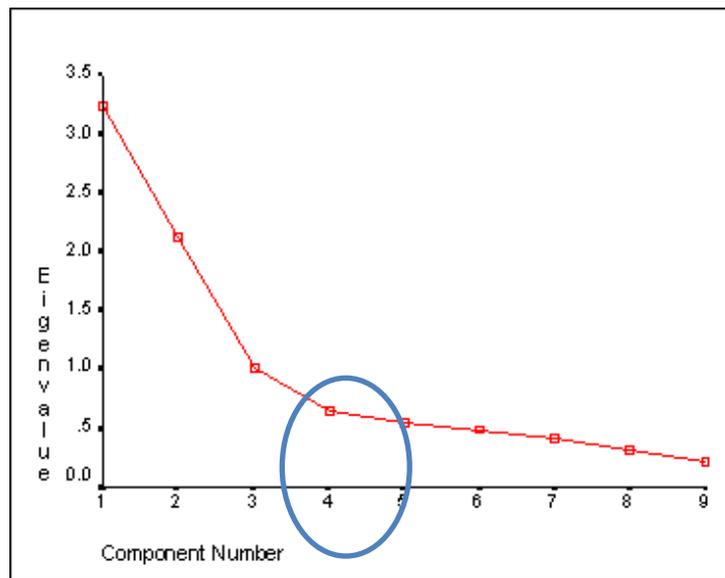
$$\frac{\partial \phi}{\partial a_2} = 2S a_2 - 2\lambda_2 a_2 = 0$$

$$S a_1 = \lambda_1 a_1$$

$$S a_2 = \lambda_2 a_2$$

observación que genera el codo, por ejemplo en la Figura 1 se observa que a partir del cuarto componente hay una caída en la capacidad predictiva, por ello para decidir cuántos componentes se utilizarán se considerarán los cuatro primeros.

**Figura 1**  
**Ejemplo de Screeplot**



2. Seleccionar componentes hasta cubrir una proporción determinada de varianza. Esta regla es arbitraria y debe aplicarse con cierto cuidado.
3. Desechar aquellos componentes asociados a valores propios inferiores a una cota, que suele fijarse como la varianza media,  $\sum \lambda_i/p$ . En particular, cuando se trabaja con la matriz de correlación, el valor medio de los componentes es 1.

# CAPÍTULO 2

---

## EL ANÁLISIS DE CONGLOMERADOS

### *Introducción*

Dado un conjunto de datos o variables que evalúa a ciertos individuos, en algunas ocasiones lo que queremos es observar la existencia de grupos, o como en este trabajo, se quiere clasificar en grupos a las empresas operadoras de agua dada la eficiencia de su desempeño. El objetivo de este método multivariado es construir conglomerados de individuos que se comporten de manera homogénea entre sí, en función de las similitudes entre ellos, y heterogéneos respecto a otros conglomerados.

Las características que se buscan en el análisis de conglomerados se puede resumir en tres puntos; 1) cada elemento pertenecerá a uno, y sólo uno, de los grupos, 2) que todo elemento quede clasificado y 3) cada grupo será internamente homogéneo.

Los métodos de conglomerados se pueden dividir en dos grupos; jerárquicos y no jerárquicos. Los primeros, son resultado de una matriz de distancias e implica que los datos se ordenen en niveles, de manera que los niveles superiores contienen a los inferiores. Estos métodos a su vez se subdividen en algoritmos aglomerativos y divisivos. El método aglomerativo agrega en grupos comenzando con los elementos individuales y va añadiendo al elemento que no han sido agregando o al grupo. Finalmente, el conglomerado divisivo, divide sucesivamente al conjunto total de los elementos hasta llegar a los elementos individuales.

Entre los métodos no jerárquicos se encuentra el de  $K$ -medias, cuyo objetivo es dividir en un número  $K$  de grupos prefijado, se dan  $K$  puntos que sirven de centros iniciales y de manera iterativa al optimizar cierta función objetivo, como minimizar la suma de varianzas al interior de los  $K$  grupos, se van reacomodando los individuos y reasignando los centros.

### *2.1. MEDIDAS DE PROXIMIDAD*

Las medidas de proximidad miden el grado de similitud y disimilitud que existe entre dos objetos de forma que, cuanto mayor es su valor, mayor es el grado de similaridad existente entre ellos y con mayor probabilidad los métodos de clasificación tenderán a ponerlos en el mismo grupo.

### 2.1.1. DISIMILARIDAD Y MEDIDAS DE DISTANCIA PARA DATOS CONTINUOS

Las medidas de disimilitud miden la distancia entre dos objetos de forma que, cuanto mayor sea su valor, más diferentes son los elementos y menor será la probabilidad de que los métodos de clasificación los pongan en el mismo grupo. Existen varias medidas de distancia, algunas de ellas son: 1) la distancia euclidiana, 2) la distancia Minkowski y 3) la distancia City Block.

#### 1. Distancia euclidiana

$$D(\underline{X}_i, \underline{X}_j) = \left[ \sum_{k=1}^p (X_{ik} - X_{jk})^2 \right]^{1/2}$$

#### 2. Distancia de *Minkowski*

$$D(\underline{X}_i, \underline{X}_j) = \left[ \sum_{k=1}^p (X_{ik} - X_{jk})^\lambda \right]^{1/\lambda}$$

Con  $\lambda$  un entero.

#### 3. Caso particular de la Distancia *City Block*

$$D(\underline{X}_i, \underline{X}_j) = \sum_{k=1}^p |(X_{ik} - X_{jk})|$$

### 2.2 MEDIDAS DE PROXIMIDAD INTRA- GRUPOS

Definir la distancia entre grupos es un concepto central en el análisis de conglomerados. La distancia o cercanía entre-grupos se puede definir de dos maneras. Supongamos que existen dos grupos A y B, el grupo A con  $n_a$  elementos y el grupo B con  $n_b$  elementos. Entonces, la distancia entre los grupos A y B se puede definir, como el resumen de las proximidades de todas las observaciones de  $n_a$  y  $n_b$  para cada grupo, por ejemplo la liga sencilla toma el mínimo entre todas las distancias posibles. Por otro lado, también se puede precisar como la proximidad entre una observación representativa o una estadística resumen, como por ejemplo, la liga centroide toma como observación representativa al centroide de cada grupo.

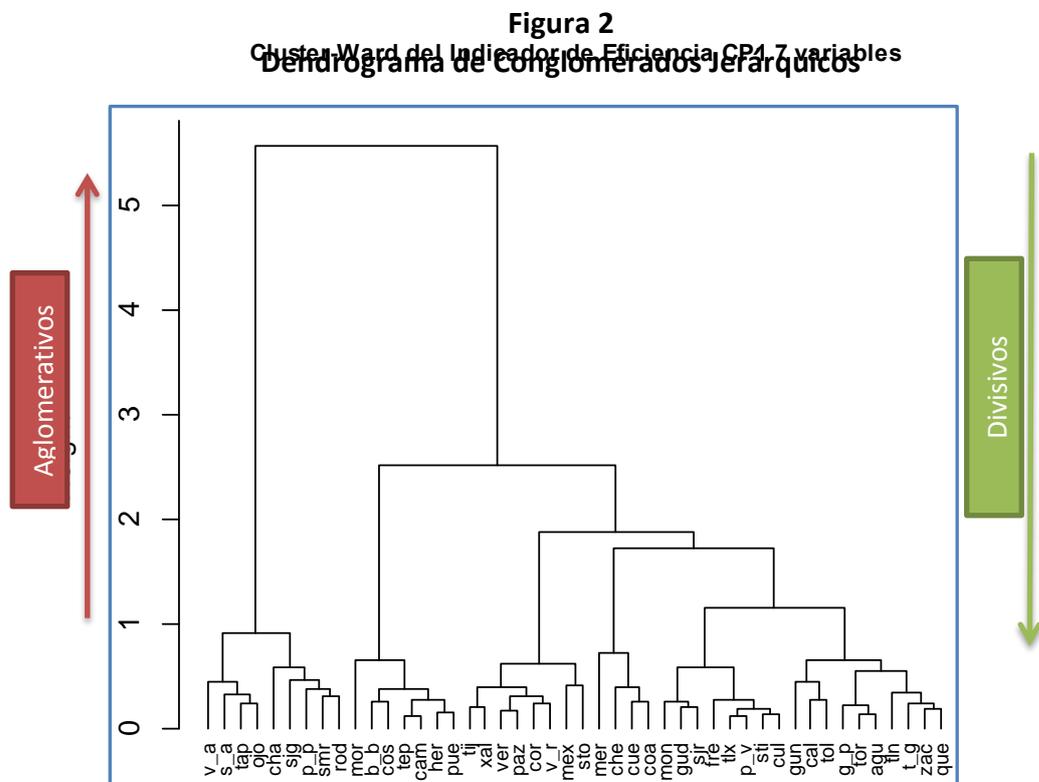
Subsecuentemente, si se quisiera calcular una nueva distancia entre el reciente grupo AB al grupo C con  $n_c$  elementos, se realizará con alguno de las maneras descritas arriba.

El cálculo de las distancias entre grupos se realiza a través de diversos métodos, como pueden ser de liga simple, de liga completa, liga de media de grupos o *Ward*; como se explica en la siguiente sección.

### 2.3 CONGLOMERADOS JERÁRQUICOS

Los conglomerados jerárquicos pueden ser de tipo aglomerativo o divisivo<sup>4</sup>, la diferencia entre ambos es que el primero empieza desde la más pequeña partición posible, es decir, que cada observación es tomada como un conglomerado; mientras que en los métodos divisivos se empieza con el grupo que contiene a todas las observaciones y se procede a dividir en grupos de menor tamaño.

Las clasificaciones jerárquicas pueden ser representadas en un diagrama bidimensional conocido como dendrograma o árbol jerárquico, el cual ilustra las divisiones hechas en cada estado del análisis y es útil cuando los puntos tienen claramente una estructura jerárquica. Véase Figura 2.



Fuente: Elaboración Propia, 2015.

4 Para mayor información sobre este tipo de método revisar Seber (1984).

```
(dist(ind_cor[, 2:8]))
hclust (*, "ward")
```

### 2.3.1 MÉTODOS AGLOMERATIVOS

La idea central es que inicialmente se tienen tantos conglomerados como objetos. Luego se van agrupando de modo que los primeros en hacerlo son los más similares y al final, todos los subgrupos se unen en un único conglomerado.

Los algoritmos utilizados en estos métodos tienen siempre la misma estructura y sólo se diferencian en la forma de calcular las distancias entre grupos. La estructura es la siguiente como lo explica (Peña, 2002).

1. Comenzar con tantas clases como elementos,  $n$
2. Seleccionar los dos elementos más próximos en la matriz de distancias formando una clase.
3. Sustituir los dos elementos utilizados, en el punto 2, para definir la clase por un nuevo elemento que represente la clase construida y así sucesivamente hasta completar todos los elementos,  $n-1$  veces, agrupados en una única clase.

### 2.3.2. CALCULO DE DISTANCIAS

Existen varias distancias como: el encadenamiento simple o vecino más cercano, encadenamiento completo o vecino más lejano, media de grupos, método centroide y método Ward, para los fines de este trabajo solo se revisará el método Ward. Para revisar el resto de los métodos se pueden ver en Morrison, (1990).

Método de *Ward*<sup>5</sup>, se define como una medida global de la heterogeneidad de una agrupación de observaciones en grupos. Esta medida  $W$  es la suma de las distancias euclídeas al cuadrado entre cada elemento y la media de su grupo.

$$W = \sum_{g=1}^G \sum_{i=1}^{n_g} (X_{ig} - \bar{X}_g)(X_{ig} - \bar{X}_g)'$$

Donde  $\bar{X}_g$  es la media del grupo  $g$ . El criterio comienza suponiendo que cada dato forma un grupo  $g$ ,  $g=n$  y, por tanto,  $W$  es cero. Después, se unen los elementos que produzcan el incremento mínimo de  $W$ . En la siguiente etapa, se tiene  $n-1$  grupos,  $n-2$  de un elemento y uno de dos elementos. A continuación se unirán nuevamente dos grupos para que  $W$  crezca lo menos posible, con lo que se pasa a  $n-2$  grupos y así sucesivamente hasta tener un único grupo. Los valores de  $W$  van indicando el crecimiento del criterio al formar grupos y puede utilizarse para decidir cuantos grupos

---

5 Método propuesto por Ward y Wishart, Peña, D. (2002)

naturales contienen nuestros datos. Para cada etapa los grupos que deben unirse para minimizar (W), sobre las distancias al cuadrado entre los grupos y sus centros de grupo

$$\min \frac{n_a n_b}{n_a + n_b} (\bar{X}_a - \bar{X}_b)' (\bar{X}_a - \bar{X}_b)^6$$

## 2.4 CONGLOMERADOS NO JERÁRQUICOS

Los métodos de partición o conglomerados no jerárquicos tienen como característica el utilizar la matriz de datos originales. El método más conocido de los conglomerados no jerárquicos es el método de k-medias cuyo objetivo es dividir la muestra en un número de grupos prefijado,  $K$ . En los siguientes pasos se puede resumir el algoritmo de k-medias:

1. Seleccionar  $K$  puntos como centros de los grupos iniciales  
Lo cual se puede hacer a) asignando aleatoriamente los objetos a los grupos y tomando los centros de los grupos formados, b) tomando como centros los  $K$  puntos más alejados entre sí, y c) construyendo unos grupos iniciales con información a priori y calculando sus centros, o bien seleccionando los centros a priori.
2. Calcular las distancias euclidianas de cada elemento a los centros de los  $K$  grupos, y asignar cada elemento al grupo a cuyo centro esté más próximo.
3. Definir un criterio de optimalidad y comprobar si reasignando alguno de los elementos mejora el criterio.

El algoritmo del método de k-medias utiliza como criterio de optimalidad, la idea es minimiza la suma de cuadrados dentro de los grupos para todas las variables; como se observa<sup>7</sup>

<sup>6</sup> En donde  $n_a$  es el número de elementos del grupo A,  $n_b$  es el número de elementos del grupo B y  $\bar{X}_a$  y  $\bar{X}_b$  es la media del grupo A y B, respectivamente.

<sup>7</sup> En donde  $x_{ijg}$  es el valor de la variable  $j$  en el elemento  $i$  del grupo  $g$  y  $\bar{x}_{jg}$  es la media de esta variable en el grupo.

El criterio de la función es equivalente a minimizar la suma ponderada de las varianzas de las variables en los grupos;

$$\min \text{SCDG} = \sum_{g=1}^K \sum_{j=1}^p n_g s_{jg}^2 \text{ En donde } S_{jg}^2 = \frac{\sum (x_{ijg} - \bar{x}_{jg})^2}{n_{1-g}}, n_g \text{ es el número de elementos del grupo } g \text{ y } s_{jg}^2$$

es la varianza de la variable  $j$  en dicho grupo.

$$\min SCDG = \min \sum_{g=1}^K \sum_{j=1}^p \sum_{i=1}^{n_g} (x_{ijg} - \bar{x}_{jg})^2$$

Otro criterio de homogeneidad, que trata de minimizar las distancias al cuadrado entre los puntos y el centro del grupo es la distancia con la norma euclídea, como se observa en la siguiente ecuación

$$\min \sum_{g=1}^K \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)' (x_{ig} - \bar{x}_g) = \min \sum_{g=1}^K \sum_{i=1}^{n_g} d^2(i, g)$$

Donde  $d^2(i, g)$  es el cuadrado de la distancia euclídea entre el elemento  $i$  del grupo  $g$  y la media de su grupo y el anterior escalar es igual a su traza<sup>8</sup>, por lo tanto;

$$\min tr(W) = \min SCDG^9$$

El criterio de la traza tiene dos propiedades importantes; primero, que no es invariante a cambios en la escala y segundo, que al minimizar la distancia euclídea produce grupos aproximadamente esféricos.

El algoritmo de k-medias busca seccionar de manera óptima con la restricción de que en cada iteración sólo se moverá un elemento de un grupo a otro.

---

<sup>8</sup> Para mayor detalle revisar el trabajo propuesto por Ward en 1963

<sup>9</sup> En donde  $W$  es la matriz de suma de cuadrados dentro de los grupos

$$W = \sum_{g=1}^K \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)'$$

# CAPÍTULO 3

---

## EL ANÁLISIS DISCRIMINANTE<sup>10</sup>

### *Introducción*

Frecuentemente es necesario identificar las características que permitan encontrar las particularidades que diferencien a dos o más grupos o sujetos. El análisis discriminante, es una técnica estadística capaz de identificar qué variables permiten diferenciar a los grupos y cuántas de estas variables son necesarias para alcanzar la mejor clasificación posible. El objetivo es encontrar la combinación lineal de las variables que mejor permita diferenciar a los grupos.

El Análisis Discriminante se le considera como un análisis de regresión donde la variable dependiente es categórica y tiene como categorías las etiquetas de cada uno de los grupos, y las variables independientes son continuas y determinan a qué grupos pertenecen los objetos. Se pretende encontrar relaciones lineales entre las variables continuas que mejor discriminen los grupos.

### *3.1. ANÁLISIS DISCRIMINANTE*

Peña, D (2002) “Se dispone de un conjunto amplio de elementos que pueden venir de dos o más poblaciones distintas. En cada elemento se ha observado una variable aleatoria  $p$ -dimensional  $x$ , cuya distribución se conoce en las poblaciones consideradas. Se desea clasificar un nuevo elemento con valores conocidos de las variables, en alguna de las poblaciones”.

#### *3.1.1 FUNCIÓN LINEAL DISCRIMINANTE*

Esta regla está basada en la normalidad multivariante de las variables consideradas. Supongamos entonces, que  $f_1$  y  $f_2$  son distribuciones normales con distintos vectores de medias pero idéntica matriz de covarianzas. También supongamos que existe un

---

<sup>10</sup> Ronald Aylmer Fisher (1890-1962) se le sitúa como el padre de este procedimiento.

elemento genérico  $x$  que pertenece a la población  $i=1,2$  y tiene la siguiente función de densidad:

$$f_i(x) = \frac{1}{(2\pi)^{p/2}|V|^{1/2}} \exp\left\{-\frac{1}{2} (x - \mu_i)' V^{-1}(x - \mu_i)\right\} \quad i = 1,2$$

En donde la función discriminante lineal<sup>11</sup>, que coincide con la planteada por Fisher, es:

$$L(x) = \left(x - \frac{1}{2} (\mu_1 + \mu_2)\right)' V^{-1} (\mu_1 + \mu_2)$$

Entonces

*y se asigna a  $f_1$  si  $L(x) > 0$ ,*

*y se asigna a  $f_2$  si  $L(x) < 0$ ,*

### 3.1.2 REGLA DE MÁXIMA VEROSIMILITUD

En términos generales, se puede decir que cuando se tienen  $g$  grupos,  $g > 2$ , se obtienen  $g-1$  funciones lineales que van a ayudar a hacer la clasificación de los individuos.

Supongamos que  $f_1$  y  $f_2$  dos poblaciones y  $X = (x_1, \dots, x_p)'$  un vector con distribución de probabilidad conocida, dependiente de un parámetro  $\phi$  que toma el valor  $\phi_1$  si  $X \in f_1$  y  $\phi_2$  si  $X \in f_2$ .

Sea  $x = (x_1, \dots, x_p)'$  el vector de observaciones de  $X$  sobre un individuo y la probabilidad de la observación  $x$  en  $f_i$  es  $L_i(x) = f(x_1, \dots, x_p; \phi_i)$ .

La regla discriminante de máxima verosimilitud consiste en asignar  $x$  a la población  $f_i$  para la cual la verosimilitud de la observación es mayor. Esta regla tiene asociada la siguiente función discriminante

$$V(x) = \log L_1(x) - \log L_2(x)$$

Si  $V(x) > 0$ , y se asigna a  $f_1$

Si  $V(x) < 0$ , y se asigna a  $f_2$

---

<sup>11</sup> Para revisar el procedimiento completo, se puede revisar a Peña. D (2002)

# CAPÍTULO 4

---

## APLICACIONES

### 4.1 CONSTRUCCIÓN DEL ÍNDICE DE EFICIENCIA

El ejercicio del análisis de las empresas operadoras de agua en este trabajo se realizó en dos pasos, en el primero, se creó un índice de eficiencia de las empresas de agua para lo que se utilizó una base de datos de las características operativas de la empresa. En segundo lugar, se realizó un análisis de regresión, en donde se pretende encontrar los elementos políticos, legales y administrativos que expliquen la eficiencia de las empresas de agua que se presentan en este trabajo, y para ello se utilizaron ocho variables explicativas.

Para la construcción del índice de eficiencia se utilizó una base de datos<sup>12</sup> de cuarenta y seis empresas operadoras de agua y siete variables (punt\_pago, v\_fact, rentdad, corrup, c\_aguap, c\_alcant y clora) que explicarán la eficiencia del servicio de la empresa. Los valores de las variables que explican la eficiencia de las empresas oscilan entre cero y uno, en donde cero es la calificación más baja y uno la más alta. En el análisis de los datos se identificaron algunos problemas de inconsistencias, para ello se utilizó un método de imputación en tres observaciones.

Las variables que se utilizan para explicar la eficiencia de las empresas de agua, es decir el índice de eficiencia ( $I_1$ )<sup>13</sup>, se tomaron de una encuesta telefónica y/o vía correo electrónico que realizó el Instituto Nacional de Ecología (INE) a cuarenta y seis empresas operadoras de agua en veinticuatro estados del país.

La encuesta consistió en preguntas contextualizadas en tres rubros; características políticas del estado o municipio en donde se encuentra la empresa, características administrativo-financieras y características institucionales de la empresa. Para el resto de las variables se utilizaron datos estadísticos públicos del Consejo Nacional de Población CONAPO, de la Comisión Nacional del Agua CNA y de la página electrónica de la Cámara de diputados, en los Institutos Estatales Electorales y en el Instituto Nacional de Estadística y Geografía INEGI.

---

<sup>12</sup> La base de datos fue resultado de la investigación y recopilación de información en un periodo de ocho meses.

<sup>13</sup> Variable que se obtendrá en la sección 4.5 de este capítulo

En la Tabla 1 se observa en la primera sección las variables con las que se construirá el índice de eficiencia y en la segunda sección, se puede observar la descripción de aquellas variables que construirán el modelo de eficiencia.

**Tabla 1**  
**Descripción de las variables que conforman el índice y el modelo de eficiencia**

	Variable	Descripción	Descripción
<i>Construcción del Índice de eficiencia</i>	<i>punt_pgo</i>	Puntualidad en el pago	Es la razón de los usuarios que pagan a tiempo y los usuarios totales
	<i>v_fact</i>	Volumen de facturación	Es la razón del volumen de agua facturado y el total de agua producido
	<i>Rentdad</i>	Rentabilidad	Es la razón de ingresos por la venta de agua y los costos de operación, mantenimiento y administración.
	<i>Corrup</i>	Corrupción	Es el índice nacional de corrupción y buen gobierno.
	<i>c_aguap</i>	Cobertura de agua potable	Porcentaje de viviendas con servicios de agua potable
	<i>c_alcant</i>	Cobertura de alcantarillado	Porcentaje de viviendas con servicios de saneamiento
	<i>c_clora</i>	eficiencia de cloración	Porcentaje de agua clorada
<i>Construcción del modelo</i>	<i>Reg</i>	Reglamento	Variable dicotómica que refleja la existencia de un reglamento administrativo, en donde 0= falta de reglamento y 1 = existencia del mismo.
	<i>prep_dir</i>	Preparación del director de la empresa	Es una variable continua, se mide como el número de años estudiados por el directivo de la empresa.
	<i>Cont</i>	Continuidad	Variable dicotómica en donde 0= a la falta de continuidad de objetivos directivos y 1= continuidad de objetivos directivos.
	<i>certificado</i>	Certificado	Variable dicotómica en donde 0= sin certificado administrativo y 1= cuenta con un certificado administrativo
	<i>tip_tarif</i>	Tipo de tarifa	Variable dicotómica en donde 0 = tarifa fija y 1= tarifa volumétrica
	<i>tarifa_dom</i>	Tarifa doméstica	Es una variable continua, medida en pesos (\$)
	<i>Zd</i>	Zona de disponibilidad	Variable dicotómica en donde 0= zonas con poca disponibilidad y 1= zonas con alta disponibilidad
	<i>Prec_mfluv</i>	Precipitación media fluvial	Variable continua, medida en milímetros cúbicos de agua.

