

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL-ESPOL**



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS  
DEPARTAMENTO DE POSTGRADO**

**PROYECTO DE TITULACIÓN  
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:  
“MAGÍSTER EN ESTADÍSTICA APLICADA”**

**TEMA:  
ESTUDIO MULTIVARIANTE PARA IDENTIFICAR  
ESTADIO DE CANCER PARA MEJORAR LA  
SUPERVIVENCIA – CALIDAD DE VIDA**

**AUTOR: VICTOR FERNANDO AVILA LINO**

**Guayaquil - Ecuador**

**2024**

## RESUMEN

En investigación los tipos de estadios de cáncer, uno de los problemas es identificar la gravedad del cáncer. En este trabajo con la aplicación de análisis multivariante permite un enfoque integral en la interpretación de datos, ayuda a identificar factores críticos, segmentar grupos, predecir tendencias y resumir la información compleja en patrones interpretables, en busca de insights más precisos y profundos, los métodos de análisis multivariantes como Método Stata, Análisis Factorial Múltiple (AFM), nos permiten explorar relaciones complejas entre múltiples variables y proporcionan una visión más completa de los fenómenos que estamos estudiando.

## **ABSTRACT**

In cancer staging research, one of the problems is to identify the severity of the cancer. In this work with the application of multivariate analysis allows a comprehensive approach to data interpretation, helps to identify critical factors, segment groups, predict trends and summarize complex information into interpretable patterns, in search of more accurate and deeper insights, multivariate analysis methods such as Statis Method, Multiple Factor Analysis (MFA), allow us to explore complex relationships between multiple variables and provide a more complete view of the phenomena we are studying.

## **DEDICATORIA**

Nuestra recompensa se encuentra en el esfuerzo y no en el resultado. Un esfuerzo total es una victoria completa.

Mahatma Gandhi

## **AGRADECIMIENTO**

A Dios por la vida

A mi tutor, el Doctor Omar Ruiz Barzola, por su constante motivación y valiosa guía para la ejecución de este proyecto de titulación

A los profesores y directivos de la ESPOL por ayudarme a incrementar mis conocimientos y crecer como profesional

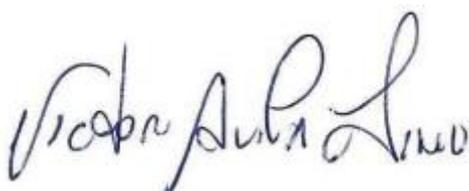
A Ernesto Burbano de Lara gran jefe, un ser humano súper amable y por darme la oportunidad de aprender este masterado

A mi familia, especialmente a mi esposa, Gabriela, por su comprensión y ánimo. Y por supuesto a mis hijos, ya que cualquier reto o esfuerzo que realizo en mi vida lo hago por ellos

## DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido

A handwritten signature in blue ink, reading "Victor Fernando Avila Lino". The signature is written in a cursive, flowing style.

Ing. Victor Fernando Avila Lino

# TRIBUNAL DE GRADUACIÓN



---

Andrea García Angulo, Ph.D.

PRESIDENTE



---

Omar Ruíz Barzola, Ph.D.

TUTOR



---

Mgtr. Francisco Moreira Villegas

DOCENTE EVALUADOR

## ABREVIATURAS O SIGLAS

T1	TUMORES DE MAMA
T2	TUMORES CABEZA Y CUELLO
T3	TUMORES ÓRGANOS DIGESTIVOS
T4	TUMORES ÓRGANOS GENITALES FEMENINOS
T5	TUMORES ÓRGANOS GENITALES MASCULINOS
T6	TUMORES PIEL SARCOMA Y PARTES BLANDAS
T7	TUMORES-MIELOMAS-LEUCEMIAS OTROS
IE0	I.E.S.S. ESTADIO 0
IE1	I.E.S.S. ESTADIO 1
IE2	I.E.S.S. ESTADIO 2
IE3	I.E.S.S. ESTADIO 3
IE4	I.E.S.S. ESTADIO 4
ME0	M.S.P. ESTADIO 0
ME1	M.S.P. ESTADIO 1
ME2	M.S.P. ESTADIO 2
ME3	M.S.P. ESTADIO 3
ME4	M.S.P. ESTADIO 4
PE0	PCTE INST. ESTADIO 0
PE1	PCTE INST. ESTADIO 1
PE2	PCTE INST. ESTADIO 2
PE3	PCTE INST. ESTADIO 3
PE4	PCTE INST. ESTADIO 4