Fuente: Elaboración Propia, 2015.

*punt\_pgo*. La puntualidad<sup>14</sup> de pago representa la proporción de los usuarios del servicio de agua que pagan de manera puntual el servicio que le ofrece la empresa de agua. Este indicador se construyó con dos variables: 1) el registro de hogares o usuarios que cuentan con servicio de agua potable en el área de incidencia de la

<sup>14</sup> La puntualidad se determinó, como el pago oportuno de la factura de agua, hasta por dos periodos de facturación, si a la tercera factura no se ha cumplido con el pago se considera un usuario moroso.

empresa y 2) el total de los hogares del área de incidencia de la empresa que pagaron a tiempo.

*v\_fact*. Volumen de facturación refleja el porcentaje del volumen de agua que la empresa operadora factura<sup>15</sup>. El indicador es la razón entre el volumen de agua facturada y el volumen de agua extraída del subsuelo por la empresa que ofrece el servicio. Mientras más cercana esté del número uno, mejores condiciones económicas de administración tendrá la empresa.

*rentdad*. La rentabilidad<sup>16</sup> es el resultado del cociente entre el ingreso (en pesos) por la venta de agua a los usuarios del servicio y los costos de administración, operación y mantenimiento (en pesos) en los que incurre la empresa para ofrecer el servicio al usuario.

*corrup*. Es una variable obtenida a nivel estatal para el año 2007 que registra la corrupción en servicios públicos ofrecidos por los tres niveles de gobierno y por particulares. La variable se obtuvo del Índice Nacional de Corrupción y Buen Gobierno que publica Transparencia Mexicana.

*c\_aguap*. Cobertura<sup>17</sup> de agua potable, muestra la proporción de la población en 2007 que habita en viviendas particulares y que cuenta con agua entubada dentro de la vivienda, dentro del terreno o de una llave pública o hidrante. Niveles altos, cercanos a uno, representan altos niveles en la extensión del servicio.<sup>18</sup>

*c\_alcant*. Cobertura de alcantarillado, mide la proporción de la población en 2007 que habita en viviendas particulares, cuya vivienda cuenta con un desagüe conectado a la red pública de alcantarillado, a una fosa séptica, a un río, lago, mar, a una barranca o grieta<sup>19</sup>. Un indicador cercano a uno, representa apreciables niveles en la extensión del alcantarillado en la población.

---

<sup>15</sup> La no facturación puede deberse a las siguientes razones: existen tomas clandestinas, fallas en el medidor del usuario, problemas de eficiencia física (fugas) en el traslado del líquido.

<sup>16</sup> Lo deseable es que las empresas cuenten con un indicador igual o mayor a uno, lo que implica que los ingresos por el pago del servicio son del mismo monto o mayor que los costos en los que incurren para ofrecer el servicio.

<sup>17</sup> Es importante resaltar que este porcentaje no muestra la distancia a la que se encuentra la llave o el hidrante, la calidad del agua que abastece o la frecuencia o tandeo con la que se ofrece el servicio.

<sup>18</sup> Sin embargo, es tan ambigua la definición que a pesar de que un hogar cuente con el servicio, no asegura que éste sea de buena calidad, sin tandeo o que se encuentre lejos de la vivienda, motivo por el cual se debe poner atención en el manejo de este indicador.

<sup>19</sup> Sin embargo, la salida de aguas residuales puede estar conectada a desagües clandestinos o bien que contaminen los mantos acuíferos subterráneos o superficiales, lo que provocaría en un mediano plazo un problema en la calidad del agua.

*c\_clora*. Eficiencia de cloración, es el parámetro que evalúa la calidad del agua para el uso y consumo humano distribuida por los sistemas de abastecimiento público. Mide la eficiencia de cloración en el suministro del servicio de agua los hogares que garantizan la calidad del servicio. El indicador se construyó a partir del promedio de las determinaciones de los niveles de concentración de cloro en agua con base al rango de concentración de entre 0.2 a 1.5 mg/l de cloro que marca la NOM-179-SSA1-1998 de calidad bacteriológica del servicio de agua.

*reg*. Reglamento, evalúa la existencia de un reglamento administrativo de la empresa operadora del agua.

*prep\_dir*. La preparación del directivo, se refiere a la preparación académica que tiene la persona que se encuentra a cargo de la empresa de agua.

*Cont. Continuidad*, representa la constancia de un plan de trabajo llevado a cabo por el mismo directivo de la empresa de agua, se consideró continuidad cuando el directivo permaneció en el puesto por más de 3 años.

*certificado*. Es una variable que representa la presencia de algún certificado por buen uso administrativo o mejoramiento de la eficiencia de la empresa de agua. Los certificados otorgados pueden ser ISO 14000 o el certificado de alguna calificadora municipal

*tip\_tarif*. Tipo de tarifas, es la variable que caracteriza a las tarifas aplicadas en el país, así, se divide en tarifa fija (precio constante independientemente del consumo) o una tarifa volumétrica (el precio depende del consumo).

*tarifa\_dom*. Tarifa doméstica del servicio de agua potable es una variable continua que refleja el precio que debe pagar el usuario del servicio de agua. Esta tarifa se relaciona con la variable tipo de tarifa y el uso de ambas variables en el modelo pretende encontrar la relación que existe con la eficiencia de la empresa.

*Zd*. Es la caracterización de la zona de disponibilidad de agua en donde se encuentra ubicada la empresa de agua, se dividió para este ejercicio en dos zonas, con disponibilidad y sin disponibilidad de agua.

*prec\_mfluv*. Precipitación media fluvial, se refiere al promedio de milímetros cúbicos de agua que han caído en un año en una zona determinada.

En la Tabla 2 se presentan la localización, el estado y municipio, de las cuarenta y seis empresas de las que tenemos información y que se incluyeron en el análisis de este

trabajo, se encuentra también la abreviatura que se utilizará a partir de este momento para nombrar a las empresas.

**Tabla 2**  
**Empresas de agua utilizadas en el estudio**

Estado	Municipio	Abreviatura	Estado	Municipio	Abreviatura
Aguascalientes	Aguascalientes	agu	Nayarit	Bahía de Banderas	b_b
	Calvillo	cal		Tepic	tep
	Cosío	cos		Xalisco	xal
	S.J. Gracia	sjg	Nuevo León	Monterrey	mon
Baja California	Mexicali	mex	Puebla	Puebla	pue
Baja California Sur	Tijuana	tij	Querétaro de	Querétaro	que
	La Paz	paz	Arteaga	San Juan del Río	sjr
Campeche	Champtón	cha	Quintana Roo	Chetumal	che
	Campeche	cam	San Luis Potosí	Santa María del Río	smr
Chiapas	Tapachula	tap		Villa de Arista	v_a
	Tuxtla Gutiérrez	t_g		Villa de Reyes	v_r
Coahuila de Zaragoza	Torreón	tor	Sinaloa	Culiacán	cul
	Saltillo	sti	Sonora	Hermosillo	her
Durango	Gómez Palacio	g_p		Puerto Peñasco	p_p
	Rodeo	rod	Tlaxcala	Tlaxcala	tlx
Guanajuato	Guanajuato	gun	Veracruz de	Córdoba	cor
Jalisco	Guadalajara	gud	Ignacio de	Veracruz	ver
	Puerto Vallarta	p_v	Llave	Coatzacoalcos	coa
	El Salto	sto	Yucatán	Mérida	mer
México	Toluca	tol	Zacatecas	Fresnillo	fre
	Tlalnepantla	tln		Zacatecas	zac
Michoacán	Morelia	mor		Ojocaliente	ojo
Morelos	Cuernavaca	cue	Sain Alto	s_a	

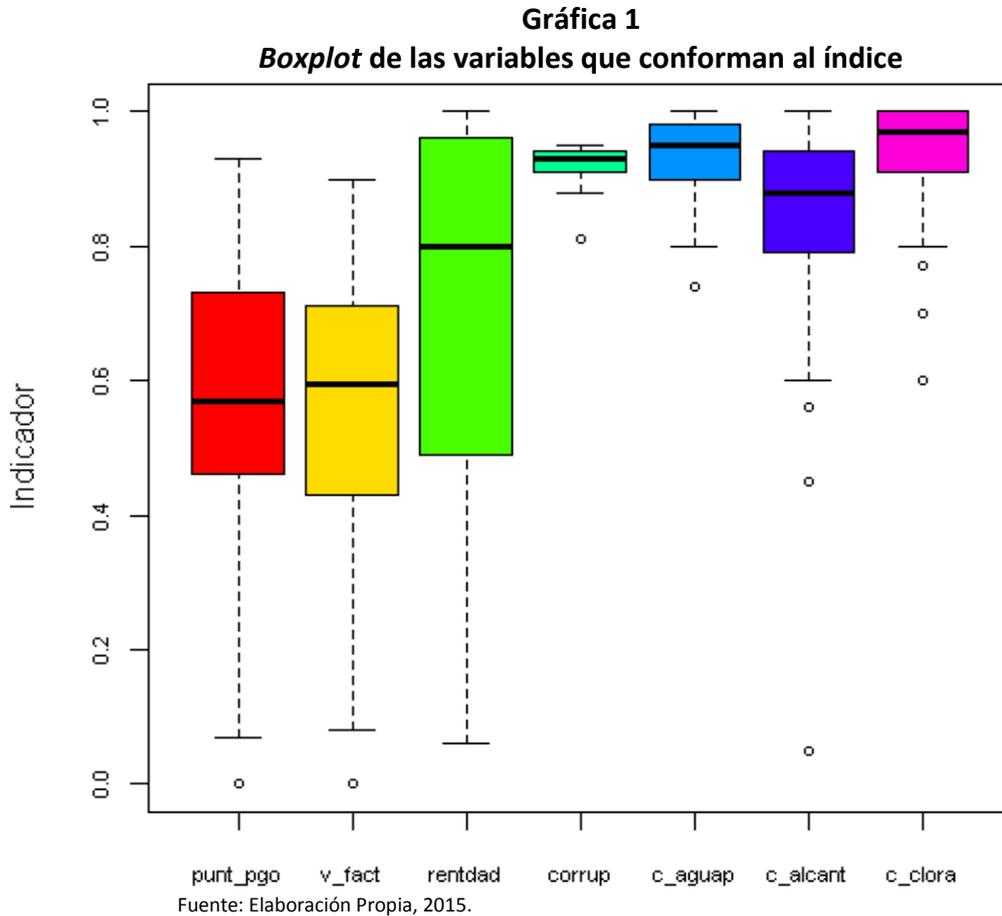
Fuente: Elaboración Propia, 2015.

#### 4.2. ANÁLISIS DESCRIPTIVO<sup>20</sup>

La primera revisión a las variables que conformarán el índice de eficiencia de las empresas de agua se obtiene al observar el Gráfica 1. Así, para la variable *punt\_pgo* la mediana se encuentra muy cerca de 0.6 y el 50% de los datos se concentran entre 0.45 y 0.7, existe datos atípicos en el más bajo nivel; algo muy parecido sucede con la variable *v\_fact*, pero a un rango ligeramente debajo de la variable anterior. Por otro lado, la variable *rentdad* es la que más resalta por la longitud de la variabilidad, en donde más de la mitad de los datos se encuentran entre 0.45 y 0.95. Por el contrario la variable *corrup* tiene una variabilidad pequeña. Comportamientos muy similares presentan las cajas de *c\_aguap* y *c\_clora* con cajas cortas que oscilan entre 0.9 y 1 y

<sup>20</sup> La base de datos se encuentra en el Anexo

ambas presentan datos atípicos, en el caso de la variable *c\_alcant* presenta también una baja variabilidad en los datos, con un rango que va entre 0.8 y 0.95 y su mediana está cerca de 0.9 y con tres datos atípicos.



En la Tabla 3 se presenta la estadística descriptiva de las siete variables que conformarán el índice de eficiencia de las 46 empresas de agua de las que se obtuvo información completa.

Se observan algunos datos dudosos como lo son los dos *ceros* absolutos en los indicadores de puntualidad de pago y en volumen de facturación. Al igual que el indicador de rentabilidad y de cobertura de alcantarillado, como se observa en la Tabla 3. De la misma forma, se debe tomar con suspicacia los datos que reporten *uno* en cualquiera de los indicadores.

**Tabla 3**  
**Medidas de tendencia central de las siete variables que conforman el índice**

	punt_pgo	v_fact	Rentdad	corrup	c_aguap	c_alcant	c_clora
Minino	0	0	0.06	0.81	0.74	0.05	0.60
1st Qu.	0.46	0.44	0.50	0.91	0.90	0.79	0.91
Media	0.57	0.60	0.80	0.93	0.95	0.88	0.97
Mediana	0.56	0.56	0.71	0.92	0.94	0.84	0.94
3rd Qu.	0.73	0.71	0.96	0.94	0.98	0.94	1.00
Máximo	0.93	1.00	1.00	0.95	1.00	1.00	1.00
Varianza	0.048	0.046	0.073	0.001	0.003	0.028	0.007

Fuente: Elaboración Propia, 2015.

La estadística descriptiva de las variables que se utilizarán en el modelo de regresión se describe en las Tablas 4 y 5.

**Tabla 4**  
**Estadística descriptiva de las variables continuas**

Variable	Media	Desviación estándar	Mínimo	Máximo
<i>prep_dir</i>	17	2.7016	12	22
<i>tarifa_dom</i>	5.2810	4.0094	0.42	19.85
<i>prec_mfluv</i>	1135.1	466.9	107.2	2110.1

Fuente: Elaboración Propia, 2015.

Nota: los valores negativos de las variables dependientes *l\_1*, representan a una eficiencia mayor en contra parte de los valores positivos.

**Tabla 5**  
**Estadística descriptiva de las variables binarias**

Variable	Media	Desviación estándar	Mínimo	Máximo
<i>Reg</i>	0.8043	0.0652	0.00	1.00
<i>Cont</i>	0.3261	0.1210	0.00	1.00
<i>certificado</i>	0.2391	0.1286	0.00	1.00
<i>tip_tarif</i>	0.9348	0.0377	0.00	1.00
<i>Zd</i>	0.2826	0.1249	0.00	1.00

Fuente: Elaboración Propia, 2015.

#### 4.2.1. ANÁLISIS GRÁFICO

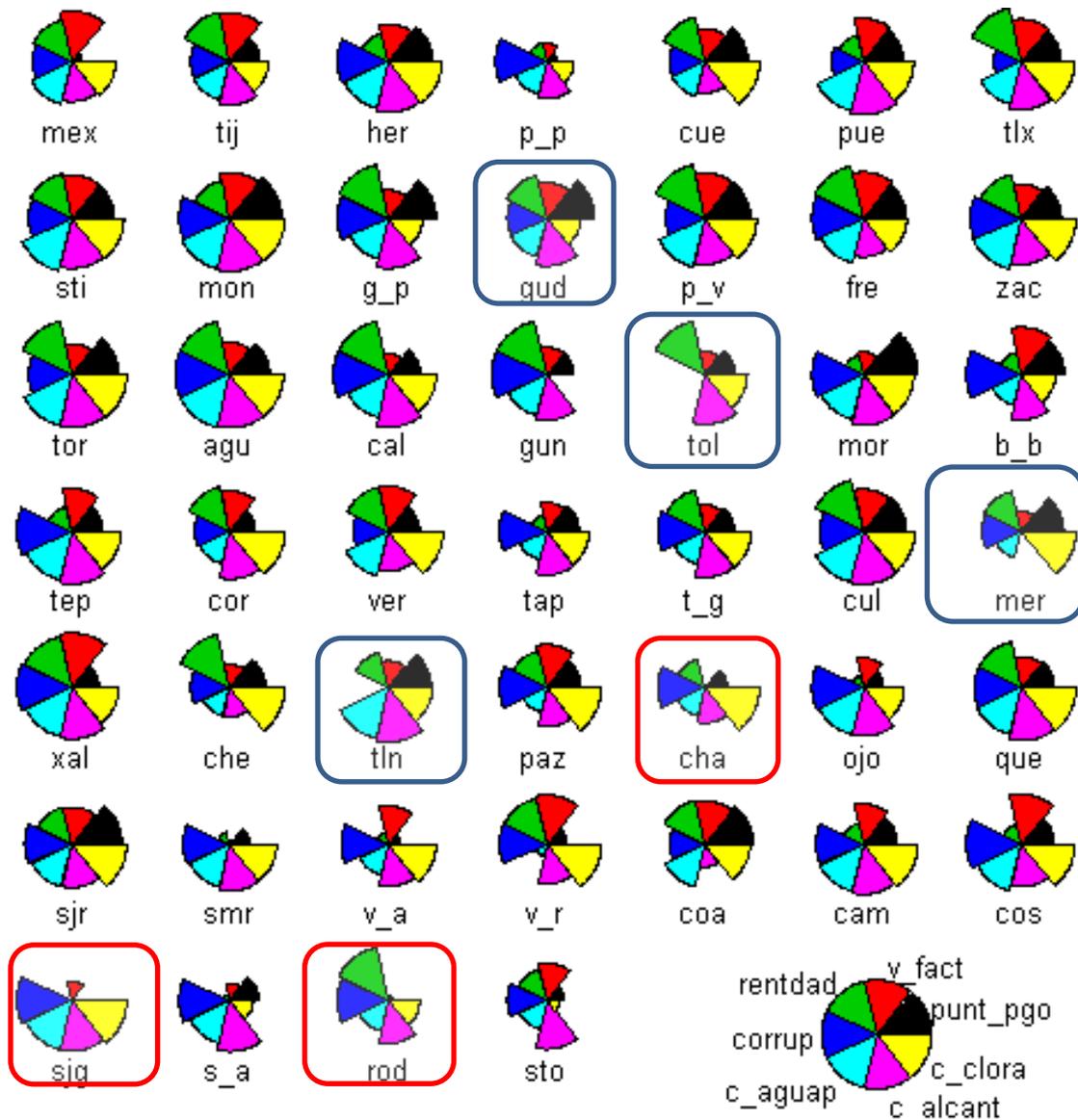
En la Gráfica 2 se observa un análisis de los datos a través de la herramienta de estrellas, en donde cada empresa está representada por una estrella, y cada color simboliza una característica medible del desempeño de la empresa. Esperaríamos que las estrellas estén completas y que el tamaño de los lados sea muy parecido, sin embargo, pocas representaciones de las empresas de agua se encuentran completas en el abanico de colores, hay empresas que carecen de una o varias características, es

el caso de las empresas que se encuentran en Toluca (tol), Mérida (mer), Guanajuato (gun), Tlalnepantla (tln) y Rodeo (rod) o bien que la proporción entre las características es demasiado heterogénea como sucede con las empresas que se encuentran en Puerto Peñasco (p\_p), Morelia (mor), Ojocaliente (ojo), Santa María del Río (smr), Villa de Arista (v\_a) y Sain Alto (s\_a).

En el gráfico se observan enmarcadas con un recuadro rojo a las empresas que les falta una o más características y enmarcadas de azul a las empresas que llaman la atención por la disparidad entre sus características. En este último grupo se observa el caso de Toluca (tol) que tiene ausentes elementos dos corrupción y cobertura de agua potable, Guanajuato (gun) en donde no aparece de cobertura de cloración, Mérida (mer) que carece de cobertura de alcantarillado y Tlalnepantla (tln) en donde no aparece en corrupción. Sin embargo para cada uno de estos datos existe una explicación, en el caso de Toluca la cobertura de agua potable es tan baja y la corrupción alta, en comparación con las otras empresas, que no aparece en el gráfico, para Guanajuato es el mismo caso, la cobertura de cloración es tan baja que no aparece en el gráfico, Mérida (mer) tiene problemas en el suelo cavernoso así que su cobertura de alcantarillado es muy baja y Tlalnepantla, que reporta corrupción alta dada la calificación de Transparencia Internacional.

Por otro lado, las empresas enmarcadas en rojo son aquellas que en donde el valor mínimo de la característica es cero y no existe razón suficiente que explique esa condición; es el caso de la empresa Champoton (cha) que declara cero en la variable volumen de facturación, lo que quiere decir que no se cobra a nadie el servicio de agua que ofrece esta empresa, para San José de Gracia (sjg) el problema es más grave, declara rentabilidad y puntualidad de pago cero, lo cual es muy extraño que ningún usuario pague a tiempo y que la empresa trabaje con números rojos y Rodeo (rod) declara que no le cobra a nadie por el servicio, esto es que su volumen de facturación es cero y que la puntualidad de pago es muy baja y sin embargo, tienen una rentabilidad perfecta (de uno). Estas últimas tres empresas están enmarcadas de rojo porque dada la imposibilidad de los datos reportados, se decidió imputarlos.

**Gráfica 2**  
**Comparación de las empresas de agua a través del gráfico de estrellas**



Nota: Negro = punt\_pgo, Rojo = v\_fact, Verde = rentdad, Azul= corrup, Verde agua= c\_aguap y Rosa = c\_alcant.  
 Fuente: Elaboración Propia, 2015.

### 4.3. ANÁLISIS DE LAS VARIANZAS Y CORRELACIÓN DE LAS VARIABLES DEL ÍNDICE

Usando la matriz de varianza y covarianza (Véase Tabla 6) se observa en la diagonal principal la varianza para cada variable del índice de desempeño. Así las mayores varianzas en la matriz de datos se encuentran en las variables *rentdad*, *punt\_pgo* y *v\_fact* (0.073, 0.048 y 0.044), mientras que las varianzas más bajas se representan por *corrup*, *c\_aguap* y *c\_clora* (0.001, 0.003 y 0.007).

**Tabla 6**

**Matriz de varianzas de los siete indicadores**

	punt_pgo	v_fact	Rentdad	corrup	c_aguap	c_alcant	c_clora
punt_pgo	<b>0.048</b>	0.014	0.020	-0.001	0.004	0.001	0.005
v_fact	0.014	<b>0.044</b>	0.013	0.000	0.003	0.005	0.003
rentdad	0.020	0.013	<b>0.073</b>	-0.002	-0.001	-0.003	0.001
corrup	-0.001	0.000	-0.002	<b>0.001</b>	0.000	0.000	0.000
c_aguap	0.004	0.003	-0.001	0.000	<b>0.003</b>	0.004	0.001
c_alcant	0.001	0.005	-0.003	0.000	0.004	<b>0.029</b>	-0.001
c_clora	0.005	0.003	0.001	0.000	0.001	-0.001	<b>0.007</b>

Fuente: Elaboración Propia, 2015.

El coeficiente de correlación es la cuantificación del grado de asociación en que dos variables continuas se relacionan suponiendo que la relación es lineal, su valor oscila entre 1 y -1. Una característica del coeficiente de correlación es que es independiente a la unidad utilizada en la medición de las variables.

En la Tabla 7 se observa que los coeficientes de correlación son bajos, sin embargo en particular, los valores más altos de la matriz se encuentran con la variable *punt\_pago* y las variables *rentdad* (0.33), *c\_aguap* (0.32), *v\_fact* (0.30) y *c\_clora* (0.28), lo que sugiere una relación positiva baja entre las variables mencionadas. Así, *punt\_pago* será la variable más importante para el análisis posterior.

Por otro lado, existe una correlación positiva entre la variable de cobertura de agua potable y la de cobertura de alcantarillado (0.4), como se esperaba, debido a que los costos fijos en cobertura de agua potable son similares a los costos de la cobertura de alcantarillado. Finalmente, la cobertura de cloración esta levemente correlacionada con la variable de puntualidad de pago (0.28).

Se esperaría que existiera una correlación entre las tres primeras variables del indicador; puntualidad de pago, volumen de facturación y rentabilidad. Lo anterior, debido a que las finanzas sanas de cualquier empresa están directamente relacionadas con el pago a tiempo del producto o servicio que ofrece y con el cobro del mismo, por lo que esperaríamos una calificación similar en las tres variables, sin embargo, el comportamiento diferenciado de las empresas ofrece datos interesantes.

**Tabla 7**

**Matriz de correlaciones de los siete indicadores**

	punt_pgo	v_fact	Rentdad	Corrup	c_aguap	c_alcant	c_clora
punt_pgo	1.00	0.30	0.33	-0.16	0.32	0.02	0.28
v_fact	0.30	1.00	0.23	0.00	0.23	0.13	0.19
rentdad	0.33	0.23	1.00	-0.27	-0.06	-0.07	0.06
corrup	-0.16	0.00	-0.27	1.00	0.29	0.05	-0.06
c_aguap	0.32	0.23	-0.06	0.29	1.00	0.40	0.19
c_alcant	0.02	0.13	-0.07	0.05	0.40	1.00	-0.08
c_clora	0.28	0.19	0.06	-0.06	0.19	-0.08	1.00

Fuente: Elaboración Propia, 2014.

#### 4.4. TRATAMIENTO DE LOS DATOS DE LAS VARIABLES DEL ÍNDICE

Debido a inconsistencias surgidas del análisis anterior se transformó el valor de los datos de tres empresas de agua ubicadas en: San José de Gracia, Champotón y Rodeo. Para la primera empresa se imputó el valor mínimo (después del cero) de la variable *punt\_pgo*, para la empresa ubicada en el municipio de Champotón también se tomó valor mínimo (después del cero) de *v\_fact* y finalmente para la empresas de Rodeo se sustituyó con el valor máximo (antes del uno) de *rentdad*. Las medidas de tendencia central de la imputación se muestran en la Tabla 8.

**Tabla 8**  
**Medidas de tendencia central después de la imputación**

	<i>punt_pgo</i>	<i>v_fact</i>	<i>Rentdad</i>	<i>Corrup</i>	<i>c_aguap</i>	<i>c_alcant</i>	<i>c_clora</i>
<i>Min.</i>	0.02	0.08	0.06	0.81	0.74	0.05	0.60
<i>1st Qu</i>	0.46	0.44	0.46	0.91	0.90	0.79	0.91
<i>Median</i>	0.57	0.60	0.78	0.93	0.95	0.88	0.97
<i>Mean</i>	0.56	0.56	0.69	0.92	0.94	0.84	0.94
<i>3rd Qu</i>	0.73	0.71	0.95	0.94	0.98	0.94	1.00
<i>Max.</i>	0.93	0.90	1.00	0.95	1.00	1.00	1.00

Fuente: Elaboración Propia, 2015.

#### 4.5. RELACIÓN ENTRE VARIABLES

En esta sección se realizará un análisis de las relaciones que existen entre algunas variables que conformarán el índice de eficiencia y así como las variables que caracterizan la administración y finanzas de la empresa y algunos elementos socioeconómicos, que se utilizarán en el modelo de regresión.

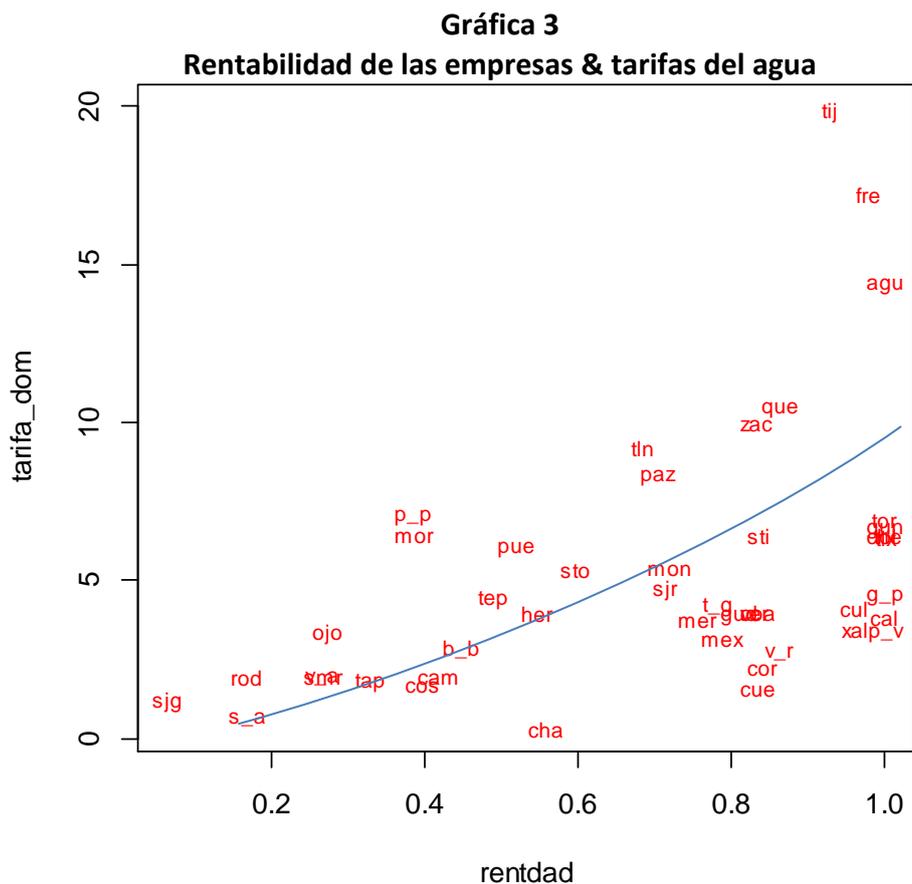
En la Gráfica 3 se observa la relación entre el precio del servicio de agua potable<sup>21</sup>, variable *tarifa\_dom*, y la rentabilidad de las 46 empresas operadoras de agua que se estudian aquí. Partiendo del supuesto que el principio básico de cualquier empresa es mantener el precio del servicio de agua potable, al menos igual al costo de producción del servicio, entonces, si el precio está por encima del costo de producción la empresa es rentable, y si se encuentra por debajo de sus costos de producción la empresa no es rentable.

En contraste con este supuesto se observa en la gráfica que hay empresas de agua, como la ubicada en Rodeo (rod) Durango, la de San Juan de Gracia (sjg) en Aguascalientes y la ubicada en Sain Alto (s\_a) Zacatecas que tienen niveles de

<sup>21</sup> También la llamamos tarifa domiciliaria

rentabilidad menores de 0.2<sup>22</sup>, y en donde el precio de servicio de agua, en estos lugares, es menor a tres pesos.

En el otro extremo de la gráfica se observa que las empresas ubicadas en Tijuana (tij) en Baja California, la de Fresnillo (fre) Zacatecas y la ubicada en Aguascalientes (agu), cabe mencionar que son estas tres empresas las que cobran los mayores precios por el agua, 19 17 y 14 pesos respectivamente, la explicación de lo anterior se estudiará más adelante.

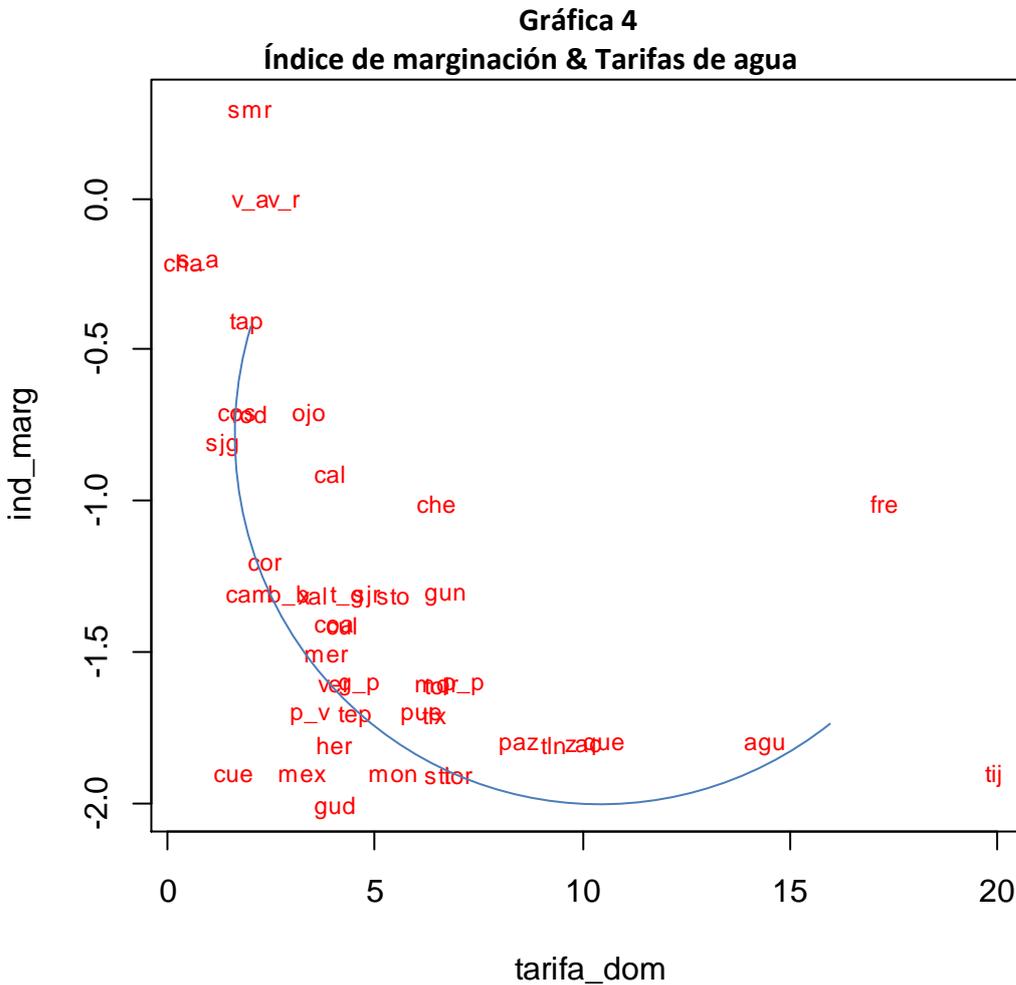


Fuente: Elaboración Propia, 2015.

En la Gráfica 4 observamos una relación inversa entre la variable de marginación, *ind\_marg*, que caracteriza el entorno socioeconómico del municipio en donde se encuentra ubicada la empresa, y la variable tarifa del servicio de agua, *tarifa\_dom*, lo que implica, que a menor marginación del municipio mayores serán las tarifas de agua. Las empresas que encabezan a los municipios con mayores índices de marginación y menores tarifas de agua, son la ubicada en Santa María del Río (smr), Villa de Arista (v\_a) y Villa de Reyes (v\_r) todas en San Luis Potosí. En el extremo inferior derecho, se ubican las empresas con menores niveles de marginación y con precios del agua altos,

<sup>22</sup> De una escala de 0 a 1

como es el caso de las empresas ubicadas en Puebla (pue), Aguascalientes (agu) y Tijuana (tij).



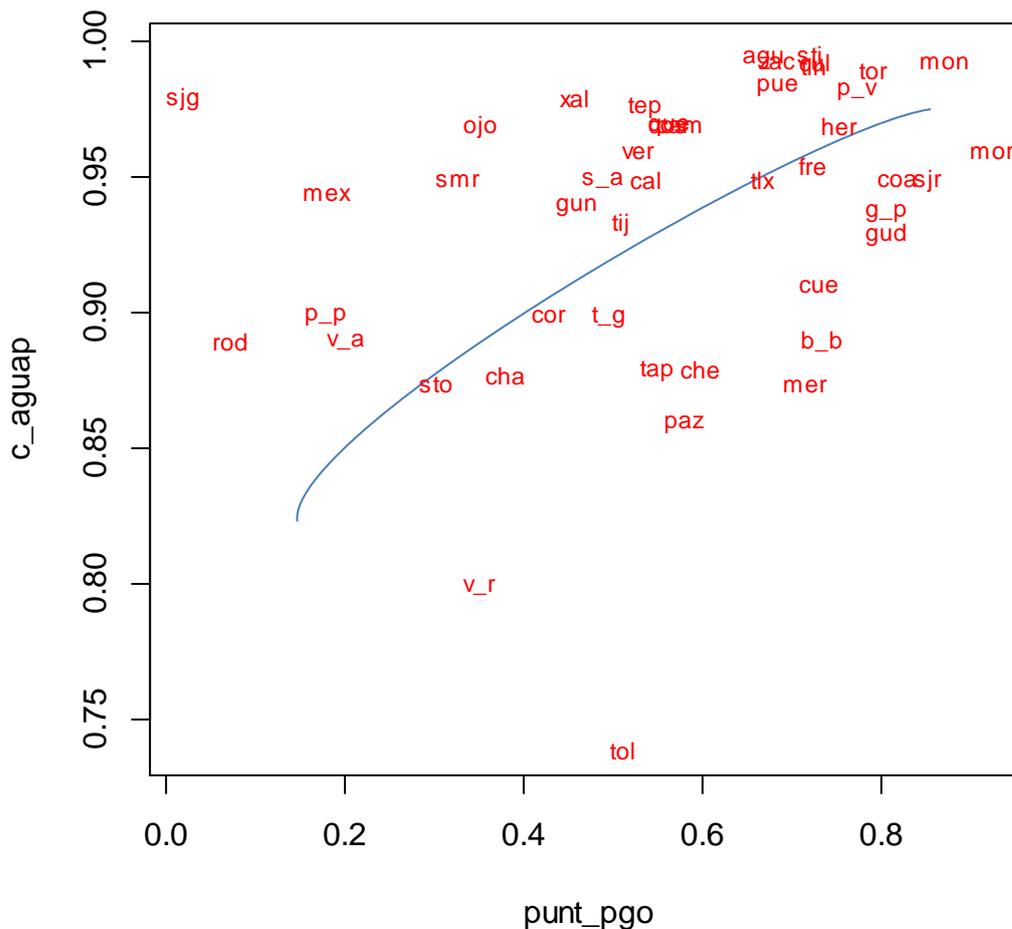
Fuente: Elaboración Propia, 2015.

En la Gráfica 5, se observa la relación positiva entre las variables puntualidad de pago (punt\_pgo) y cobertura de agua potable (c\_aguap), muestra la importancia del pago oportuno por parte de los usuarios del servicio de agua potable para que exista una mayor cobertura del mismo servicio, esto es, que a mayor puntualidad de pago, mayor es la cobertura de agua potable que ofrecen las empresas al municipio en el que se ubican.

Las empresas con mayor cobertura de agua potable y con mayores niveles de puntualidad son las que se ubican en el extremo derecho y arriba de la Gráfica 5, como lo son las ubicadas en Monterrey (mon), Torreón (tor) y Puerto Vallarta (p\_v). Por otro lado, en el extremo contrario, encontramos a las empresas con menores niveles de cobertura de agua potable y de puntualidad de pago, en este caso se encuentra en Villa de Reyes (v\_r), El Salto (sto) en Jalisco y Champotón (cha) en Campeche.

Dos datos *outliers* interesantes son: tol y v\_r, en Toluca (tol) se observa la empresa con menores niveles de cobertura de agua potable, a pesar de los intentos llevados a cabo en los últimos años por aumentar la cobertura, aún no es suficiente para satisfacer la demanda de su población. Por otro lado, Villa de Reyes (v\_r), también tiene una de las coberturas de agua más bajas de nuestra lista de empresas y no obstante también es una de las empresas con peores niveles de puntualidad de pago.

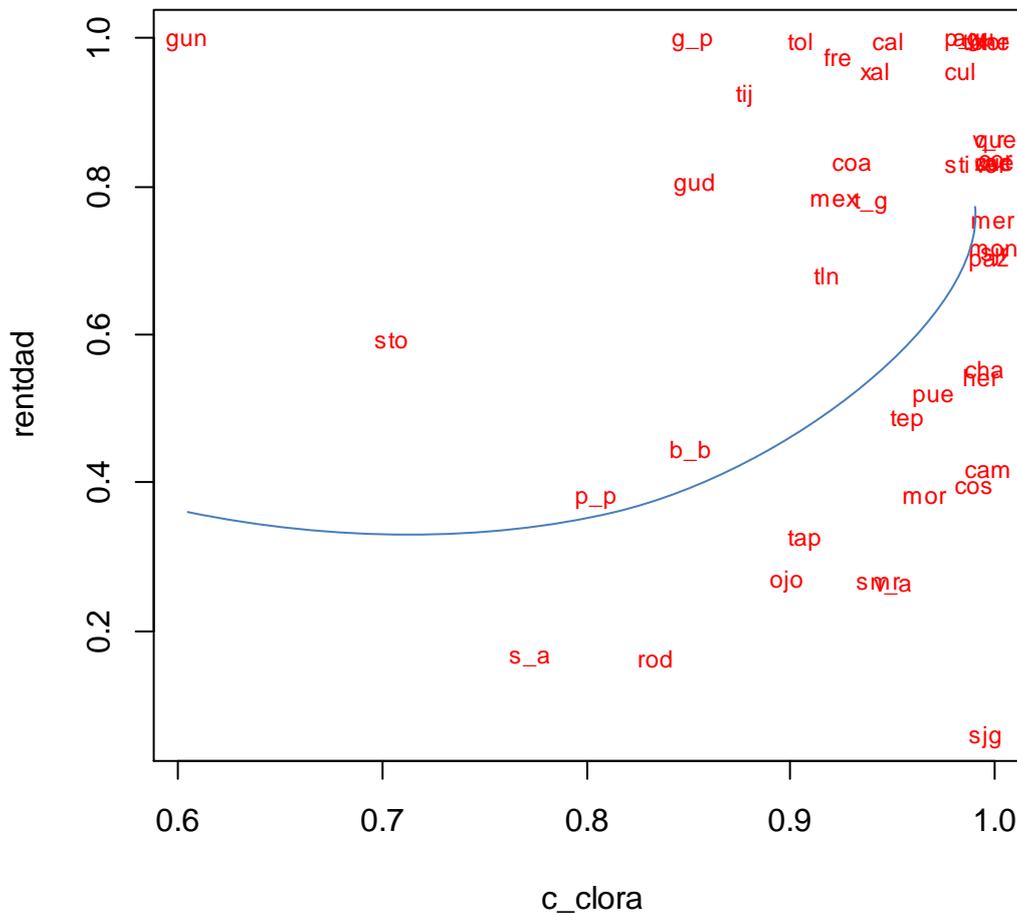
**Gráfica 5**  
**Puntualidad de pago vs cobertura de agua**  
**Puntualidad de pago & Cobertura de agua potable**



Fuente: Elaboración Propia, 2015.

Podemos ver en la Gráfica 6 la relación entre la rentabilidad de las empresas y la calidad del agua que ofrecen esas mismas empresas a los usuarios de los municipios en donde se ubican. Las empresas con niveles mayores de cloración y rentabilidad son, entre otras, Puerto Vallarta (p\_v) y Cuernavaca (cue). Por otro lado, las empresas con menores niveles de rentabilidad y con calidad de agua pobre, son: Sain Alto (s\_a), Rodeo (rod) y Puerto Peñasco (p\_p).

**Gráfica 6**  
**Rentabilidad de la empresa & Calidad del agua**



Fuente: Elaboración Propia, 2015.

#### 4.6. ANALISIS DE COMPONENTES PRINCIPALES

La construcción del índice de desempeño de las empresas que ofrecen el servicio de agua se realizó a través de la metodología de componentes principales, dicho método busca una función que refleje y conserve ciertas particularidades de las siete variables recopiladas para ese fin. La función deberá mostrar la importancia de cada una de las variables y establecer una nueva expresión del problema en menos dimensiones.

El objetivo del análisis de componentes principales es encontrar una nueva variable que es la combinación lineal de *punt\_pgo*, *v\_fact*, *rentdad*, *corrup*, *c\_aguap*, *c\_alcant* y *c\_clora*. Resultado de esa combinación lineal es el índice de desempeño buscado. Para ello se utilizó un Análisis de Componentes Principales usando la matriz de covarianza con la ayuda del lenguaje *R*.

En la Tabla 9 se observa que los cuatro primeros componentes principales explican 95% de la varianza de las siete variables y los tres primeros componentes el 82% de la varianza. Por otro lado, el primer componente principal aporta el 44% de la variabilidad total.

**Tabla 9**  
**Varianza explicada y proporción acumulada por componente principal**

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Desviación estándar	0.300	0.212	0.181	0.162	0.079	0.046	0.025
Proporción de varianza	0.444	0.222	0.160	0.129	0.031	0.010	0.003
Proporción acumulada	0.444	0.666	0.827	0.956	0.987	0.997	1.000

Fuente: Elaboración Propia, 2015.

En los vectores propios se observa que el primer componente principal está compuesto principalmente de las variables puntualidad de pago, volumen facturado y rentabilidad, cabe señalar que son éstas las variables más correlacionadas. La información de estas tres variables juntas podría explicar la administración de las empresas de agua. Cabe señalar que la variable con mayor carga es rentabilidad, *rentdad*.

En la Tabla 10 se observa, en valor absoluto, que en el primer componente la variable rentabilidad es la de mayor peso seguida por las variables puntualidad de pago y volumen facturado. Nótese también, que en términos generales el primer componente es negativo lo que implica que las empresas con mejores puntuaciones en la escala tendrán valores negativos altos.