## TABLA DE CONTENIDO

Tabla 1 Resumen de las variables.....	29
Tabla 2 Coeficiente de correlación vectorial.....	32
Tabla 3 Coseno <sup>2</sup> .....	32
Tabla 4 Pesos k-tablas .....	32
Tabla 5 Valores propios para la inter-estructura.....	34
Tabla 6 Valores propios para el compromiso .....	35
Tabla 7 K tablas Análisis factorial múltiple .....	36
Tabla 8 Valores propios del ACP global .....	37
Tabla 9 Valores propios del ACP del periodo 1- ene1 .....	37
Tabla 10 Coordenadas-Correlación -Cos <sup>2</sup> -Contribucion del ACP del periodo 1- ene1 con relación a los diagnósticos .....	38
Tabla 11 Coordenadas-Correlación -Cos <sup>2</sup> -Contribucion del ACP del periodo 1- mes1 con relación a los cliente-estadios .....	39
Tabla 12 Tabla resumen de las variables .....	41

## LISTADO DE FIGURAS

Figura 1 Ejemplo método Statis.....	15
Figura 2 Esquema general del funcionamiento del AFM.....	16
Figura 3 Matriz X conformada por la yuxtaposición de tablas.....	17
Figura 4 Estructura metodológica Statis .....	24
Figura 5 Análisis Factorial Múltiple .....	25
Figura 6 Esquema AFM / STATIS .....	26
Figura 7 Obtención de la Data en Power BI.....	27
Figura 8 Estructura matriz Método Statis en R Studio .....	28
Figura 9 Resumen Boxplot variable IE0.....	29
Figura 10 Boxplot diagnóstico con relación IE0 .....	30
Figura 11 Interestructura Statis.....	31
Figura 12 Coseno2 .....	33
Figura 13 Matriz Compromiso .....	33
Figura 14 Proyección de Componentes.....	34
Figura 15 Intra-estructura .....	35
Figura 16 Scree plot Análisis factorial múltiple.....	42
Figura 17 Visualiza los resultados de las variables.....	43
Figura 18 Visualiza los resultados de los individuos .....	44
Figura 19 Longitud de los vectores.....	45
Figura 20 Clúster agrupamiento jerárquico.....	46
Figura 21 Clúster nube de puntos.....	47

# TABLA DE CONTENIDO

Contenido	
RESUMEN .....	II
ABSTRACT .....	III
DEDICATORIA.....	IV
AGRADECIMIENTO.....	V
DECLARACIÓN EXPRESA.....	VI
TRIBUNAL DE GRADUACIÓN.....	VII
ABREVIATURAS O SIGLAS .....	VIII
TABLA DE CONTENIDO.....	IX
LISTADO DE FIGURAS .....	X
TABLA DE CONTENIDO.....	XI
<b>CAPÍTULO 1</b> .....	1
1. INTRODUCCION.....	1
1.1 Antecedente .....	1
1.2 Descripción del problema .....	4
1.3 Justificación.....	5
1.4 Objetivo General .....	6
1.5 Objetivos Especificos .....	6
1.6 Hipótesis .....	7
1.7 Alcance .....	7
<b>CAPÍTULO 2</b> .....	8
2. MARCO TEÓRICO.....	8
2.1 ESTADIO DE CANCER.....	8
2.2 ANÁLISIS UNIVARIANTE Y BIVARIANTE .....	9
2.3 ANÁLISIS MULTIVARIANTE.....	11
2.4 METODO STATIS.....	12
2.5 ANALISIS FACTORIAL MULTIPLE (AFM) .....	16