**Tabla 10**  
**Cuadro de Cargas por componente principal**

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
punt_pgo	-0.48	-0.371	0.774	-0.103	-0.13		
v_fact	-0.383	-0.658	-0.514	0.388			
Rentdad	-0.786	0.553	-0.232	-0.136			
Corrup						0.318	0.947
c_aguap		-0.114		-0.103	0.11	0.93	-0.313
c_alcant		-0.32	-0.272	-0.896		-0.125	
c_clora					0.981	-0.106	
SS loadings	1	1	1	1	1	1	1
Proportion Var	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Cumulative Var	0.143	0.286	0.429	0.571	0.714	0.857	1

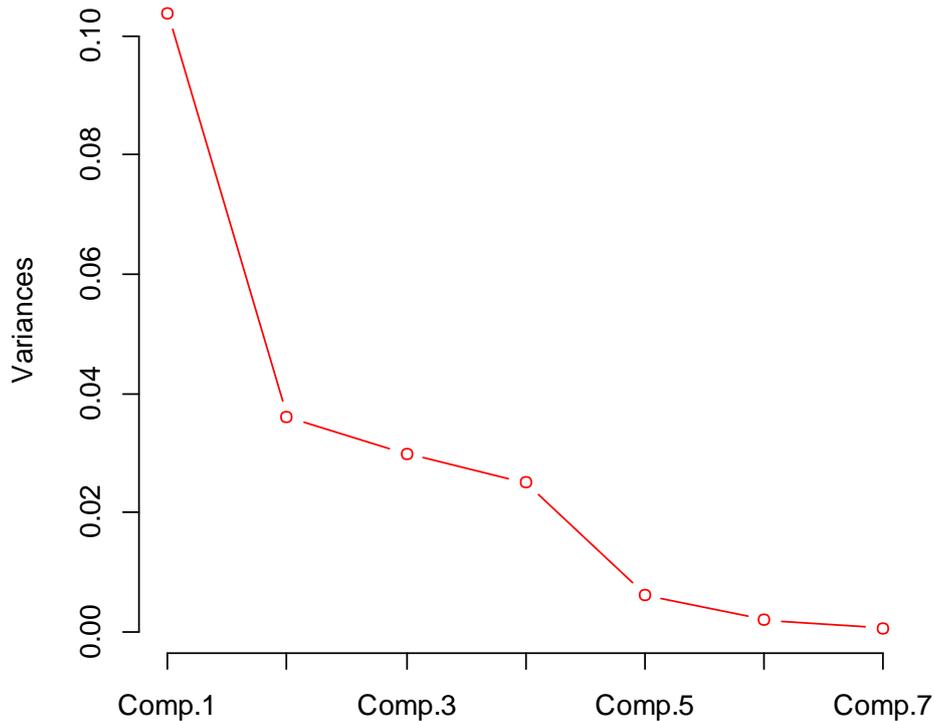
Fuente: Elaboración Propia, 2015.

Nota: los números que no aparecen en la Tabla 10 son tan pequeños que no pueden ser tomados como Ceros.

La gráfica de codo sugiere el número de componentes a considerar. En la Gráfica 7, se observa una súbita caída de la aportación de la varianza, lo que nos deja ver que los

dos codos importantes se encuentran entre el primer y el tercer componente. Hecho que se considerará para la construcción del índice de desempeño.

**Gráfica 7**  
**Variabilidad explicada por componente principal**



Fuente: Elaboración Propia, 2015.

#### 4.6.1 El índice de desempeño

El índice ( $I_1$ ) se define como una función lineal de las características del desempeño de las empresas operadoras de agua. Se conformará del primer componente principal debido a que tiene la mayor capacidad de explicación de los datos, que en este caso alcanza el 44.4% del total. Por otro lado, las variables que conforman el índice son consistentes con la teoría económica, ya que el desempeño de una empresa se refleja en finanzas adecuadas y los indicadores rentabilidad, facturación y puntualidad de pago son muestra de ello. Además, para el administrador de la empresa, el manejo de solo tres variables es más práctico que involucrar un número mayor de variables.

El primer *score* o  $I_1$ , tiene un rango de entre -0.419 y 0.846, en donde el primer valor está representado por la empresa ubicada en Puerto Vallarta y el último la empresa de San José Gracia en Aguascalientes. Debido a los signos de las cargas del primer

componente principal Puerto Vallarta es la empresa, de nuestro estudio, con mayor desempeño en el servicio de agua. En contraparte, San José de Gracia representa a la empresa con el peor desempeño.

Para calificar el desempeño de las empresas operadoras de agua se ordenaron los valores del índice de menor a mayor, como se observa en la Tabla 11, en donde las empresas con mayores índices de desempeño, al menos cuatro de cinco empresas, se ubican en el norte del país. Por otro lado, los menores índices de desempeño pertenecen a cuatro empresas de agua que se ubican en el centro del país, la información completa se puede revisar en la Tabla 1.A. del Anexo.

**Tabla 11**  
**Municipios con mayor y menor índice de desempeño del servicio de agua**

Municipios	Índice (alto)	Municipios	Índice (bajo)
Puerto Vallarta	-0.419	San José de Gracia	0.534
Fresnillo	-0.382	Santa María del Río	0.561
Tlaxcala	-0.352	Sain Alto	0.626
Culiacán	-0.348	Puerto Peñasco	0.55
Torreón	-0.344	Villa de Arista	0.50

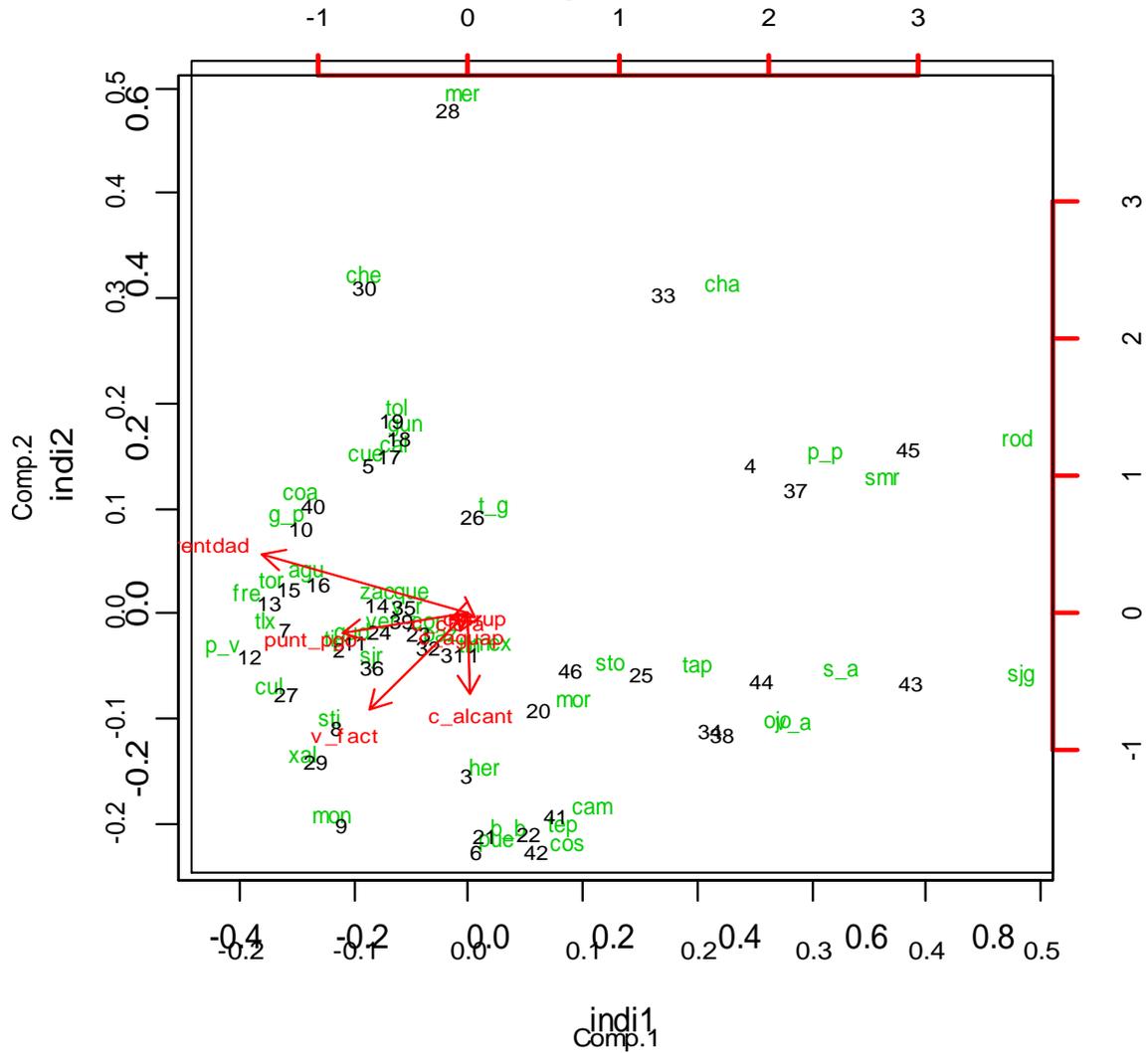
Fuente: Elaboración Propia, 2015.

Aprovechando el ejercicio anterior se le aplicó una tasa de crecimiento al índice de eficiencia ordenado, con el fin de observar los saltos más representativos y tener una idea empírica de por donde deberían de formarse los grupos, los rangos formados fueron los mostrados en la siguiente tabla. Para observar las tasas de crecimiento revisar la Tabla 1.A. del Anexo.

En el *Biplot* mostrado en la Figura 3, se describirá solo el comportamiento del eje de las abscisas por ser éste el eje que representa al índice de eficiencia. Como se observa, las variables rentabilidad (*rentdad*), puntualidad de pago (*punt\_pg*) y volumen de facturación (*v\_fact*) son las variables con mayor longitud de vector, lo que indica el peso o la importancia que estas variables tienen en la construcción del índice, siendo *rentdad* el componente más representativo ubicando a la izquierda a las empresas más rentables, que en este caso sería Puerto Vallarta y a la menos rentable San José de Gracia.

Es importante notar que las empresas con mayor eficiencia se encuentran a la izquierda y abajo de la gráfica en donde los valores de ese cuadrante son negativos. De tal forma que la empresa con mayor eficiencia es la que se encuentra más hacia la izquierda y la peor será aquella que se encuentre más pegada a la derecha.

**Figura 3**  
**Biplot de las 46 empresas**  
**Indicadores del indicador corregido de eficiencia en 46 municipios**



Fuente: Elaboración Propia, 2015.

Rango de empresas con eficiencia alta	Rango de empresas con eficiencia media	Rango de empresas con ineficiencia
-0.42 a -0.03	-0.01 a 0.33	0.37 a 0.85

#### 4.7. ANALISIS DE CONGLOMERADOS

Uno de los objetivos de esta investigación es clasificar al grupo de empresas de agua estudiadas, por lo que, después de la construcción del índice se clasificará por categorías a las empresas. La idea de las categorías es conservar la mayor homogeneidad posible dentro del grupo y a su vez resaltar la heterogeneidad entre grupos.

Para la primera etapa del análisis de conglomerados se consideró al índice de eficiencia, el primer componente, aplicado los siguientes métodos: *Average*, *Centroide*, *Single*, *Complete*, *Ward* (los dendogramas se encuentran en el Anexo Figura 3.A a Figura 12.A) y *K-Means* y se probó con cinco y tres grupos respectivamente.

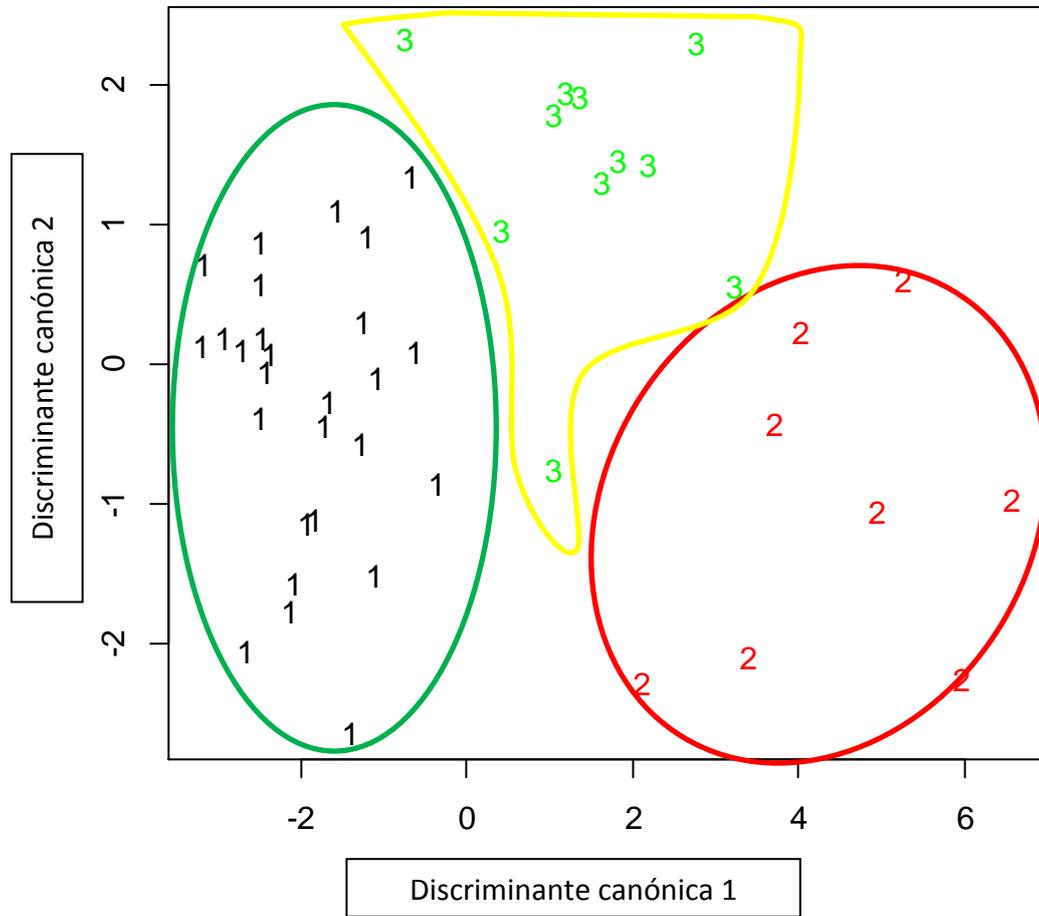
Los resultados de dichos métodos se compararon entre sí a través de la salida del primer componente principal, que se tomó como el índice de desempeño de las empresas, se ordenó de menor a mayor y se identificaron los saltos más representativos, a través de la tasa de crecimiento del índice, para visualizar de manera empírica a los primeros grupos, los resultados de la comparación se observan en la Tabla 2.A y Tabla 3.A. del anexo.

Comparando los resultados de las diversas metodologías de *cluster*, se observó que los mejores agrupamientos se encuentran en los métodos de *Ward* y *K-mean*, como se deja ver en los dendogramas del Anexo. Para determinar entre estos dos métodos, cuál es el que mejor agrupa a las empresas, se realizó un nuevo análisis, pero ahora, se consideraron a los tres primeros vectores de los componentes principales, que explican el 83% de la variabilidad, para realizar un nuevo *cluster* con tres grupos.

En la Figura 4 se observan los tres grupos que generó el *cluster* con el método de *Ward* considerando los tres primeros componentes principales, el primer cuadro, de izquierda a derecha, representa a las empresas con peores niveles de eficiencia, el cuadro del centro representa a las empresas con mediana eficiencia y el cuadro derecho representa a las empresas con mayor eficiencia, tomando la información del índice de eficiencia.



**Figura 5**  
**Cluster K- mean de los tres primeros componentes principales**



Fuente: Elaboración Propia, 2015.

Para realizar la comparación entre ambos métodos se unificaron los grupos y se contrastó con el *score* que se usará como índice de eficiencia. El primer grupo aglomera a las empresas de agua con niveles de eficiencia alta, el grupo dos nos muestra a aquellas empresas con un mediano nivel de eficiencia y finalmente el grupo tres conglomerar a las empresas con niveles de ineficiencia, ver Tabla 12.

**Tabla 12**  
**Resultados de algunos métodos del análisis de *Cluster***

Índice de eficiencia	7 var 3 CP'S		CP1	7 var 3 CP'S	
	<i>Ward</i>	<i>K-mean</i>		<i>Ward</i>	<i>K-mean</i>
-0.419	p_v	p_v	-0.097	Cor	cor
-0.382	fre	Fre	-0.075	Paz	paz
-0.352	tlx	Tlx	-0.039	Mer	mer
-0.348	cul	Cul	-0.031	Tln	tln
-0.344	tor	Tor	-0.006	Her	her
-0.319	g_p	g_p	0.007	mex	mex
-0.297	coa	Coa	0.008	t_g	t_g
-0.293	xal	Xal	0.014	pue	pue
-0.287	agu	Agu	0.033	b_b	b_b
-0.253	sti	Sti	0.117	tep	tep
-0.249	tij	Tij	0.128	cos	cos
-0.247	mon	Mon	0.136	mor	mor
-0.215	gud	Gud	0.165	cam	cam
-0.197	che	Che	0.193	sto	sto
-0.192	cue	Cue	0.331	tap	tap
-0.184	sjr	Sjr	0.370	cha	cha
-0.176	zac	Zac	0.460	ojo	ojo
-0.170	ver	Ver	0.483	v_a	v_a
-0.151	cal	Cal	0.534	p_p	p_p
-0.145	tol	Tol	0.561	s_a	s_a
-0.131	gun	Gun	0.626	Smr	smr
-0.128	v_r	v_r	0.838	Rod	rod
-0.122	que	Que	0.846	Sjg	sjg

	Grupo 1		Grupo 2		Grupo 3
--	---------	--	---------	--	---------

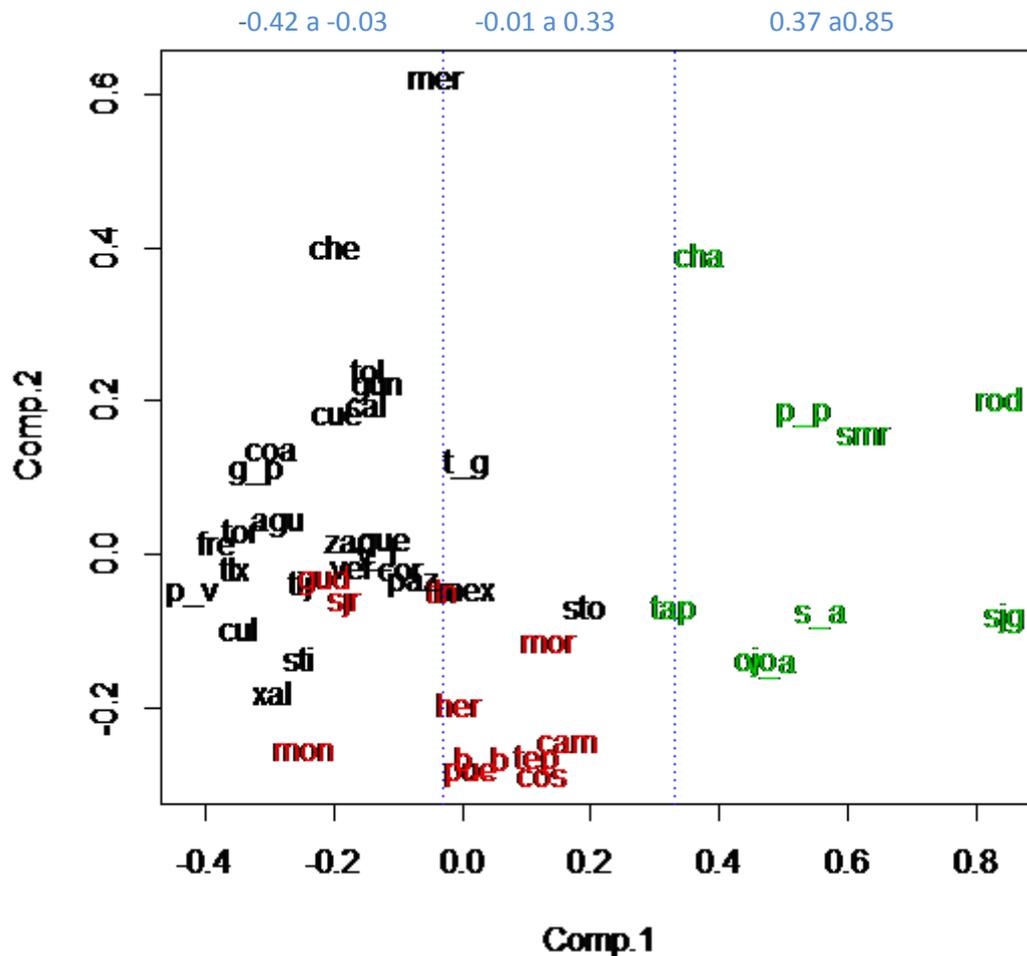
Fuente: Elaboración Propia, 2015.

Como se observa en la Tabla 12 existen algunas empresas que no terminan por definir el grupo al que pertenecen, si comparamos el resultado del *cluster* con el índice de eficiencia. Con la clasificación de *Ward* son seis las empresas (mon, gud, sjr, tln, her, sto) que se encuentran entre el primero y el segundo grupo, así mismo, con el método de *K-mean* son tres las empresas (mon, tln, her) que pertenecen al grupo uno y dos, sin embargo ambos métodos coinciden en agrupar a las empresas ineficientes, en el grupo 3.

La Gráfica 8, se observa en los ejes a los dos primeros componentes principales y cada punto representa una empresa derivada del resultado del análisis de *Cluster* con los tres primeros componentes principales utilizando el método de *Ward* con 3 grupos, para mostrar los límites entre cada segmento se utilizó los rangos que se observan en los saltos de la tasa de crecimiento del índice de eficiencia. Los resultados del primero y segundo componente principal, dejan ver que tres empresas del grupo 1 (Mexicali, El salto y Tuxtla Gutiérrez) clasificadas como empresas eficientes por el primer componente, se ubican físicamente en el grupo 2 o grupo de eficiencia media, y que tres empresas del grupo 2 (San José de Gracia, Monterrey y Guadalajara) se encuentran en el grupo 1, esta distorsión, podría ser causada porque alguna variable

con mayor peso en la segunda componente empuja a la empresa en cuestión a una ubicación errónea. Sin embargo, el grupo 3, de empresas ineficientes se encuentra bien definido, como se observa.

**Gráfica 8**  
**Índice de eficiencia agrupado por el método Ward**

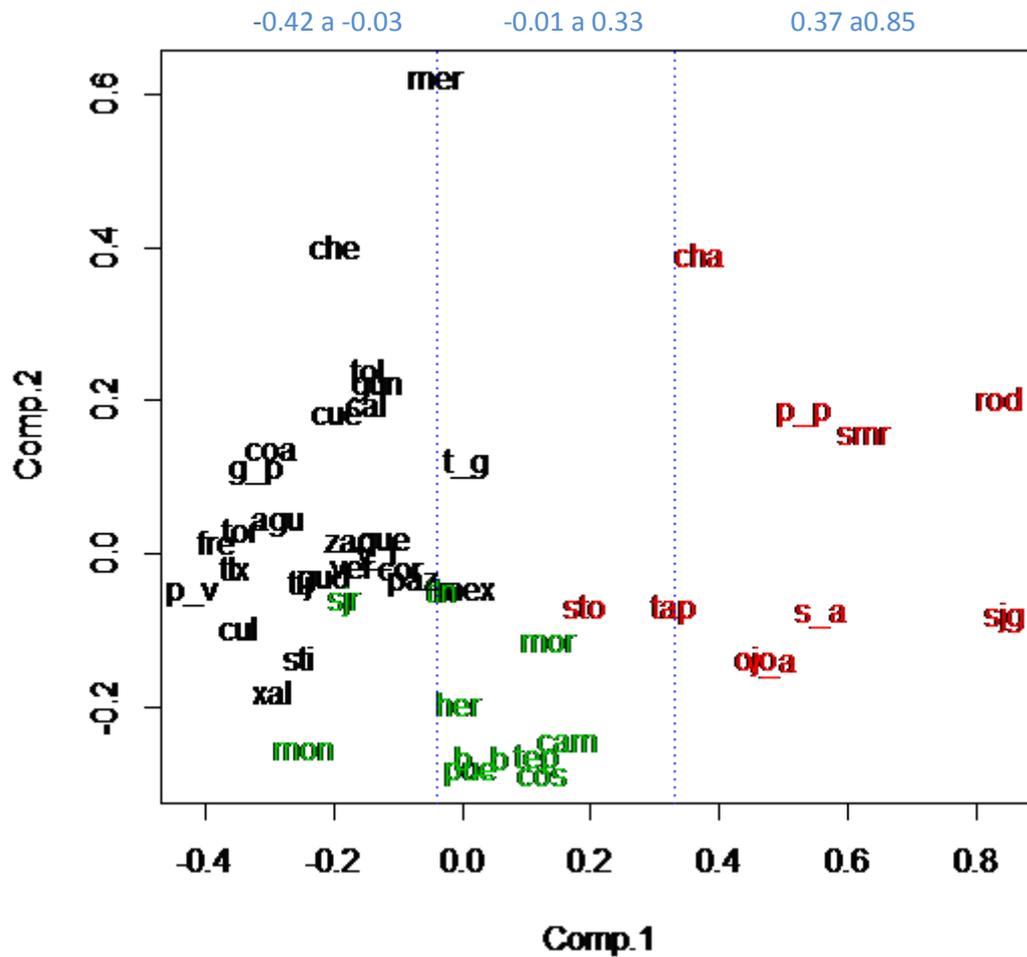


Nota: Cada color representa un grupo resultado del análisis de Cluster.  
 En este caso el programa por default nombra al Grupo1, de color negro, como las empresas eficientes, al Grupo 2, de color rojo, como las empresas Regulares y al Grupo 3, de color verde, como las empresas ineficientes.  
 Fuente: Elaboración Propia, 2015.

La Gráfica 9, representa a las empresas ordenadas por el método de *K-mean* dado el primero y segundo componente principal, en donde el primer grupo de izquierda a derecha representa al grupo con un desempeño mayor, en el grupo del medio (los regulares) se ubican 10 empresas de las cuales, tres de ellas están mal ubicadas, aquellas de color diferente al rojo (Tuxtla Gutiérrez, El Salto y Mexicali), y las empresas

con más bajos desempeños son nueve y se encuentran representadas por el grupo de la derecha, color verde.

**Gráfica 9**  
**Índice de eficiencia agrupado por el método *K-mean***



Nota: Cada color representa un grupo resultado del análisis de Cluster.  
 En este caso el programa por default nombra al Grupo1, de color negro, como las empresas eficientes, al Grupo 2, de color rojo, como las empresas Regulares y al Grupo 3, de color verde, como las empresas ineficientes.  
 Fuente: Elaboración Propia, 2015.

El resumen de este apartado de análisis de *cluster* es: se realizaron seis métodos diferentes de *cluster*, de estos, se comparó entre ellos y se consideró que los mejores agrupamientos son los de los métodos *Ward* y *K-mean*. Para determinar qué método se usaría para clasificar la eficiencia de las empresas de agua, se realizó un nuevo ejercicio de *cluster* con los tres primeros vectores de los componentes principales, sin embargo, se observa que los dos métodos tienen ciertos problemas para clasificar al

menos a tres empresas Monterrey (mon), Tlanepantla (tln) y Hermosillo (her). Aparentemente, el método de *K-mean* agrupa mejor que el método de *Ward*, sin embargo, para asegurar una correcta clasificación de las empresas se realizará un Análisis Discriminante en la siguiente sección.

#### 4.8. ANALISIS DISCRIMINANTE

En el Análisis de *Cluster* que se realizó en la sección anterior se identificó a los dos métodos que mejor agrupan a las empresas de agua según su eficiencia. Por ello, el objetivo de esta sección es seleccionar un método de *Cluster* que represente mejor la clasificación de las empresas a través de la técnica del Análisis Discriminante.

El Análisis Discriminante se puede considerar como un análisis de regresión donde la variable dependiente es categórica y tiene como categorías a los grupos formados en el análisis de *Cluster*, y las variables independientes son continuas y determinan a qué grupos pertenecen los objetos, para ello se utilizaron las siete variables que conforman el índice de eficiencia. En esta sección se pretende encontrar relaciones lineales entre las variables continuas que mejor discriminen a los grupos.

Para realizar este ejercicio se tomaron las siete variables del índice de eficiencia y se caracterizó con los grupos formados por el método de *Ward* y el método de *K-means*, los datos se observan en la Tabla 13, en donde el grupo 1 representa a las empresas con más altos grados de eficiencia, el grupo 2 es aquel en donde las empresas tienen una mediana eficiencia y finalmente el grupo 3 muestra a aquellas empresas con ineficiencia.

**Tabla 13**  
**Las observaciones de las variables del índice y los grupos formados**

empresa	punt_pago	v_fact	rentdad	Corrup	c_aguap	c_alcant	Grupo <i>Ward</i>	Grupo <i>K means</i>
mex	0.177	0.834	0.785	0.912	0.945	0.75	1	1
tij	0.508	0.801	0.927	0.912	0.934	0.806	1	1
her	0.754	0.616	0.547	0.94	0.97	0.94	2	2
p_p	0.177	0.301	0.383	0.94	0.9	0.7	3	3
cue	0.731	0.538	0.835	0.902	0.911	0.6	1	1
pue	0.683	0.713	0.518	0.89	0.984	0.982	2	2
tlx	0.666	0.699	1	0.883	0.95	0.89	1	1
sti	0.719	0.714	0.835	0.916	0.996	0.934	1	1
mon	0.87	0.747	0.718	0.94	0.993	0.979	2	2

g_p	0.805	0.461	1	0.935	0.938	0.929	1	1
gud	0.806	0.584	0.808	0.912	0.93	0.89	2	1
p_v	0.771	0.741	1	0.912	0.983	0.863	1	1
fre	0.725	0.76	0.979	0.941	0.955	0.728	1	1
zac	0.683	0.558	0.833	0.941	0.993	0.86	1	1
tor	0.793	0.513	1	0.916	0.99	0.98	1	1
agu	0.668	0.526	1	0.953	0.995	0.981	1	1
cal	0.537	0.342	1	0.953	0.95	0.97	1	1
gun	0.458	0.451	1	0.949	0.94	0.85	1	1
tol	0.511	0.375	1	0.812	0.74	0.92	1	1
mor	0.925	0.365	0.385	0.943	0.96	0.94	2	2
b_b	0.733	0.78	0.446	0.948	0.89	0.835	2	2
tep	0.534	0.699	0.489	0.948	0.977	0.977	2	2
cor	0.428	0.663	0.84	0.903	0.9	0.89	1	1
ver	0.529	0.741	0.831	0.903	0.96	0.75	1	1
tap	0.548	0.465	0.327	0.929	0.88	0.82	3	3
t_g	0.495	0.433	0.782	0.929	0.9	0.86	1	1
cul	0.727	0.694	0.96	0.919	0.993	0.977	1	1
mer	0.715	0.32	0.755	0.911	0.874	0.048	1	1
xal	0.457	0.9	0.961	0.948	0.98	0.95	1	1
che	0.598	0.38	1	0.911	0.88	0.564	1	1
tln	0.725	0.441	0.683	0.812	0.992	1	2	2
paz	0.578	0.706	0.702	0.927	0.86	0.74	1	1
cha	0.38	0.081	0.557	0.928	0.878	0.666	3	3
ojo	0.35	0.487	0.272	0.941	0.97	0.9	3	3
que	0.562	0.509	0.862	0.93	0.97	0.98	1	1
sjr	0.853	0.613	0.713	0.93	0.95	0.81	2	2
smr	0.325	0.081	0.268	0.932	0.95	0.87	3	3
v_a	0.2	0.627	0.265	0.932	0.89	0.79	3	3
v_r	0.35	0.805	0.863	0.932	0.8	0.75	1	1
coa	0.82	0.712	0.834	0.903	0.95	0.45	1	1
cam	0.576	0.656	0.419	0.928	0.97	0.93	2	2
cos	0.562	0.819	0.397	0.953	0.97	0.8	2	2
sjg	0.0177	0.309	0.063	0.953	0.98	0.93	3	3
s_a	0.488	0.286	0.167	0.941	0.95	0.936	3	3
rod	0.071	0.078	0.167	0.935	0.89	0.8	3	3
sto	0.3	0.621	0.596	0.912	0.875	0.9	1	3

Fuente: Elaboración Propia, 2015.

#### 4.8.1. DISCRIMINANTE CON CLASIFICACIÓN DE WARD

Se realizó un Análisis Discriminante lineal de *Fisher* dadas las siete variables que conforman el índice de eficiencia utilizando los resultados obtenidos en el Análisis de *Cluster* con el método de *Ward*. El resultado discriminante de las empresas de agua se observa con las siguientes funciones, es notorio que en la primera función, la variable *rentdad*, rentabilidad, es la que más aporta al método para discriminar.

El análisis discriminante se realizará a tres grupos, por lo tanto, se requieren dos funciones discriminantes que se llamas; *LD1* y *LD2*.

Línea discriminante 1

$$LD1 = 0.656 \text{ puntptgo} - 2.535 \text{ vfact} - 8.039 \text{ rentdad} - 0.938 \text{ corrup} \\ + 0.832 \text{ c\_aguap} + 1.373 \text{ c\_alcant} - 0.067 \text{ c\_clora}$$

Para la segunda función la variable que con mayor peso para discriminar, según el método de *Ward* fue *corrup*, corrupción, como se observa en la siguiente función.

Línea discriminante 2

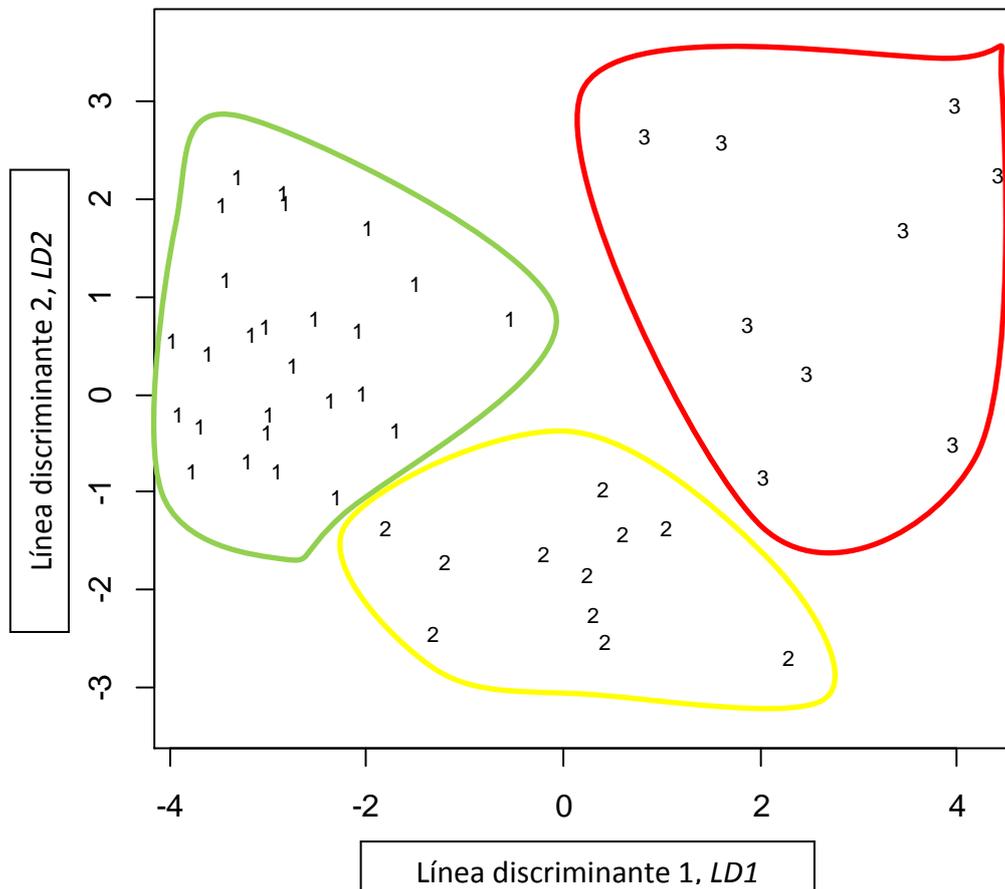
$$LD2 = -6.306 \text{ puntptgo} - 3.297 \text{ vfact} + 3.499 \text{ rentdad} + 6.950 \text{ corrup} \\ + 2.133 \text{ c\_aguap} - 2.266 \text{ c\_alcant} + 0.279 \text{ c\_clora}$$

Por otro lado, se calcularán los casos correctamente clasificados (dentro de la diagonal principal) así como los clasificados incorrectamente, en la siguiente matriz se observa que fuera de la diagonal principal existen solo 4 datos mal clasificados en los grupos eficientes y regulares.

GRUPOS	EFICIENTES	REGULARES	INEFICIENTES
EFICIENTES	25	1	0
REGULARES	2	9	0
INEFICIENTES	0	1	8

En la Gráfica 10 se observa el resultado de la clasificación de las dos funciones que realiza el Análisis Discriminante, dados los grupos formados por el *cluster* de Ward.

**Gráfica 10**  
**Representación gráfica de los tres grupos formados en las funciones discriminantes**



Fuente: Elaboración Propia, 2015.

Por otro lado, para comprobar la capacidad predictiva de las funciones discriminantes<sup>23</sup>, se realizó una validación cruzada y el porcentaje total de clasificación correcta, se puede leer en la Tabla 14. Como se observa, la capacidad predictiva es lo suficientemente aceptable 0.91, así mismo de manera individual el porcentaje de clasificación correcta en los tres niveles, es mayor a 0.8.

<sup>23</sup> Si se requiere se puede revisar en la Tabla 4.A. del Anexo la Probabilidad posterior por empresa

**Tabla 14**  
**Validación Cruzada**

Niveles		
1°	2°	3°
0.962	0.818	0.889
Porcentaje correcto total		
0.9130435		

#### 4.8.2. DISCRIMINANTE CON CLASIFICACIÓN DE K-MEAN

El Análisis Discriminante lineal de *Fisher* se utilizará para evaluar los resultados del *cluster* con el método de *K-mean* se obtienen las funciones discriminantes que se leen enseguida, como se observa la primera función tiene el mismo comportamiento que la función generada por el método *Ward*, en donde la variable *rentdad*, rentabilidad, es quien más aporta al método para discriminar.

Línea discriminante 1

$$LD1 = -0.064656 \text{ punt}pgo - 1.530 \text{ vfact} - 8.346 \text{ rentdad} - 3.506 \text{ corrup} \\ - 0.327 \text{ c\_aguap} + 1.568 \text{ c\_alcant} - 3.354 \text{ c\_clora}$$

Para la segunda función, las variables con mayor peso para discriminar, según el método de *K-mean* utilizado para clasificar son: *corrup* y *punt\_pgo*, corrupción y puntualidad de pago, como se observa en la siguiente función.

Línea discriminante 2

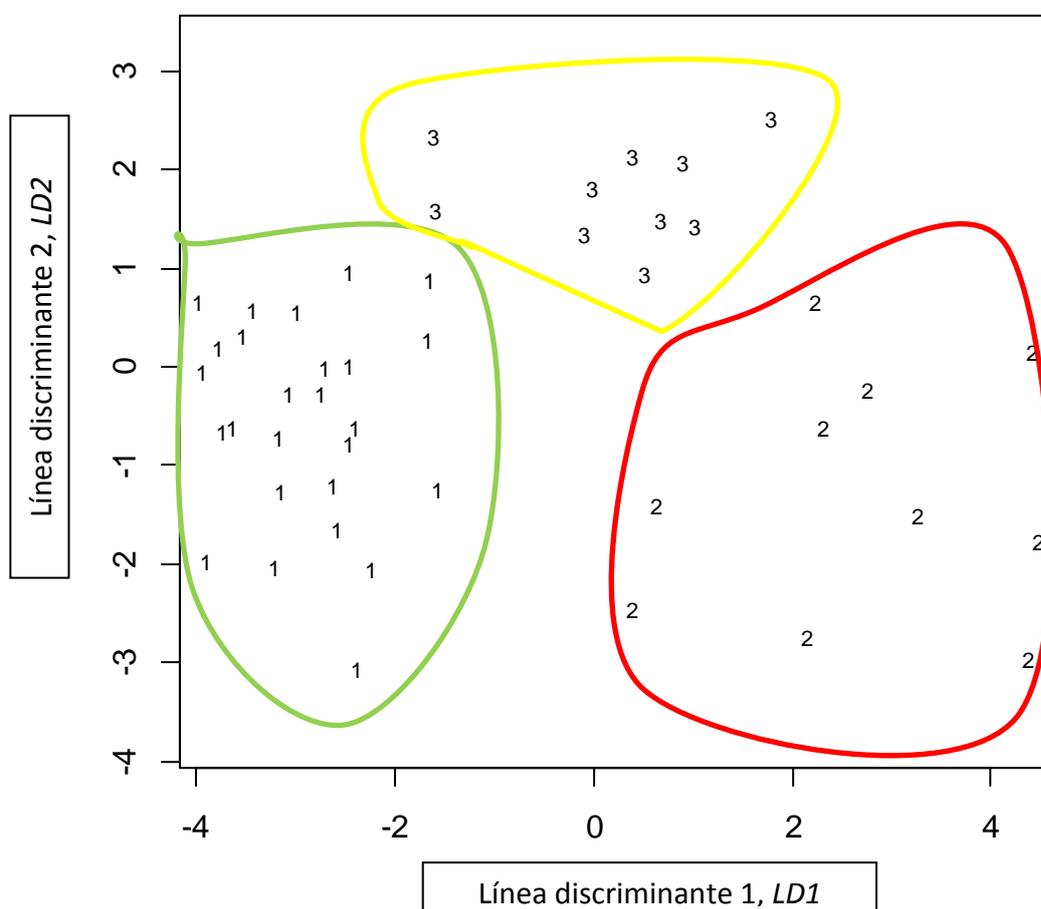
$$LD2 = 5.768 \text{ punt}pgo + 3.114 \text{ vfact} - 3.591 \text{ rentdad} - 6.241 \text{ corrup} \\ - 0.860 \text{ c\_aguap} + 2.066 \text{ c\_alcant} + 2.076 \text{ c\_clora}$$

De igual forma, se calcularán los casos correctamente clasificados (dentro de la diagonal principal) así como los clasificados incorrectamente, en la siguiente matriz se observa que fuera de la diagonal principal existen solo 5 datos mal clasificados, uno más en comparación con el método *Ward*, estos datos se encuentran en los grupos eficientes e ineficientes.

GRUPOS	EFICIENTES	REGULARES	INEFICIENTES
EFICIENTES	26	0	0
REGULARES	2	7	1
INEFICIENTES	2	0	8

En la Gráfica 11 se observa el resultado gráfico de la clasificación de las dos funciones que realiza el Análisis Discriminante dados los grupos formados por el *cluster* de *K-mean*.

**Gráfica 11**  
**Representación gráfica de las tres variables discriminantes canónicas**



Fuente: Elaboración Propia, 2015.

Por otro lado, para comprobar la capacidad predictiva de las funciones discriminantes<sup>24</sup>, se realizó una validación cruzada y el porcentaje total, se puede leer

<sup>24</sup> Si se requiere se puede revisar en la Tabla 5.A del Anexo la Probabilidad posterior por empresa

en la Tabla 15, como se observa, la capacidad predictiva es lo suficientemente aceptable 0.89, sin embargo, es menor al porcentaje total en donde se utilizó una clasificación con el método de *Ward*, aunado a ello, de manera individual el porcentaje de clasificación correcta en los tres niveles es mayor a 0.7.

**Tabla 15**  
**Validación Cruzada**

Niveles		
1°	2°	3°
1	0.7	0.8
Porcentaje correcto total		
0.8913043		

Con los resultados obtenidos de los dos Análisis Discriminantes, se puede concluir que la aglomeración de *Ward* tiene una mejor clasificación en el porcentaje correcto total, fue de 91%, que la aglomeración de *K-means*, en donde su porcentaje correcto total fue de 89%, esto es, que *K-means* tienen más datos mal clasificados, lo que implica que esos datos mal clasificados van a dar clases contiguas, como se observa en el análisis.

#### 4.9. ANALISIS DE EFICIENCIA DE LAS EMPRESAS DE AGUA EN MÉXICO

##### 4.9.1. MODELO DE EFICIENCIA DE LAS EMPRESAS

La variable que representa al sistema legal del agua se llama *Reg*, es binaria y busca capturar a aquellas empresas que cuentan con un reglamento interno que dirija el funcionamiento del servicio.

Las variables del sector administrativo son cuatro y muestran las características más representativas de la administración de las empresas de agua, para ello se investigó la variable *tarifa\_dom*, que es el precio del servicio de agua consumiendo 30 m<sup>3</sup> (consumo promedio de una familia de cuatro personas). La variable que complementa a la información anterior es *tip\_tarif*, que se construyó de forma binaria y en donde la unidad representa a la tarifa que cobra dependiendo del número de metros cúbicos consumidos. La variable *cont* es binaria y la unidad representa la continuidad de los objetivos en la administración. Por otro lado, la variable *certificado* es una variable binaria que caracteriza a las empresas que han invertido en su empresa para que ésta sea certificada por alguna entidad calificadoradora y finalmente la variable *prep\_dir* nos permite observar si hay una relación entre los años estudiados del director de la empresa y la eficiencia de la misma.

Las variables exógenas utilizadas en este estudio fueron de tipo medio ambiental y tienen el objetivo de mostrar que las características geográficas de un lugar, afectan la eficiencia de una empresa de agua. La variable *zd* caracteriza la zona de disponibilidad y se construyó como una variable dicotómica que divide al país en dos zonas geográficas; la zona 1 que caracteriza a la zona de escasas y la zona 0 que describe la menor escasez de agua. En la Figura 13A, del anexo, se pueden ver los 24 estados en donde se encuentran las empresas de agua que están dentro del este estudio, de color morado se caracterizan a las empresas con alta escasez y de verde las empresa con menor escasez.

Así mismo la variable *prec\_mfluv*, es una variable continua que mide la cantidad de agua de lluvia que cae en el municipio en donde se encuentra ubicada la empresa. De manera que el modelo de regresión por ajustar es el siguiente:

$$I_{1i} = \alpha_0 + \alpha_1 Reg_i + \alpha_2 prep\_dir_i + \alpha_3 cont_i + \alpha_4 certificado_i + \alpha_5 tip_{tarif_i} \\ + \alpha_6 tarifa_{dom_i} + \alpha_7 zd_i + \alpha_8 prec\_mfluv_i + U_i$$

#### 4.10 RESULTADOS

En la Tabla 16 se muestran los resultados del análisis de la regresión, en general se observa que el modelo que explica la eficiencia es un modelo adecuado, debido a que el coeficiente  $R^2$  corregida, es de 0.66, lo que implica que la mayor parte de la variabilidad de la variable eficiencia está explicada por las variables seleccionadas del modelo.

Por otro lado, se confirma la relación lineal entre la variable eficiencia y las variables explicativas del modelo, ya que la prueba de significancia de la regresión deja ver que es altamente significativa, *p-value*: 1.257e-07.

De manera particular, todas las betas son significativas para  $\alpha=0.10$  y los resultados por variable, son los siguientes: la variable *Reg* es significativa, lo que implica que las reglas formales como los reglamentos internos de una empresa son importantes para mejorar la eficiencia de una empresa de agua.