2.6 Paquetes de R.....	18
2.6.1 Stats .....	18
2.6.2 MultibiplotGUI .....	18
2.6.3 biplotbootGUI.....	18
2.6.4 GGEBiplot .....	19
2.6.5 ade4 .....	19
2.6.6 MultBiplotR .....	19
2.6.7 Esquisse.....	20
2.6.8 Factoextra .....	20
2.6.9 SISTEMA DE INFORMACIÓN HOSPITALARIO (HIS)....	20
2.6.10 ESTANDARIZACION.....	20
2.6.11 Algoritmo de Ward .....	21
2.6.12 CLUSTER.....	21
<b>CAPÍTULO 3</b> .....	<b>23</b>
3. METODOLOGIA.....	23
3.1 Descripción de Investigación.....	26
3.2 Autorización Información .....	26
3.3 Obtención de la Data.....	27
<b>CAPÍTULO 4</b> .....	<b>29</b>
RESULTADOS .....	29
4.1 Análisis Explicativo .....	29
4.2 Análisis Método Statis .....	31
4.3 Análisis Factorial Múltiple .....	36
<b>CAPÍTULO 5</b> .....	<b>48</b>
5. CONCLUSIONES Y RECOMENDACIONES.....	48
5.1 Conclusiones Base de datos .....	48
5.2 Análisis Método Statis .....	48
5.3 Análisis Factores Múltiples .....	48
5.4 Conclusión final .....	49
5.5 Recomendaciones.....	49

<b>CAPÍTULO 6</b> .....	50
6. REFERENCIAS .....	50
Bibliografía .....	50
<b>CAPÍTULO 7</b> .....	53
7. APENDICE Y ANEXOS .....	53
CODIGO R .....	53

# CAPÍTULO 1

## 1. INTRODUCCION

### 1.1 Antecedente

En salud, el término estadificación se refiere a la extensión del cáncer, como el tamaño del tumor y si el cáncer se diseminó (Aguirre-Loaiza et al., 2017). Saber el estadio del cáncer sirve para que los médicos puedan determinar la gravedad del cáncer y las probabilidades de supervivencia, planificar el mejor tratamiento, así como identificar estudios clínicos que podrían ser opciones de tratamiento.

El cáncer siempre lleva el nombre del estadio en el momento del diagnóstico, aunque luego empeore o se disemine. La información sobre los cambios en el cáncer que ocurren con el tiempo se añade a la estadificación inicial (Núñez et al., 2017). Es decir, aunque cambie el cáncer, el estadio no cambia.

Hay varios sistemas de estadificación (Aranda Aguilar 2004); algunos se usan de forma general para muchos tipos de cáncer, como el sistema de estadificación tumor, ganglio linfático, metástasis (TNM). Otros sistemas son para tipos específicos de cáncer como:

**Estadio 0.** Este estadio describe el cáncer in situ. “In situ” significa “en el lugar”. Los cánceres de estadio 0 aún se ubican en el lugar en el que se iniciaron. No se diseminaron a los tejidos adyacentes. Este estadio del cáncer suele ser curable. Por lo general, se puede extirpar todo el tumor con una cirugía.

**Estadio I.** Por lo general, este estadio es un cáncer que no ha crecido profundamente en los tejidos adyacentes. Además, no se ha diseminado a los ganglios linfáticos ni a otras partes del cuerpo. A menudo se le denomina cáncer en estadio temprano.

**Estadio II y Estadio III.** En general, estos 2 estadios son cánceres que han crecido más profundamente en los tejidos adyacentes. También es posible que se hayan diseminado a los ganglios linfáticos, pero no a otras partes del cuerpo.

**Estadio IV.** Este estadio significa que el cáncer se ha diseminado a otros órganos o partes del cuerpo. También se le puede denominar cáncer avanzado o metastásico.

Mientras más alto es el número, más avanzado es el cáncer. Es frecuente el uso de letras y números después del primer número para describir más detalles del cáncer. Por ejemplo, el estadio 2 del cáncer de próstata podría ser 2A, 2B o 2C.

El estadio de un cáncer determinado en el momento del diagnóstico y de los tratamientos iniciales no cambia. Esto es así para que los médicos puedan comprender el progreso médico de una persona, ayudar a comprender el pronóstico y conocer cómo afecta el tratamiento a muchas personas.

### **Revisión Bibliográfica**

Aguirre-Loaiza (et al., 2017) Calidad de vida según el estadio del cáncer de seno en mujeres: “El cáncer se clasifica según criterios histológicos en estadios (0 o in situ, I, II, III y IV), los cuales se asocian a las tasas de sobrevida”

Aranda Aguilar, E. (2004). El cáncer colorrectal ocupa el tercer lugar en frecuencia en el hombre y el segundo en la mujer. Es la cuarta causa de mortalidad por cáncer en los países occidentales,

N del Castillo (et al., 2020) Predicción en el diagnóstico de tumores de cáncer de mama empleando métodos de clasificación, se usan las matrices de confusión para conocer las medidas de los modelos de pronósticos y la curva ROC, e recomienda el empleo de estas técnicas en los hospitales y laboratorios donde se realice la detección de esta enfermedad

Castillo, (et al., 2023). Utilizaron Tratamiento del cáncer de mama luminal metastásico y valor de la data del mundo real. El avance en la terapia del cáncer de mama receptor hormonal positivo y HER2 negativo ha mejorado la supervivencia libre de progresión y la supervivencia global en algunos casos.

Además, Sánchez, (et al., 2018) aplicaron tres modelos para pronosticar cáncer de mama precoz, Aplicar tres modelos de pronósticos “ online ” (índices pronósticos de Nottingham (NPI),

¡Adjuvantonline! (AO) y PREDICT utilizados en la práctica oncológica para estratificar a pacientes y definir el uso de terapias adyuvantes en pacientes con cáncer de mama (CM) precoz, para evaluar su evaluación y predicción de sobrevida en nuestra población.

Mientras que, en 2022, García, desarrolló un Modelo predictivo del riesgo de padecer cáncer cervicouterino en la población femenina de Puerto Padre. Con la aplicación de la Regresión Logística Múltiple quedó diseñado un modelo predictivo del riesgo de padecer cáncer cervicouterino en la población femenina del municipio Puerto Padre, al estimar la carga y la correlación de sus factores de riesgo.

## 1.2 Descripción del problema

Aunque cada persona es diferente y cada experiencia con el cáncer es única, se pueden hacer algunas generalizaciones sobre los tipos y estadios del cáncer. La estadificación del cáncer es una forma de agrupar los casos y predecir las probabilidades de remisión, lo que significa ausencia de signos de cáncer en el cuerpo, y supervivencia. También puede ayudar a predecir si el cáncer crecerá, se diseminará o reaparecerá después del tratamiento, lo que se conoce como recidiva.

El estadio del cáncer le indica dónde se encuentra y su tamaño, hasta dónde creció en los tejidos adyacentes y si se diseminó a los ganglios linfáticos adyacentes o a otras partes del cuerpo. Antes de iniciar cualquier tratamiento contra el cáncer, los médicos pueden usar exámenes físicos, exploraciones por imágenes y otras pruebas para determinar el estadio del cáncer. Es posible que la estadificación no se complete hasta que hayan finalizado todas estas pruebas.

Cada uno de los exámenes, pruebas, exploraciones u otras pruebas ofrecen información a los médicos – variables aleatorias para este estudio –; sin embargo, toda esta información podría no ser utilizada al máximo por los expertos, ya que podrían existir relaciones no detectables a la vista que pudiera ser claramente identificadas a través de estudios multivariantes que indique con mayor nivel de precisión el estadio de cáncer de forma eficaz, mejorando así el diagnóstico y posible tratamiento que ofrezca al paciente la posibilidad de mejorar su supervivencia así como su calidad de vida.

### **1.3 Justificación**

Esta investigación se dirigirá en estudiar las diversas opciones del análisis de medición con características multivariantes, debido a varios métodos de medición se utilizarán el Método Statis y Análisis Factorial Múltiple (AFM). Así, el presente trabajo permitirá mediante característica multivariante la comparación de los diferentes tipos de cáncer, enviados por las instituciones de salud en un tiempo determinado y poder precisar el método apropiados de acuerdo con el análisis a realizar.