Por el lado de las variables administrativas, la variable preparación del director, *prep\_dir*, es significativa y tiene el signo esperado, esto es que a mayor estudio por parte del director de la empresa, mayor será la eficiencia de empresa que ofrece el servicio. Otro de los problemas administrativos importantes es la rápida rotación del personal directivo de las empresas de agua, lo anterior se debe, a que cada tres años existe cambio de administración sin considerar el plan de trabajo previo. La variable

*cont* desea capturar el efecto de la continuidad de los objetivos administrativos y el resultado del modelo nos dice que las empresas con un seguimiento administrativo han tenido eficiencias más altas que las empresas sin continuidad. *Certificado*, es la variable que muestra el efecto de una inversión administrativa por parte de la empresa y el modelo indica que las empresas con algún tipo de certificado tienen eficiencias mayores que aquellas empresas que no invierten en una certificación. Para la variable *tarifa\_dom* el signo es el esperado lo que implica que a precios altos por el servicio de agua, la eficiencia de empresa mejorará. En el caso de la variable *tip\_tarif*, se lee que una tarifa volumétrica, en donde el precio por metro cúbico consumido de agua es diferenciado, es decir, depende de la cantidad de agua consumida por el usuario, incrementa la eficiencia administrativa de la empresa.

Finalmente las variables de tipo ambientales, expresan el comportamiento subjetivo del consumidor ante la escasez, así la variable zona de disponibilidad, *zd*, expresa que la ubicación en donde se encuentra la empresa que abastece el servicio influye en la eficiencia del servicio, si la empresa se ubica en una zona de disponibilidad de agua baja, la eficiencia será mayor que las empresas ubicadas en zonas disponibilidad alta. Así mismo, si la precipitación media fluvial, *prec\_mfluv*, si llueve poco, la eficiencia de la empresa mejora, como es el caso del Norte de la República Mexicana.

**Tabla 16**  
**Resultados del modelo**

Coefficiente	Estimador	Error estándar	t value	Pr(> t )	
(Intercept)	1.75E+00	2.55E-01	6.856	5.87E-08	***
Reg	-4.02E-01	7.65E-02	-5.255	7.42E-06	***
prep_dir	-3.82E-02	1.20E-02	-3.173	0.003135	**
Cont	-1.60E-01	6.28E-02	-2.545	0.015503	*
Certificado	-2.75E-01	7.53E-02	-3.656	0.000833	***
tip_tarif	-4.73E-01	1.73E-01	-2.736	0.009691	**
tarifa_dom	-2.44E-01	8.13E-02	-2.999	0.004958	**
Zd	-4.70E-01	1.91E-01	-2.46	0.018957	*
prec_mfluv	-1.27E-04	7.19E-05	-1.762	0.086774	.
tip_tarif*tarifa_dom	2.30E-01	8.22E-02	2.796	0.008354	**
Zd*prec_mfluv	3.71E-04	1.57E-04	2.365	0.023719	*

Nota: Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1885 on 35 degrees of freedom

Multiple R-squared: 0.7398, Adjusted R-squared: 0.6654

F-statistic: 9.95 on 10 and 35 D, p-value: 1.257e-07

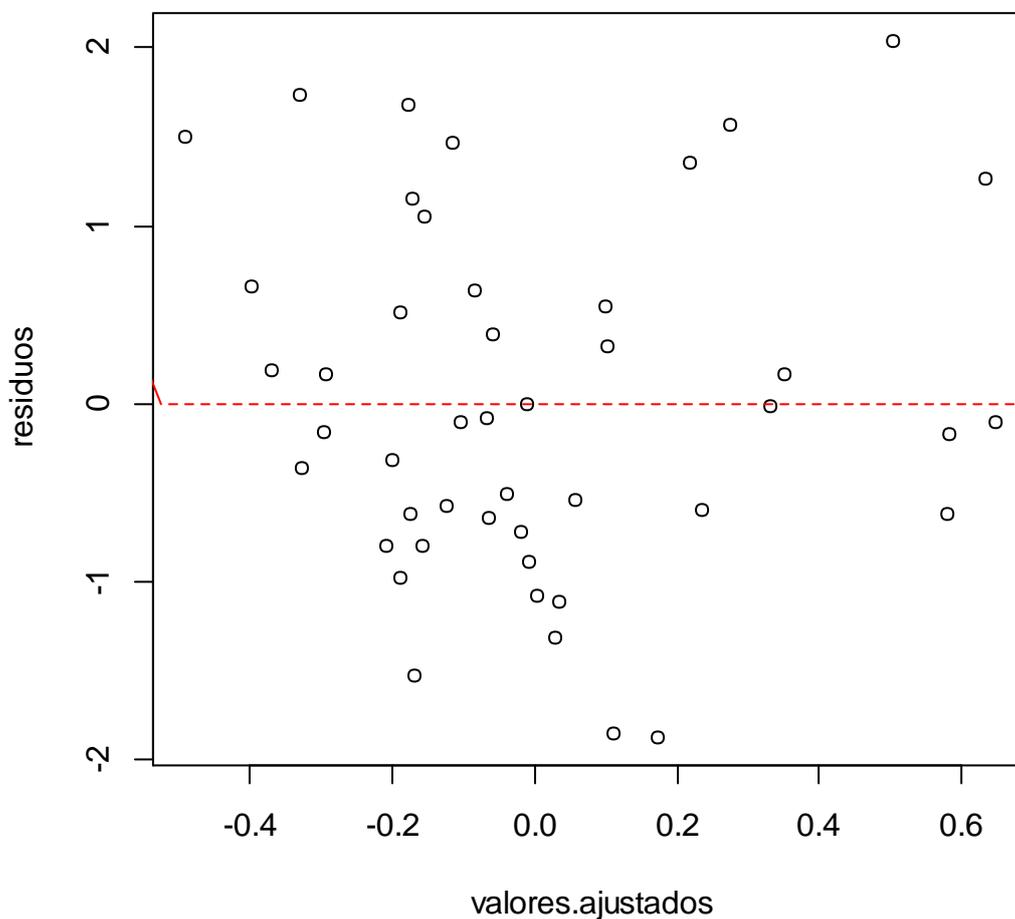
Fuente: Elaboración Propia, 2015.

#### 4.10.1. DIAGNÓSTICO DEL MODELO

Para evaluar si el modelo cumple los supuestos, se analizarán los gráficos de residuales contra los valores ajustados, se revisará la normalidad y se cotejará las observaciones de los valores ajustados contra los valores observados.

El análisis de los valores ajustados y los residuos estandarizados, se puede observar en la Gráfica 12. No hay ningún patrón especial, esto es, que las observaciones se encuentran entre el -2 y el 2 y no tienen ninguna tendencia, por lo que tanto, podemos afirmar que la homoscedasticidad y, la linealidad resultan hipótesis razonables.

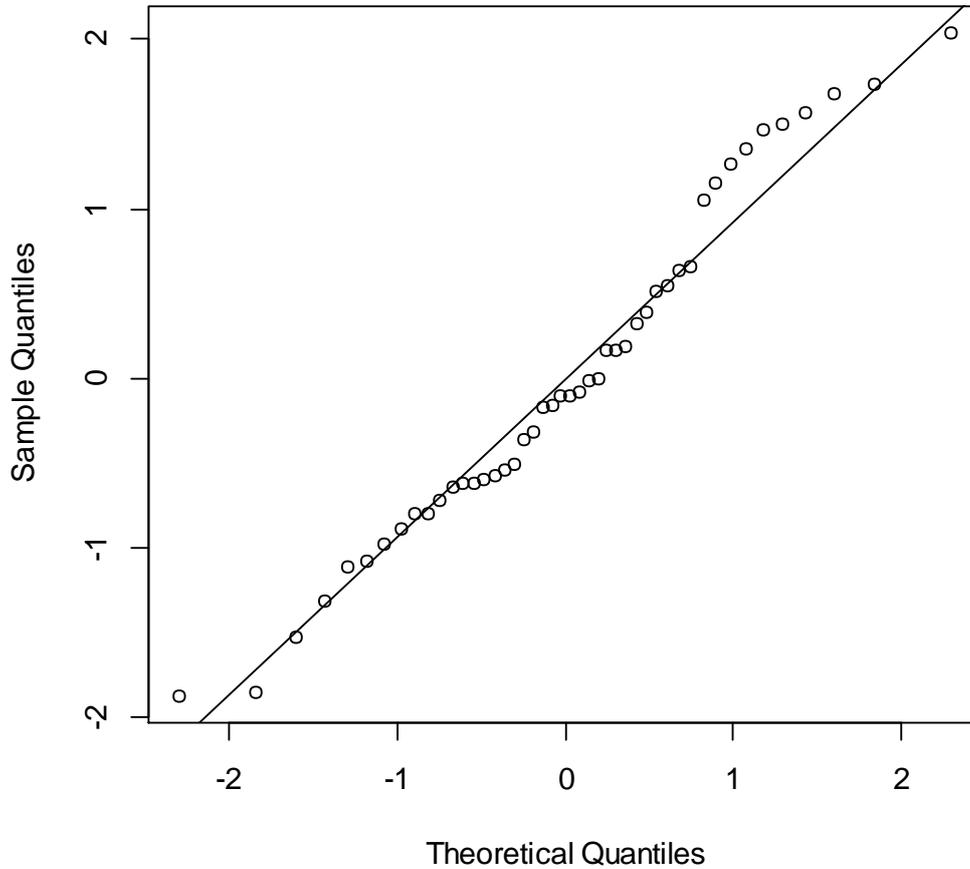
**Gráfica 12**  
**Valores ajustados & residuos**



Fuente: Elaboración Propia, 2015.

Por otro lado, la hipótesis de normalidad se puede verificar mediante un *QQ plot* de los residuales, como se observa en la Gráfica 13

**Gráfica 13**  
**Normal QQ plot**



Fuente: Elaboración Propia, 2015.

Podemos notar que los puntos están bastante alineados en algunas secciones, sin embargo, hay secciones de la gráfica un tanto dudosas, así que consideré conveniente hacer la prueba *jarque.bera.test* del paquete *t-series*, los resultados se muestran en la siguiente tabla.

**Tabla 17**  
**Prueba Jarque-Bera**

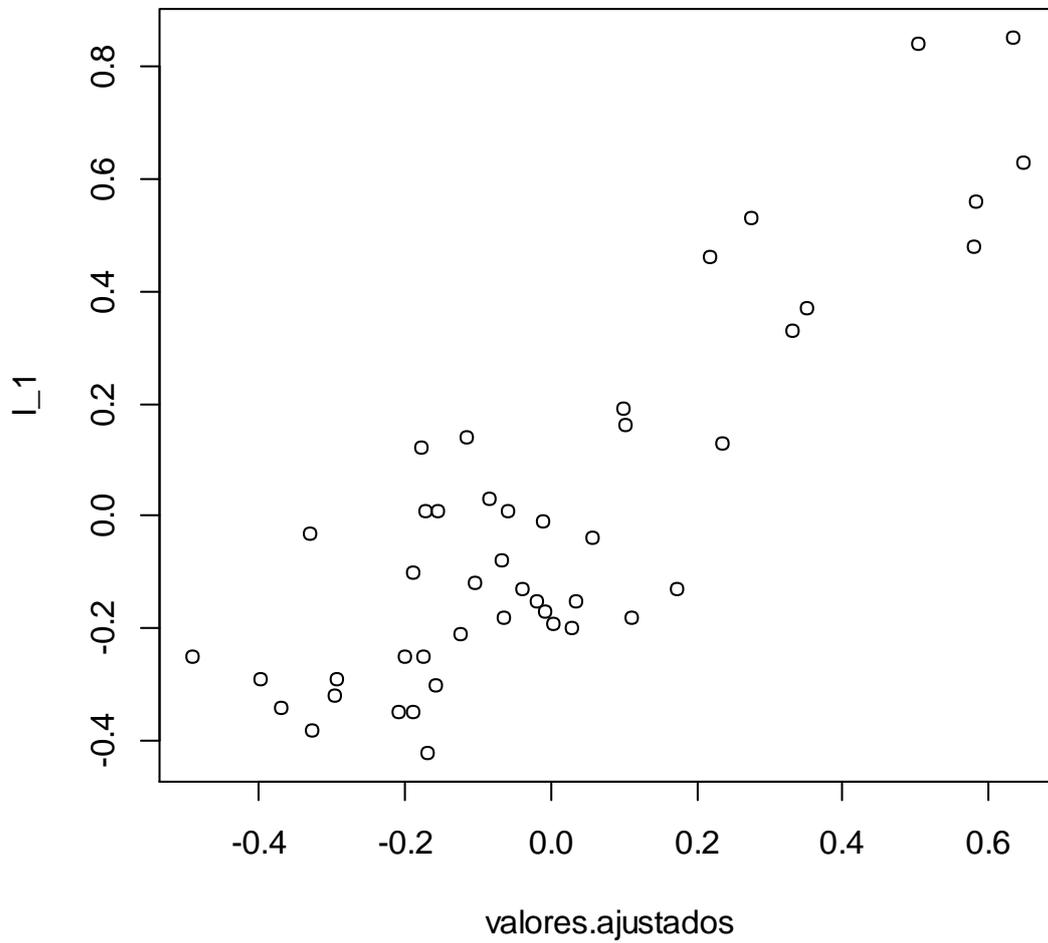
X-squared	Df	p-value
1.3574	2	0.5073

Dados los resultados del test y el QQ plot podemos concluir que la normalidad también es aceptable.

Por otro lado, ilustrará la bondad de ajuste en la Gráfica 14, en donde se observa a la variable explicada y la observada, *I\_1*, es notorio que no existen grandes discrepancias

entre los valores observados y los valores esperados, no se ve ningún outlier, por lo que se concluye que el modelo tiene un ajuste bastante aceptable.

**Gráfica 14**  
 **$Y_i$  vs  $\hat{Y}_i$**



Fuente: Elaboración Propia, 2015.

# CONCLUSIONES

---

La prestación del servicio de agua potable y saneamiento es un tema importante para la salud pública, la equidad social, el desarrollo económico y la sustentabilidad ambiental en cualquier nación.

Debido a la carencia de un parámetro que evalúe la eficiencia de las empresas que ofrecen el servicio de agua en México, el objetivo de este trabajo fue construir un índice que refleje esa eficiencia. Para ello se utilizó una base de datos que contenía 46 empresas operadoras de agua y 15 variables, siete de estas variables se usaron para la construcción del índice ya que éstas caracterizan el servicio de agua y el resto de las variables explicó el comportamiento de las empresas.

La construcción del índice se llevó a cabo, a través de una Análisis de Componentes Principales, los resultados del análisis consideran que las variables clave fueron: la puntualidad de pago, la rentabilidad de una empresa y el volumen de facturación de la misma, estas variables conforman el primer componente y éste explica el 44% de la variabilidad total de la base de datos.

El índice tiene un rango que va de -0.419 a 0.846, en donde el primer valor lo representa la empresa con mayor nivel de eficiencia y el último valor representa a la empresa menos eficiente. Los resultados del índice ubican como a las cinco empresas más eficientes a Puerto Vallarta, Fresnillo, Tlaxcala, Culiacán y Torreón; en contra parte las empresas con eficiencias más bajas son: San José de Gracia, Santa María del Río, Sain Alto, Puerto Peñasco y Villa Arista.

Una vez construido el índice de eficiencia se realizó un Análisis de *Cluster* con las siete variables que caracterizan a las empresas, con el objetivo de que nos ayudara a generar grupos de empresas homogéneas, esto es, que se pudieran separar a las empresas con altas eficiencias de las de eficiencias medias y de las empresas ineficientes. Para ello se usaron seis métodos: *Average*, *Centroide*, *Single*, *Complete*, *Ward* y *K-Means* y los resultados muestran que los mejores métodos para conglomerar fueron *Ward* y *K-Means*, ya que a pesar de tener claras diferencias entre los grupos, eran los métodos que mejor agruparon. Finalmente se efectuó un nuevo análisis de *cluster* para los mejores métodos, *Ward* y *K-Means*, pero esta vez se utilizaron los primeros tres componentes principales para construir los grupos, los resultados entre estos dos métodos fueron parecidos.

Debido a lo anterior, se hizo un Análisis Discriminante lineal que nos ayudó a definir al mejor método de conglomeración. Los resultados del análisis muestran que el método

de aglomeración de *Ward* tiene una mejor clasificación (0.91% correcto total), que la aglomeración de *K-means* (0.89% correcto total), es decir, que *K-means* tiene más datos mal clasificados. Por lo tanto, se optó por agrupar a las empresas de agua a través del método *Ward*, en donde se clasifica a las empresas como eficientes, regulares e ineficientes.

Finalmente, el modelo de regresión mostró que las variables que inciden en la eficiencia administrativa de una empresa de agua en México se pueden resumir en 8 variables:

- La variable legal, reglamento interno *Reg*
- Las variables administrativas, preparación del director *prep\_dir*, continuidad *cont*, certificado *certificado*, tarifa de agua *tarifa\_dom*, tipo de tarifa *tip\_tarif*
- Las variables exógenas, zona de disponibilidad *zd* y precipitación media fluvial *prec\_mfluv*

Dados los resultados encontrados podemos concluir que la eficiencia de las empresas que ofrecen el servicio de agua en México no sólo depende de la naturaleza de la organización económico-administrativa de la empresa, también depende de factores legales tales como un reglamento interno de la empresa; de factores administrativos tales como son el costo de la tarifa del agua y la estructura de la tarifa (precio fijo o precio por volumen) y la inversión de la empresa en certificar su proceso de manera externa, de la continuidad directiva de la empresa y de la preparación académica del director; depende también de las variables de entorno ambiental, esto es de la zona en donde se ubique la empresa y de la precipitación media.

El uso de la interacción de las variables *tipo de tarifa* y *costo de tarifa* nos ayuda a explicar de qué forma el precio del agua y la manera de cobrar el líquido, definen la eficiencia en una empresa de agua. Así mismo, la interacción de las variables *zona de disponibilidad de agua* y *precipitación fluvial*, miden el comportamiento de la empresa ante la escasez o abundancia del líquido y cómo afecta esto la eficiencia de una empresa.

**Tabla 18**  
**Comparación del número de elementos por grupo**  
**utilizando diversos métodos de Cluster**

Método	3 Grupos	5 Grupos
Average	G1: 1, G2:5 y G3:40	G1: 1, G2:1, G3:4, G4:28 y G5:12
Centride	G1: 1, G2:1 y G3:44	G1: 1, G2:1, G3:9, G4:1 y G5:34
Single	G1: 1, G2:1 y G3:44	G1: 1, G2:1, G3:1, G4:1 y G5:42
Complete	G1: 9, G2:4 y G3:33	G1: 9, G2:4, G3:10, G4:6 y G5:17
Ward	G1: 9, G2:7 y G3:30	G1: 9, G2:7, G3:8, G4:4 y G5:18
k-mean	G1: 27, G2:11 y G3:8	

Nota: G1: Grupo 1, G2: Grupo 2, G3: Grupo 3, G4: Grupo 4 y G5: Grupo 5.  
Fuente: Elaboración propia, 2015.

# BIBLIOGRAFÍA

---

Andrei, Jouravlev (2004) /Los servicios de agua potable y saneamiento en el umbral del siglo XXI. CEPAL- Naciones Unidas, Santiago de Chile.

Cámara de Diputados

<http://www.diputados.gob.mx/LeyesBiblio/index.htm>

Comisión Federal para la Protección contra los Riesgos Sanitarios COFEPRIS

[http://www.cofepris.gob.mx/wb/cfp/agua\\_de\\_calidad\\_bacteriologica/rid/1947?page=3](http://www.cofepris.gob.mx/wb/cfp/agua_de_calidad_bacteriologica/rid/1947?page=3)

[http://10.0.253.229/wb/cfp/agua\\_de\\_calidad\\_bacteriologica](http://10.0.253.229/wb/cfp/agua_de_calidad_bacteriologica)

Comisión Estatal del Agua de Guanajuato (2009)/ *Diagnóstico Sectorial de Agua Potable y Saneamiento 2008*, Ed. CEAG, México, Guanajuato.

Comisión Nacional del Agua (2011)/*Programa de seguimiento de indicadores de gestión para el cumplimiento de meta de eficiencia global. Manual de indicadores de gestión*, Ed. CNA, México.

<http://www.conagua.gob.mx/CONAGUA07/Publicaciones/Publicaciones/SGAPDS-11-11.pdf>

Comisión Nacional del Agua (2006)/*Situación del subsector Agua Potable, Alcantarillado y Saneamiento*, Ed. CNA, México.

<http://www.conagua.gob.mx/CONAGUA07/Publicaciones/Publicaciones/Subsector2006.pdf>

Comisión Nacional del Agua (2007)/*Situación del subsector Agua Potable, Alcantarillado y Saneamiento*, Ed. CNA, México.

<http://www.conagua.gob.mx/CONAGUA07/Publicaciones/Publicaciones/Subsector2007.pdf>

Comisión Nacional del Agua (2008)/*Situación del subsector Agua Potable, Alcantarillado y Saneamiento*, Ed. CNA, México.

<http://www.conagua.gob.mx/CONAGUA07/Publicaciones/Publicaciones/Subsector2008.pdf>

Comisión Nacional del Agua (2009) / *Situación del Subsector de Agua Potable, Alcantarillado y Saneamiento*, en

<http://www.conagua.gob.mx/CONAGUA07/Publicaciones/Publicaciones/LibroAnexosYTablas-Situaci%C3%B3nSAPAS.pdf>

Consejo Consultivo del Agua, (2010) / *La gestión del agua en las ciudades de México: Indicadores de desempeño de Organismos Operadores*

<http://www.imta.gob.mx/compaps/images/stories/pdf/indicadorescca2010.pdf>

Institutos Estatales Electorales

<http://www.marketingpolitico.com.mx/Institutoseselectorales.htm>

Instituto Nacional de Ecología, Dirección de Economía Ambiental (2010)/ Estudio de la Eficiencia de los Organismos Operadores de Agua en México.

Instituto Nacional de Estadística y Geografía (2004) /*Panorama censal de los organismos operadores de agua en México. Censos económicos 2004*. INEGI

Instituto Mexicano de Tecnología del Agua IMTA, Programa de Indicadores de Gestión de Organismos Operadores

<http://www.pigoo.gob.mx/>

Morrison (1990)/ *Multivariate Statistical Methods*. Mc Graw Hill.

R version 3.0.0 (2013-04-03) -- "Masked Marvel"

Copyright (C) 2013 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

<http://www.r-project.org/>

Saleth y Dinar (2004) / *The Institutional Economics Of Water A Cross-Country Analysis Of Institutions And Performance*. The World Bank.

Seber, G.A.F. (1984)/ *Multivariate Observation*. Wiley, New York

SEMARNAT (2011), Sistema Nacional de Información Ambiental y de Recursos Naturales, en

[http://www.semarnat.gob.mx/informacionambiental/Documents/sniarn/index\\_estadistica.html](http://www.semarnat.gob.mx/informacionambiental/Documents/sniarn/index_estadistica.html)

Secretaría de Salud. Vigilancia y evaluación del control de calidad del agua para uso y consumo humano, distribuida por sistemas de abastecimiento público, NORMA Oficial Mexicana NOM-179-SSA1-1998.

México. Reglamento de la Ley de Aguas Nacionales, Diario Oficial de la Federación, 29 de agosto de 2002.

Peña, Daniel (2002) / *Análisis de datos multivariantes*. Mc Graw Hill, España.