## **1.4 Objetivo General**

Analizar los tipos de estadio de cáncer en pacientes de la institución prestadora de servicios de salud, mediante la aplicación de modelos estadísticos multivariantes paramétricas y/o no paramétricos; obteniendo con ello información relevante para gestión administrativa y financiera

### **Objetivos Específicos**

- Obtener una base de datos estructurada que cumpla con los requerimientos para el análisis, a través del apoyo de una institución prestadora de servicios de salud.
- Modelar el comportamiento estadístico de pacientes según su tipo de estadio de cáncer, con la aplicación de metodologías estadísticas multivariante de última generación
- Analizar los modelos estadísticos multivariante aplicados, identificando los factores significativos asociados a los tipos de estadio de cáncer en pacientes, con el enfoque de la mejora de la supervivencia y la calidad de vida de los pacientes.

## **1.5 Hipótesis**

Los métodos de análisis multivariantes son ampliamente utilizados en diversas disciplinas, entonces se podrá elegir los métodos más precisos para comprender qué tipo de cáncer evoluciona desde una etapa temprana hasta una enfermedad maligna avanzada, lo que puede ayudar en el desarrollo de estrategias de prevención, diagnóstico y tratamiento más efectivas.

## **1.6 Alcance**

La determinación del problema en el contexto de la aplicación de los modelos multivariantes donde nos permita identificar la necesidad de comprender los factores asociados a los diferentes niveles de gravedad de esta enfermedad y cómo pueden influir en su progresión y manejo. Los modelos multivariantes son herramientas estadísticas valiosa para analizar esta relación y proporcionar información relevante para la toma de decisiones clínicas. El objetivo del estudio evaluar la influencia de diferentes variables en la progresión de la enfermedad del cáncer. El estudio se realizará con datos provenientes de una institución de salud, prestadora de servicios de salud oncológica de la ciudad de Guayaquil, recopilados durante el año 2023.

# CAPÍTULO 2

## 2. MARCO TEÓRICO

### 2.1 ESTADIO DE CANCER

En forma global, los distintos tipos de cáncer son una de las principales causas de muerte. En el 2007 se presentó un total de 7,9 millones de muertes por cáncer a nivel mundial, representando aproximadamente el 13% de todas las muertes. Cerca de 72% de las muertes por cáncer se producen en países de ingresos medios y bajos. Se estima que el número de muertes anuales por cáncer seguirá en aumento, proyectándose un incremento de 45% entre los años 2007 y 2030, llegando a 11,5 millones de muertes. Adicionalmente, 11,3 millones de personas fueron diagnosticadas de cáncer el año 2007, estimándose un incremento de casos nuevos anuales a 15,5 millones hacia el año 2030. En los hombres, los tipos de cáncer más frecuentes son el de próstata, seguido del pulmonar, mientras que en las mujeres son el cáncer de mama seguido del de cuello uterino. (Gutiérrez et al., 2017).

Existen varios sistemas de estadificación de cáncer (Aranda Aguilar 2004); algunos se usan de forma general para muchos tipos de cáncer, como el sistema de estadificación tumor, ganglio linfático, metástasis (TNM). Así como también los tipos de Estadios 0, I, II, III, IV.

Además, podemos indicar que la calidad de vida en pacientes con cáncer se encuentra representado por tres dimensiones (componente relacionado con el cuidado, bienestar y componente sociofamiliar-alimentario); esto ha sido reconocido en diferentes estudios en los que se considera que el estado de salud es uno de los más importantes determinantes de la calidad de vida general (Seidl y Zannon, 2004).

## 2.2 ANÁLISIS UNIVARIANTE Y BIVARIANTE

El análisis univariante, también conocido como estadística descriptiva, se utiliza para estudiar el comportamiento de una variable individualmente. Como indican (Walpole, Myers, Myers, & Ye, 2012): Hay ocasiones en las que no se desea llegar al detalle de la estadística inferencial, sino que únicamente se desea obtener la información resumen de un conjunto de datos representados a través de la muestra; para estos casos se utiliza un análisis univariante. Se utilizará el análisis univariante para realizar una evaluación inicial de los datos obtenidos.