Transparencia Internacional (2011)/ Índice Nacional de Corrupción y Buen Gobierno, en <http://www.transparenciamexicana.org.mx/historia/>

Ward, J. H., Jr. (1963)/ "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, 58, 236–244

# ANEXOS

## Figura 1.A. Cuestionario aplicado a las empresas operadoras de agua

### Sistema Municipal de Agua Potable y Alcantarillado De Tuxtla Gutiérrez, Chiapas

#### Preguntas:

1. ¿A cuánto asciende la cobertura de agua potable y alcantarillado? (desde 2007)
2. ¿Cuál es el número de usuarios registrados en su padrón? (desde 2007)
3. ¿Número de Usuarios que pagan a tiempo? (desde 2007)
4. ¿Cuál es el volumen de agua en m<sup>3</sup> producido anualmente? (desde 2007)
5. ¿Cuál es el volumen de agua en m<sup>3</sup> facturada anualmente? (desde 2007)
6. ¿Cuales son los ingresos en (pesos) por venta de agua? (desde 2007)
7. ¿Cuales son los costos por operación, Mantenimiento y Administración? (desde 2007)
8. ¿Cuál es el tiempo promedio de la rotación del director (años)?
9. ¿Cuál es el tipo de administración que llevan acabo?
  - » Descentralizado Municipal
  - » Descentralizado Intermunicipal
  - » Directo Municipal
  - » Descentralizado Estatal
10. La administración del servicio ah ganado algún tipo de reconocimiento, concurso por buen funcionamiento?
11. ¿Actualmente se encuentra en algún programa de eficiencia?
12. ¿La actual administración ah ganado algún tipo de premio por eficiencia como: CIDE, ANEAS, FIDE u otro?
13. ¿Actualmente ah concesionado algún tipo de operación en el servicio que ofrece?
14. ¿A cuánto asciende los egresos recuperados de los costos de operación y servicio? (desde 2007)
15. ¿A cuánto asciende el subsidio al año que recibe su administración para funcionar? (desde 2007)
16. ¿Tienen algún tipo de certificación?
17. ¿Qué tipo de tarifas aplican? a) fijas o b) servicio medido
18. ¿A cuánto asciende la tarifa doméstica? (desde 2007)
19. ¿Cuantas quejas reciben por parte de los usuarios al año? (desde 2007)
20. ¿Los acuerdos de la administración son accesibles (publicados) para el público en general? ¿en donde se publican los acuerdos?
21. ¿En los últimos años han adquirido equipo eficiente para sistemas de bombeo?
22. ¿Existen suspensiones del servicio por falta de pago? ¿Cuántas en un año? (desde 2007)
23. ¿Hay recargos aplicables a los usuarios por falta de pago? ¿Cuántos en un año? (desde 2007)
24. ¿Cuál el numero de micro medidores que administra su organismo operador? (desde 2007)
25. ¿Cuántas macro mediciones surten a su organismo operador? (desde 2007)
26. ¿Cuál es la eficiencia física, Comercial y Global del organismo operador que maneja? (desde 2007)
27. ¿A cuánto asciende los egresos totales por el servicio total que ofrece? (desde 2007)
28. ¿Cuál es el monto de la inversión total (maquinaria, equipo, capacitación, actualización, etc.) que realiza anualmente? (desde 2007).

#### Respuestas:

1. En 2007 = 107,081 descargas En 2008 = 111,116 Descargas En 2009 = 113,157 descargas,  
En 2007= 109,892 tomas en 2008 =113,388 en 2009=114,811
2. » En 2007 – 111,614 usuarios » en 2008 – 115,670 » en 2009 – 117,881 » mayo-2010 – 119,412
3. » En 2007 – 55,230 » en 2008 – 55022 » en 2009 – 56,024 » mayo-2010 – 56,797
4. » En 2007 – 44,948,418 » en 2008 – 68,291,981 » en 2009 – 63,896,753
5. » En 2007 – 19,467,905 » en 2008 – 20,781,647 » en 2009 – 18,953,767
6. » En 2007 – \$ 170,488,346 » en 2008 – \$ 207,287,025 » en 2009 – \$ 234,766,173
7. En 2007 = \$ 218,105,246.92, en 2008 = \$ 437,393,310.22, en 2009 = \$ 328,741,478.49
8. Cada 3 años
9. Descentralizado Municipal



### Base de datos que generó al índice de eficiencia

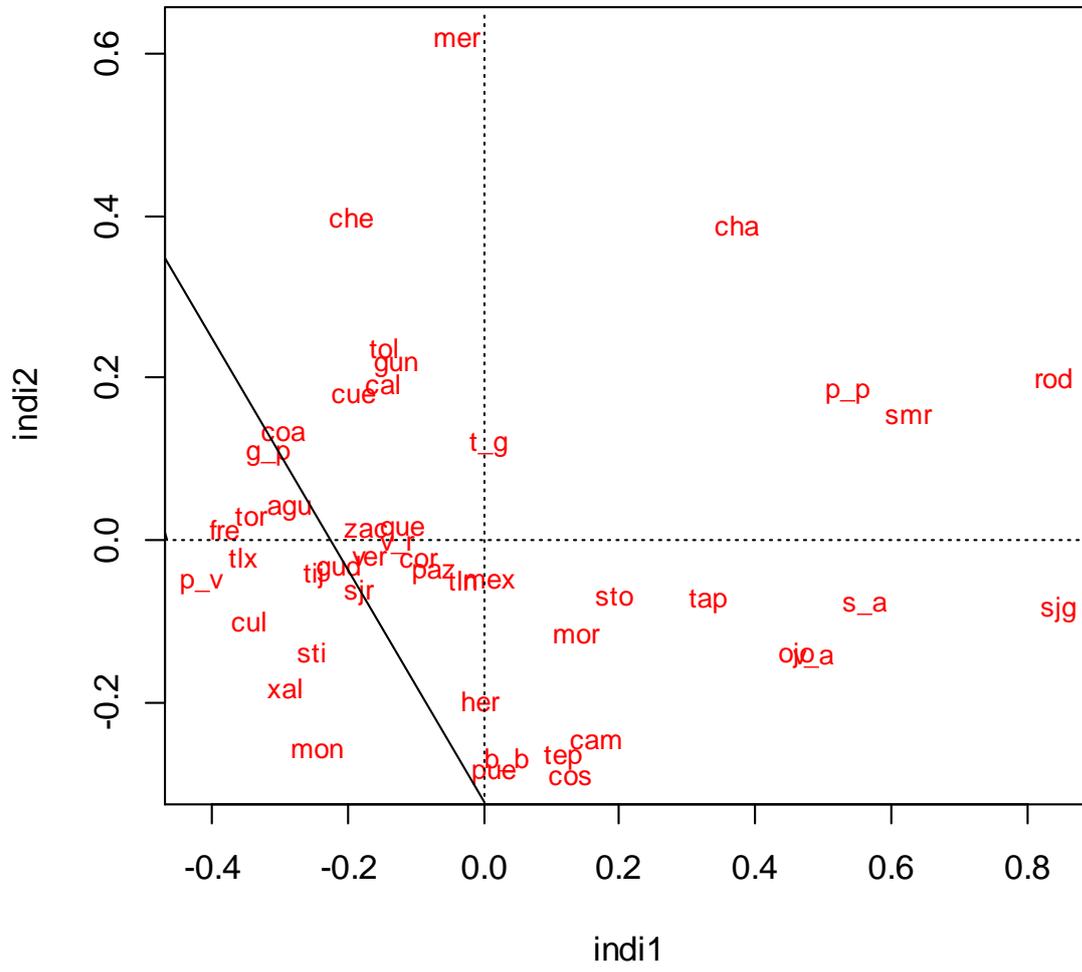
Ubicación de la empresa	Usuarios	Cobertura		Volumen anual (m3)	ingresos (\$)	Costos (\$)	índice de corrupción	Eficiencia de cloración
	pagan a tiempo	agua potable	alcantarillado	facturado	por venta de agua	operación y mantenimiento		
Mexicali	52870	0.95	0.75	81093089	698445905	889347326	8.8	91.98
Tijuana	239556	0.93	0.81	89844000	1322210398	1425854674	8.8	87.71
Hermosillo	161972	0.97	0.94	52477201	199996623	365600000	6	99.31
Puerto Peñasco	2576	0.9	0.7	3418645	17608622	45953644	6	80.43
Cuernavaca	65810	0.91	0.6	25984911	97936015	117274543	9.8	100
Puebla	264123	0.98	0.98	84070851	294576608	569046234	11	96.9
Tlaxcala	8790	0.95	0.89	4664596	17659210	10078426	11.7	98.88
Saltillo	136107	1	0.93	31295000	163034857	195135820	8.4	98.05
Monterrey	840369	0.99	0.98	242028085	1727480818	2405292000	6	99.85
Gómez Palacio	60131	0.94	0.93	18325000	156618000	156618000	6.5	85.12
Guadalajara	783612	0.93	0.89	162523256	968803241	1198943932	8.8	85.24
Puerto Vallarta	46563	0.98	0.86	22017398	155187000	127325000	8.8	98.41
Fresnillo	26589	0.96	0.73	8654450	34956566	35698452	5.9	92.27
Zacatecas	60096	0.99	0.86	14040756	74972988	90030824	5.9	99.89
Torreón	136824	0.99	0.98	32395387	411439715	357694862	8.4	100
Aguascalientes	142325.8	1	0.98	40835161	388021693	341193334	4.7	98.88
Calvillo	3000	0.95	0.97	1968185	15004825	11604929	4.7	94.76
Guanajuato	14251	0.94	0.85	5378333	56878232	43720260	5.1	60.4
Toluca	66915	0.74	0.92	23400000	223106942	181705516	18.8	90.55
Morelia	178707	0.96	0.94	33830519	198965944	516230000	5.7	96.61
Bahía de Banderas	16699	0.89	0.84	7000000	16068322	36009985	5.2	85.04
Tepic	52483	0.98	0.98	27496000	67099690	137328032	5.2	95.71
Córdoba	18413	0.9	0.89	10661818	35664000	42475000	9.7	100
Veracruz	99437	0.96	0.75	62114240	225727486	271552413	9.7	99.79
Tapachula	30524	0.88	0.82	10260023	15390015	47041300	7.1	90.7
Tuxtla Gutiérrez	55230	0.9	0.86	19467905	170488346	218105247	7.1	93.94
Culiacán	174642	0.99	0.98	43749058	384167930	400078236	8.1	98.34
Mérida	189356	0.87	0.05	32156086	171186465	226841258	8.9	99.95
Xalisco	5200	0.98	0.95	5598721	11197335	11651386	5.2	94.07
Chetumal	87332	0.88	0.56	23453504	229588793	206000000	8.9	99.78
Tlanepantla	103475.6	0.99	1	34527561	364488829	533856396	18.8	91.7
La Paz	43118	0.86	0.74	16478852	136118403	194000661	7.3	99.6
Chamotón	2201	0.88	0.67	0	2513217	4515955	7.2	99.53
Ojocaliente	1921.1	0.97	0.9	1198560	897564	3294725	5.9	89.84
Querétaro		0.97	0.98	36391052	437471274	507515628	7	99.97
San Juan del Río	50286	0.95	0.81	9814633	76122382	106788372	7	100
Santa María del Río	803	0.95	0.87	71539	207637	773748	6.8	94.27
Villa de Arista	312.8	0.89	0.79	201640	172429	649931	6.8	94.95
Villa de Reyes	997.85	0.8	0.75	976980	948423	1099081	6.8	99.75
Coatzacoalcos	55064	0.95	0.45	17825402	83526211	100126814	9.7	93.01
Campeche	36809	0.97	0.93	17368172	18278938	43600146	7.2	99.7
Cosío		0.97	0.8	361496	490830	1236727	4.7	98.98
San José de Gracia		0.98	0.93	467720	100648	1604552	4.7	99.5
Sain Alto	692	0.95	0.94	229860	166080	995609	5.9	77.19
Rodeo	128	0.09	0.8	7200	4088860	4005190	6.5	83.38
El Salto	20433	0.88	0.9	8106687	10323010	17327185	8.8	70.44

**Tabla 1.A Resultaos del Primer Score**

Orden	Score 1er componente	Ubicación de la empresa	Orden	Score 1er componente	Ubicación de la empresa
1	-0.419	p_v	24	-0.097	cor
2	-0.382	fre	25	-0.075	paz
3	-0.352	tlx	26	-0.039	mer
4	-0.348	cul	27	-0.031	tln
5	-0.344	tor	28	-0.006	her
6	-0.319	g_p	29	0.007	mex
7	-0.297	coa	30	0.008	t_g
8	-0.293	xal	31	0.014	pue
9	-0.287	agu	32	0.033	b_b
10	-0.253	sti	33	0.117	tep
11	-0.249	tij	34	0.128	cos
12	-0.247	mon	35	0.136	mor
13	-0.215	gud	36	0.165	cam
14	-0.197	che	37	0.193	sto
15	-0.192	cue	38	0.331	tap
16	-0.184	sjr	39	0.370	cha
17	-0.176	zac	40	0.460	ojo
18	-0.170	ver	41	0.483	v_a
19	-0.151	cal	42	0.534	p_p
20	-0.145	tol	43	0.561	s_a
21	-0.131	gun	44	0.626	smr
22	-0.128	v_r	45	0.838	rod
23	-0.122	que	46	0.846	sjg

Fuente: Elaboración propia, 2014

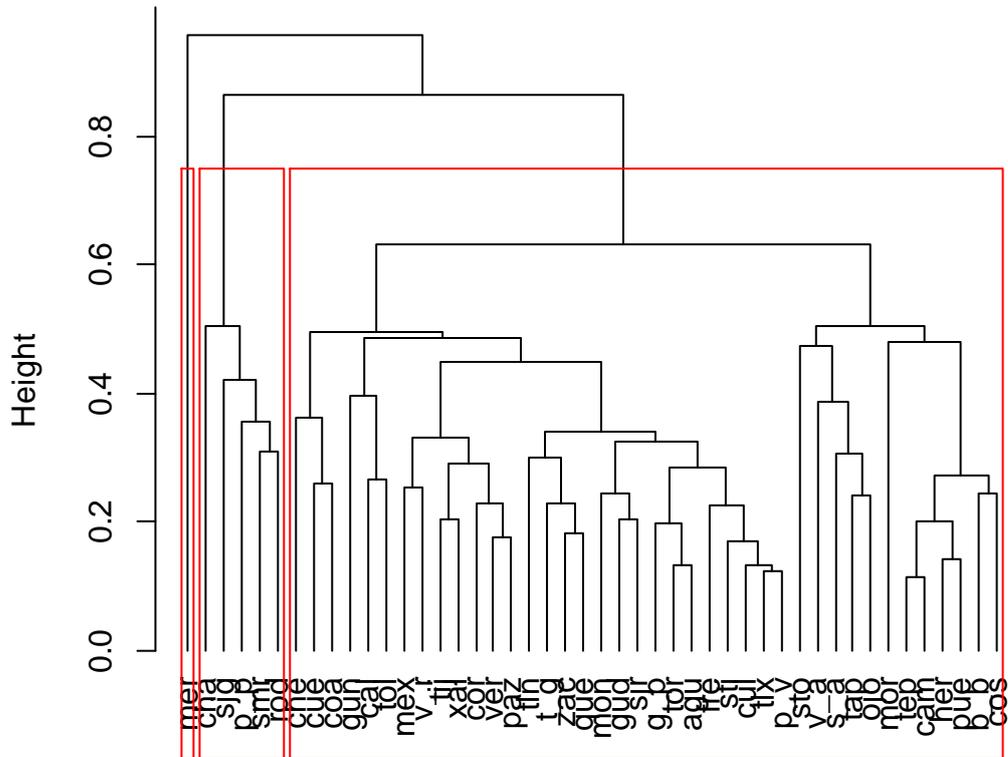
**Figura 2.A**  
**Primero y Segundo Componente Principal de las 46 empresas**



Fuente: Elaboración Propia, 2014.

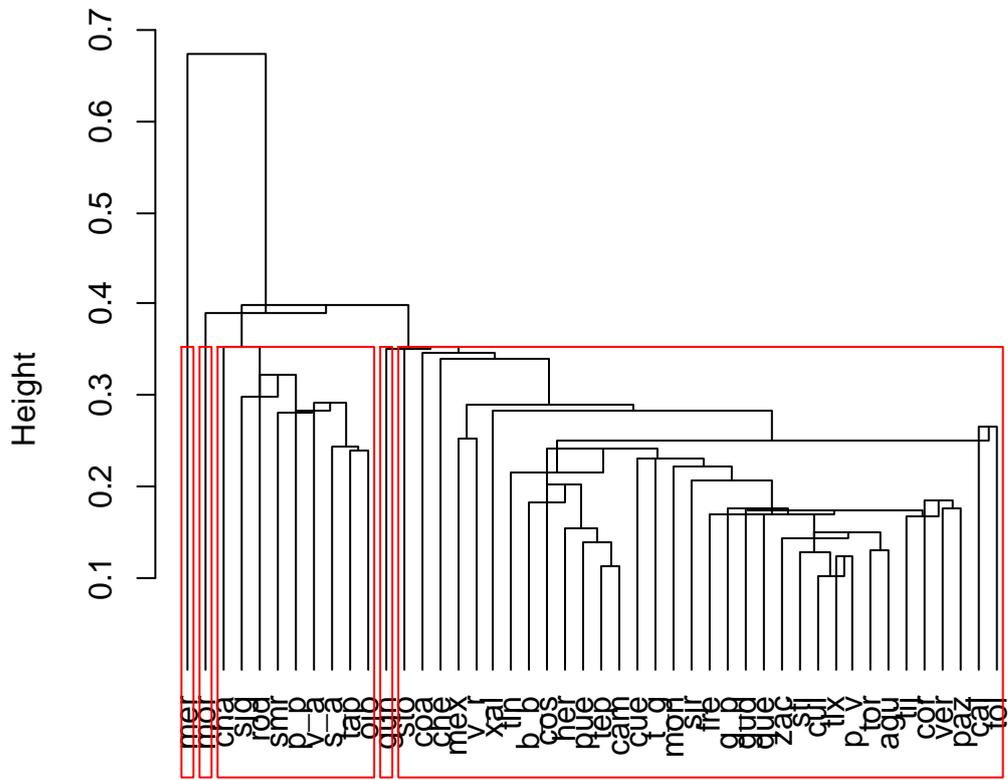


**Figura 4.A**  
**Con 3 grupos**  
**Cluster Average del índice de eficiencia en 46 empresas**



Fuente: Elaboración Propia, 2014.

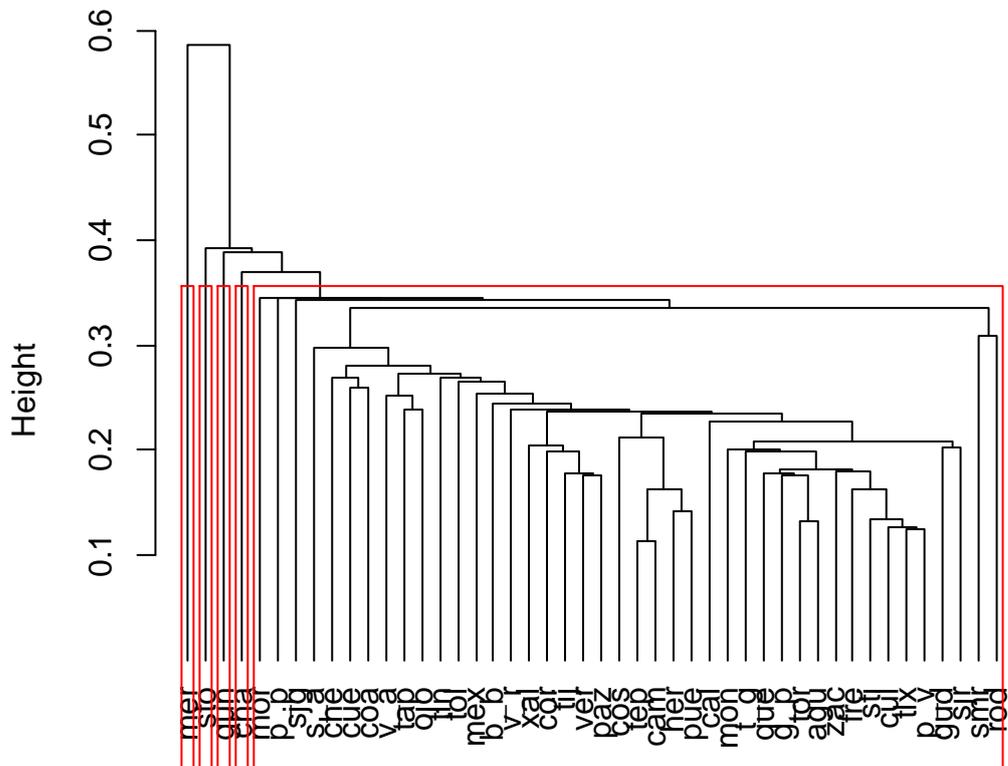
**Figura 5.A**  
**Con 5 grupos**  
**Cluster Centroide del índice de eficiencia en 46 empresas**



Fuente: Elaboración Propia, 2014.

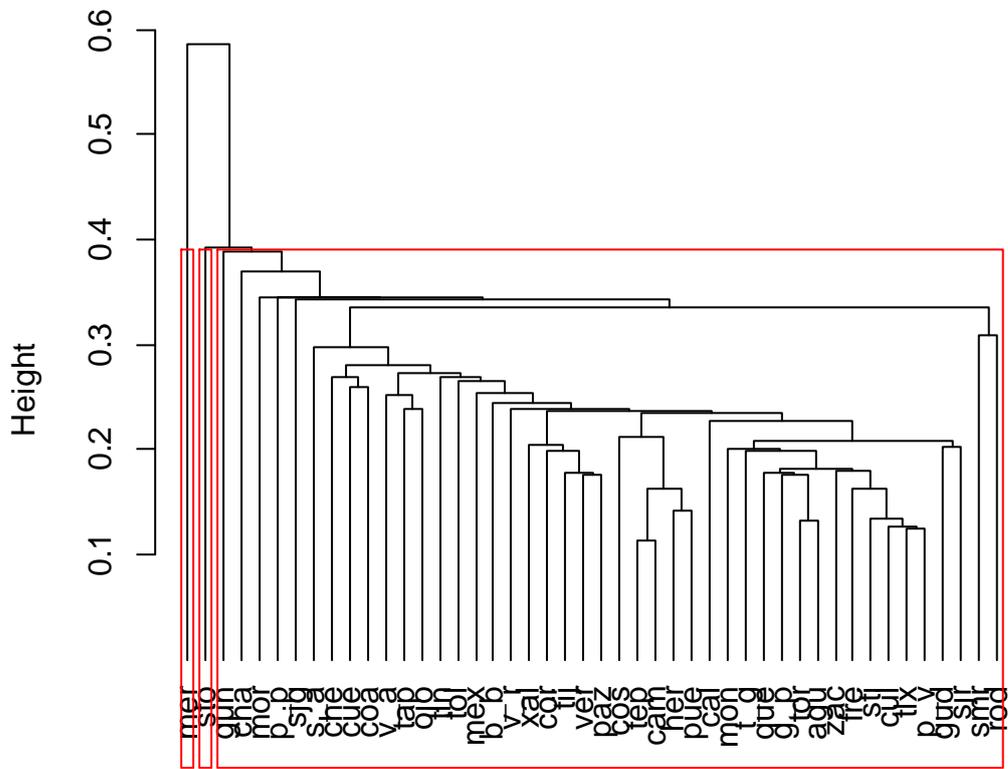


**Figura 7.A**  
**Con 5 grupos**  
**Cluster Single del índice de eficiencia en 46 empresas**



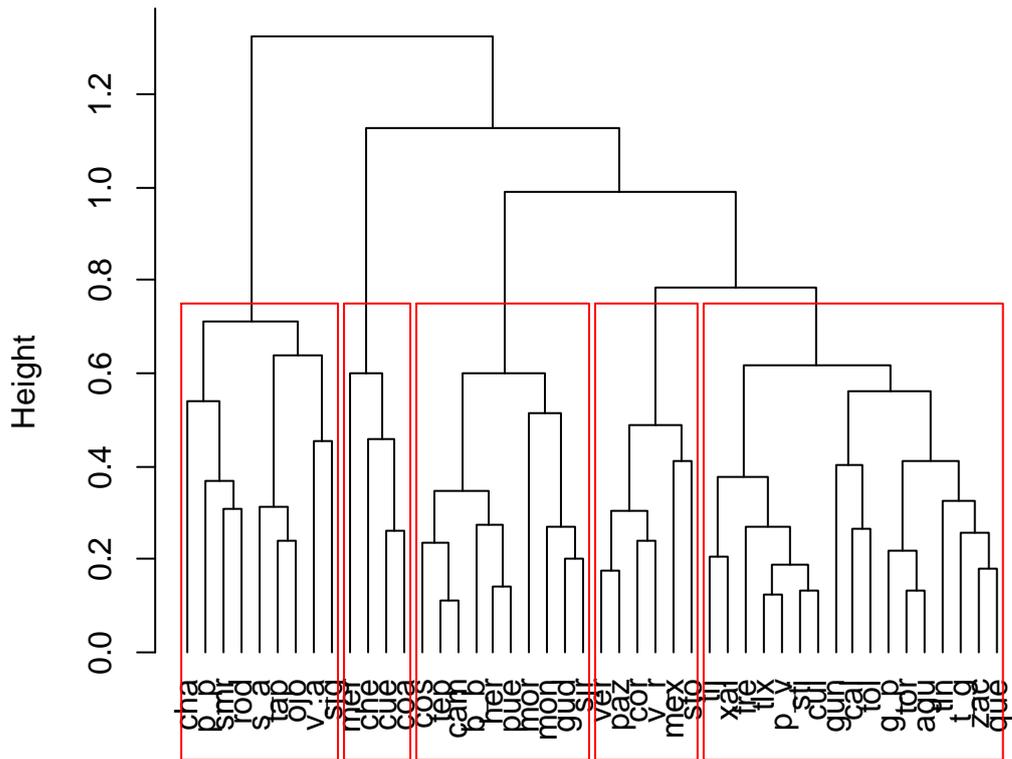
Fuente: Elaboración Propia, 2014.

**Figura 8.A**  
**Con 3 grupos**  
**Cluster Single del índice de eficiencia en 46 empresas**



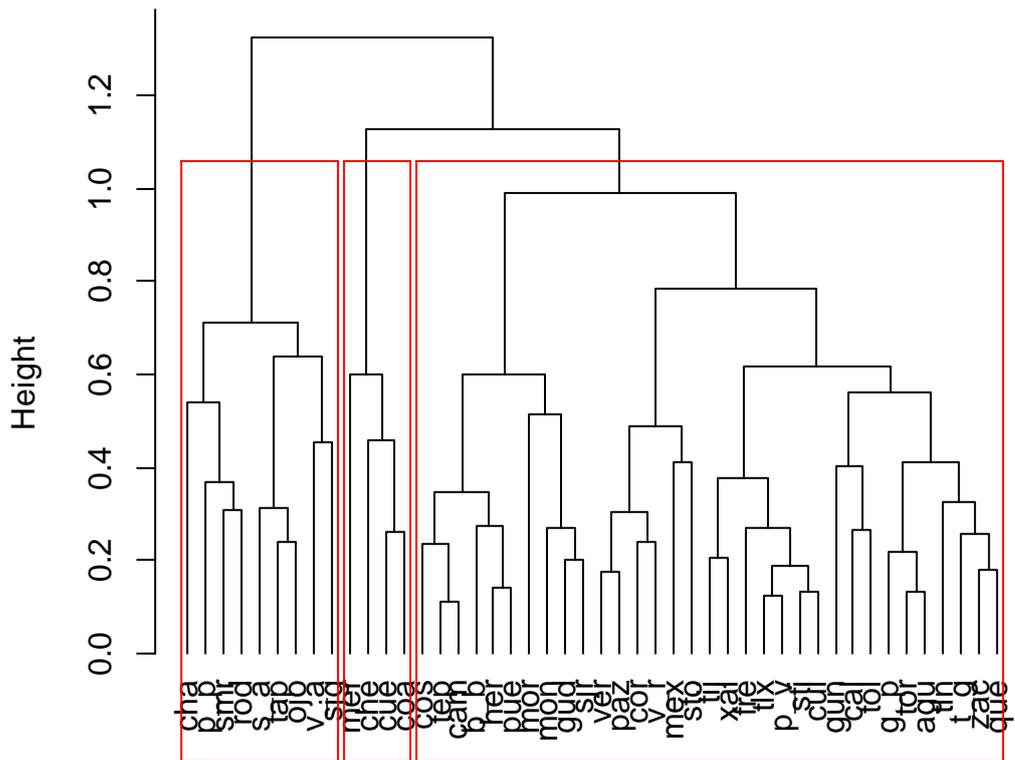
Fuente: Elaboración Propia, 2014.

**Figura 9.A**  
**Con 5 grupos**  
**Cluster Complete del índice de eficiencia en 46 empresas**



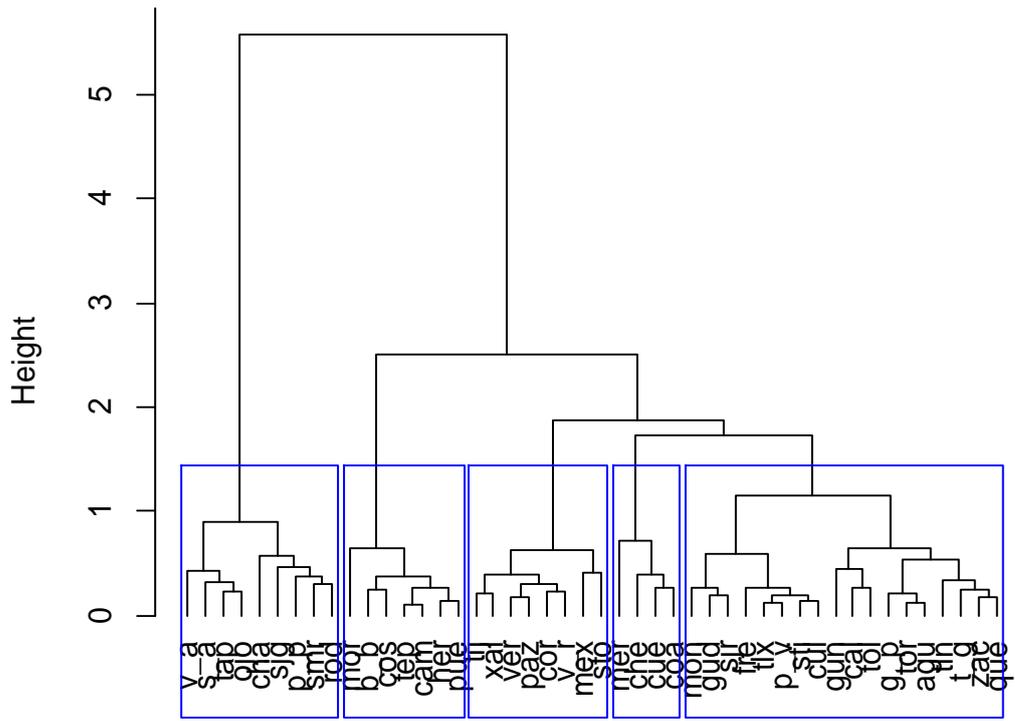
Fuente: Elaboración Propia, 2014.

**Figura 10.A**  
**Con 3 grupos**  
**Cluster Complete del índice de eficiencia en 46 empresas**



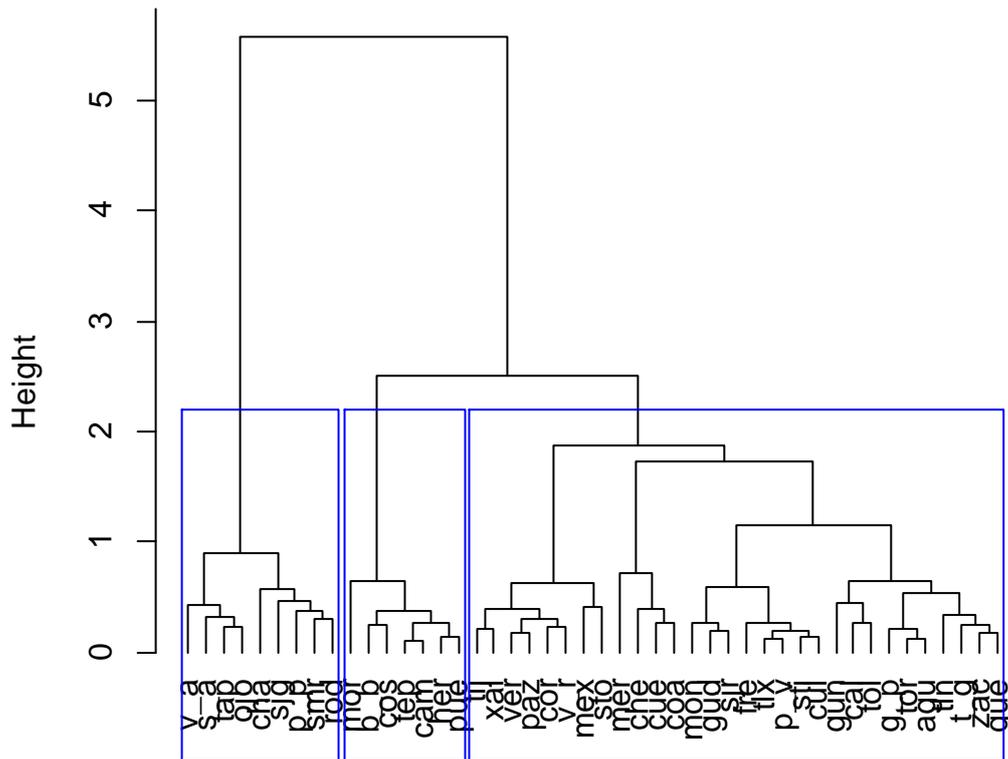
Fuente: Elaboración Propia, 2014.

**Figura 11.A**  
**Con 5 grupos**  
**Cluster Ward del índice de eficiencia CP1-7 Variables en 46 empresas**



Fuente: Elaboración Propia, 2014.

**Figura 12.A**  
**Dendograma *Ward* con 3 grupos**  
**Cluster Ward del índice de eficiencia CP1-7 variables en 46 empresas**  
**Cluster-Ward del Indicador de Eficiencia CP1 7 variables**



Fuente: Elaboración Propia, 2014.

```
(dist(ind_cor[, 2:8]))
hclust (*, "ward")
```

**Figura 13.A**  
**Mapa de clasificación de escasez de agua**



Estado	Municipio	Eficiencia física	Estado	Municipio	Eficiencia física
Aguascalientes	Aguascalientes	63	Nayarit	Bahía de Banderas	90
	Calvillo	32		Tepic	69.9
	Cosío	3		Xalisco	64
	S.J. Gracia	31	Nuevo León	Monterrey	79
Baja California	Mexicali	83	Puebla	Puebla	63
	Tijuana	83	Querétaro de Arteaga	Querétaro	52
Baja California Sur	La Paz	51		San Juan del Río	80
Campeche	Champotón	33	Quintana Roo	Chetumal	75
	Campeche	68	San Luis Potosí	Santa María del Río	8.07
Chiapas	Tapachula	50		Villa de Arista	32
	Tuxtla Gutiérrez	51		Villa de Reyes	40
Coahuila de Zaragoza	Torreón	52	Sinaloa	Culiacán	67
	Saltillo	62.35	Sonora	Hermosillo	56
Durango	Gómez Palacio	47		Puerto Peñasco	57.45
	Rodeo	90	Tlaxcala	Tlaxcala	70
Guanajuato	Guanajuato	59.93		Córdoba	34
Jalisco	Guadalajara	63	Veracruz de Ignacio de Llave	Veracruz	74.93
	Puerto Vallarta	93		Coatzacoalcos	65
	El Salto	40	Yucatán	Mérida	31
México	Toluca	37	Zacatecas	Fresnillo	50
	Tlalnepantla	36		Zacatecas	52
Michoacán	Morelia	39		Ojocaliente	39
Morelos	Cuernavaca	57		Sain Alto	30

Fuente: Elaboración propia con base a información de CONAGUA

**Tabla 2.A.**  
**Comparación del Cluster para cinco grupos**

	CP1	Average	Centroide	Single	Ward	K-mean	Complete
EFICIE NTES	-0.419	p_v	p_v	p_v	p_v	p_v	p_v
	-0.382	fre	fre	fre	fre	fre	fre
	-0.352	tlx	tlx	tlx	tlx	tlx	tlx
	-0.348	cul	cul	cul	cul	cul	cul
	-0.344	tor	tor	tor	tor	tor	tor
	-0.319	g_p	g_p	g_p	g_p	g_p	g_p
	-0.297	coa	coa	coa	coa	coa	coa
	-0.293	xal	xal	xal	xal	xal	xal
	-0.287	agu	agu	agu	agu	agu	agu
	-0.253	sti	sti	sti	sti	sti	sti
	-0.249	tij	tij	tij	tij	tij	tij
	-0.247	mon	mon	mon	mon	mon	mon
	-0.215	gud	gud	gud	gud	gud	gud
	-0.197	che	che	che	che	che	che
	-0.192	cue	cue	cue	cue	cue	cue
	-0.184	sjr	sjr	sjr	sjr	sjr	sjr
	-0.176	zac	zac	zac	zac	zac	zac
	-0.170	ver	ver	ver	ver	ver	ver
	-0.151	cal	cal	cal	cal	cal	cal
	-0.145	tol	tol	tol	tol	tol	tol
	-0.131	gun	gun	gun	gun	gun	gun
	-0.128	v_r	v_r	v_r	v_r	v_r	v_r
	-0.122	que	que	que	que	que	que
	-0.097	cor	cor	cor	cor	cor	cor
	-0.075	paz	paz	paz	paz	paz	paz
	-0.039	mer	mer	mer	mer	mer	mer
	-0.031	tln	tln	tln	tln	tln	tln
	-0.006	her	her	her	her	her	her
	0.007	mex	mex	mex	mex	mex	mex
	0.008	t_g	t_g	t_g	t_g	t_g	t_g
	0.014	pue	pue	pue	pue	pue	pue
	0.033	b_b	b_b	b_b	b_b	b_b	b_b
	0.117	tep	tep	tep	tep	tep	tep
	0.128	cos	cos	cos	cos	cos	cos
	0.136	mor	mor	mor	mor	mor	mor
	0.165	cam	cam	cam	cam	cam	cam
	0.193	sto	sto	sto	sto	sto	sto
	0.331	tap	tap	tap	tap	tap	tap
	0.370	cha	cha	cha	cha	cha	cha
	0.460	ojo	ojo	ojo	ojo	ojo	ojo
0.483	v_a	v_a	v_a	v_a	v_a	v_a	
0.534	p_p	p_p	p_p	p_p	p_p	p_p	
0.561	s_a	s_a	s_a	s_a	s_a	s_a	
0.626	smr	smr	smr	smr	smr	smr	
0.838	rod	rod	rod	rod	rod	rod	
0.846	sjg	sjg	sjg	sjg	sjg	sjg	

Tabla 3.A.  
Comparación del *Cluster* para tres grupos

	CP1	Average	Centroide	Single	Ward	K-mean	Complete
EFICIENTES	-0.419	p_v	p_v	p_v	p_v	p_v	p_v
	-0.382	fre	fre	fre	fre	fre	fre
	-0.352	tlx	tlx	tlx	tlx	tlx	tlx
	-0.348	cul	cul	cul	cul	cul	cul
	-0.344	tor	tor	tor	tor	tor	tor
	-0.319	g_p	g_p	g_p	g_p	g_p	g_p
	-0.297	coa	coa	coa	coa	coa	coa
	-0.293	xal	xal	xal	xal	xal	xal
	-0.287	agu	agu	agu	agu	agu	agu
	-0.253	sti	sti	sti	sti	sti	sti
	-0.249	tij	tij	tij	tij	tij	tij
	-0.247	mon	mon	mon	mon	mon	mon
	-0.215	gud	gud	gud	gud	gud	gud
	-0.197	che	che	che	che	che	che
	-0.192	cue	cue	cue	cue	cue	cue
	-0.184	sjr	sjr	sjr	sjr	sjr	sjr
	-0.176	zac	zac	zac	zac	zac	zac
	-0.170	ver	ver	ver	ver	ver	ver
	-0.151	cal	cal	cal	cal	cal	cal
	-0.145	tol	tol	tol	tol	tol	tol
-0.131	gun	gun	gun	gun	gun	gun	
-0.128	v_r	v_r	v_r	v_r	v_r	v_r	
-0.122	que	que	que	que	que	que	
-0.097	cor	cor	cor	cor	cor	cor	
-0.075	paz	paz	paz	paz	paz	paz	
EFICIENCIA MEDIA	-0.039	mer	mer	mer	mer	mer	mer
	-0.031	tln	tln	tln	tln	tln	tln
	-0.006	her	her	her	her	her	her
	0.007	mex	mex	mex	mex	mex	mex
	0.008	t_g	t_g	t_g	t_g	t_g	t_g
	0.014	pue	pue	pue	pue	pue	pue
	0.033	b_b	b_b	b_b	b_b	b_b	b_b
	0.117	tep	tep	tep	tep	tep	tep
	0.128	cos	cos	cos	cos	cos	cos
	0.136	mor	mor	mor	mor	mor	mor
INEFICIENTES	0.165	cam	cam	cam	cam	cam	cam
	0.193	sto	sto	sto	sto	sto	sto
	0.331	tap	tap	tap	tap	tap	tap
	0.370	cha	cha	cha	cha	cha	cha
	0.460	ojo	ojo	ojo	ojo	ojo	ojo
	0.483	v_a	v_a	v_a	v_a	v_a	v_a
	0.534	p_p	p_p	p_p	p_p	p_p	p_p
	0.561	s_a	s_a	s_a	s_a	s_a	s_a
	0.626	smr	smr	smr	smr	smr	smr
	0.838	rod	rod	rod	rod	rod	rod
0.846	sjg	sjg	sjg	sjg	sjg	sjg	

**Tabla 4.A.**  
**Probabilidad posterior por empresa y grupo clasificado (Ward)**

Empresa	Grupos			Empresa	Grupos		
	1°	2°	3°		1°	2°	3°
1	1.0000	0.0000	0.0000	24	0.9989	0.0011	0.0000
2	0.9999	0.0001	0.0000	25	0.0001	0.9068	0.0931
3	0.0019	0.9977	0.0003	26	0.9938	0.0060	0.0002
4	0.0003	0.0002	0.9996	27	0.9971	0.0029	0.0000
5	0.9921	0.0079	0.0000	28	0.9243	0.0014	0.0743
6	0.0007	0.9991	0.0001	29	1.0000	0.0000	0.0000
7	0.9997	0.0003	0.0000	30	1.0000	0.0000	0.0000
8	0.9149	0.0851	0.0000	31	0.6616	0.3348	0.0036
9	0.0575	0.9425	0.0000	32	0.8439	0.1561	0.0000
10	0.9979	0.0021	0.0000	33	0.1309	0.0041	0.8650
11	0.8559	0.1441	0.0000	34	0.0000	0.0226	0.9774
12	0.9993	0.0007	0.0000	35	0.9965	0.0035	0.0000
13	0.9999	0.0001	0.0000	36	0.1811	0.8188	0.0000
14	0.9846	0.0154	0.0000	37	0.0000	0.0000	1.0000
15	0.9972	0.0028	0.0000	38	0.0003	0.0532	0.9465
16	0.9999	0.0001	0.0000	39	1.0000	0.0000	0.0000
17	1.0000	0.0000	0.0000	40	0.9771	0.0229	0.0000
18	1.0000	0.0000	0.0000	41	0.0006	0.9855	0.0139
19	0.9997	0.0003	0.0000	42	0.0025	0.9919	0.0056
20	0.0000	0.9577	0.0423	43	0.0000	0.0000	1.0000
21	0.0003	0.9996	0.0001	44	0.0000	0.0221	0.9779
22	0.0113	0.9791	0.0097	45	0.0000	0.0000	1.0000
23	0.9993	0.0007	0.0000	46	0.1662	0.5941	0.2396

**Tabla 5.A.**  
**Probabilidad posterior por empresa y grupo clasificado (*K-mean*)**

Empresa	Niveles			Empresa	Niveles		
	1°	2°	3°		1°	2°	3°
<b>1</b>	0.9999	0.0000	0.0001	<b>24</b>	0.9991	0.0000	0.0009
<b>2</b>	0.9999	0.0000	0.0001	<b>25</b>	0.0000	0.3310	0.6690
<b>3</b>	0.0034	0.0003	0.9963	<b>26</b>	0.9958	0.0002	0.0040
<b>4</b>	0.0000	1.0000	0.0000	<b>27</b>	0.9992	0.0000	0.0008
<b>5</b>	0.9980	0.0000	0.0020	<b>28</b>	0.9998	0.0001	0.0001
<b>6</b>	0.0005	0.0003	0.9992	<b>29</b>	1.0000	0.0000	0.0000
<b>7</b>	0.9999	0.0000	0.0001	<b>30</b>	1.0000	0.0000	0.0000
<b>8</b>	0.9631	0.0000	0.0369	<b>31</b>	0.8586	0.0190	0.1223
<b>9</b>	0.2156	0.0000	0.7844	<b>32</b>	0.8597	0.0001	0.1402
<b>10</b>	0.9996	0.0000	0.0004	<b>33</b>	0.8162	0.1722	0.0116
<b>11</b>	0.6027	0.0000	0.3973	<b>34</b>	0.0000	0.9867	0.0133
<b>12</b>	0.9998	0.0000	0.0002	<b>35</b>	0.9987	0.0000	0.0013
<b>13</b>	1.0000	0.0000	0.0000	<b>36</b>	0.5555	0.0000	0.4445
<b>14</b>	0.9961	0.0000	0.0039	<b>37</b>	0.0000	0.9999	0.0001
<b>15</b>	0.9997	0.0000	0.0003	<b>38</b>	0.0000	0.9802	0.0197
<b>16</b>	1.0000	0.0000	0.0000	<b>39</b>	1.0000	0.0000	0.0000
<b>17</b>	1.0000	0.0000	0.0000	<b>40</b>	0.9929	0.0000	0.0071
<b>18</b>	0.9948	0.0052	0.0000	<b>41</b>	0.0003	0.0119	0.9878
<b>19</b>	0.9994	0.0001	0.0005	<b>42</b>	0.0011	0.0066	0.9923
<b>20</b>	0.0000	0.0108	0.9892	<b>43</b>	0.0000	1.0000	0.0000
<b>21</b>	0.0001	0.0038	0.9960	<b>44</b>	0.0000	0.9982	0.0018
<b>22</b>	0.0059	0.0158	0.9783	<b>45</b>	0.0000	1.0000	0.0000
<b>23</b>	0.9993	0.0000	0.0007	<b>46</b>	0.5372	0.1343	0.3285