Los análisis univariantes más frecuentes son:

- Tendencia Central: Media, moda y mediana
- Dispersión: Rango, varianza, máximo, mínimo, cuartiles e intercuartílico
- Gráficas: Histogramas, gráfico de barras, polígonos de frecuencia y gráficos circulares.

El análisis univariante es la forma más sencilla de análisis de datos, en la que los datos analizados sólo contienen una variable. Al tratarse de una sola variable, no se ocupa de las causas ni de las relaciones. El objetivo principal del análisis univariante es describir los datos y encontrar los patrones que existen en ellos.

El análisis bivariante consiste en estudiar las relaciones que existen entre variables observadas o agrupadas de dos en dos, así como el comportamiento de una variable en función de otra. Se caracteriza por tener una variable predictora y otra de respuesta. Algunas de las metodologías son:

- Diagrama de dispersión: Este gráfico muestra la relación entre dos variables. En el eje de abscisas se presentan los valores de una variable, y en el eje de ordenadas, los valores de la otra. El patrón del diagrama revela qué tipo de relación existe entre las dos variables y su intensidad.
- Coeficientes de correlación: Estos coeficientes miden la fuerza y la dirección de la relación entre dos variables. El coeficiente de correlación de Pearson evalúa la relación lineal, mientras

que el coeficiente de correlación de Spearman se base en el orden de los datos.

- Regresión lineal simple: este análisis modela la relación entre una variable independiente (X) y una variable dependiente (Y). Permite hacer predicciones sobre el valor de y basándose en los cambios en X.

El análisis bivalente se puede crear un gráfico de dispersión trazando una variable frente a otra en un plano cartesiano. Si los datos parecen ajustarse a una línea o curva, entonces existe una relación o correlación entre las dos variables.

## 2.3 ANÁLISIS MULTIVARIANTE

El análisis multivariante se refiere a diferentes métodos que estudian y examinan el efecto simultáneo de múltiples variables. Estos métodos estadísticos multivariados se utilizan para analizar el comportamiento conjunto de más de una variable aleatoria.

Los métodos de análisis multivariantes se utilizan en la evaluación y recopilación de datos estadísticos para aclarar y explicar las relaciones entre las diferentes variables que se asocian a estos datos

Las pruebas multivariantes se utilizan siempre cuando intervienen más de tres variables y el contexto de su contenido no está claro. El objetivo es tanto detectar una estructura, como comprobar los datos en busca de estructuras.

En análisis multivariantes es de carácter multidimensional, por lo tanto, la geometría, el cálculo matricial y las distribuciones multivariantes juegan un papel fundamental (Cuadras, 2004). Debido a la cantidad de datos y variables que intervienen, este análisis se realiza mediante el uso de software.

Según (Peña, 2002), los objetivos del análisis multivariante son los siguientes:

- Crear nuevas variables a partir del conjunto original. Estas variables deben ser menos y corresponderán a transformaciones de las originales, perdiendo la menor cantidad de información posible.
- Evaluar si es posible crear grupos con los datos
- Utilizar agrupaciones definidas para clasificar nueva información.
- Relacionar dos conjuntos de variables.

Los métodos multivariantes pueden subdividirse según diferentes aspectos. Se diferencian en función de si el objetivo es descubrir una estructura dentro de la combinación de datos, o si se trata de comprobar los datos con una determinada estructura a estructura.

## 2.4 METODO STATIS

El método Statis como revelan (Gonzalez et al., 1995), (Estructura estadística de una tabla de tres índices) parte de un grupo de tablas de datos (individuos x variables), obtenidas en situaciones diferentes, un caso común es el de disponer de las mismas informaciones en diferentes instantes. Disponemos entonces de  $r$  tablas  $X_1, X_2, \dots, X_r$ , donde  $X_k$  es de tamaño  $n \times p_k$  que tienen asociadas métricas  $M_k$  para los espacios de los individuos  $IR^{p_k}$  y de una métrica de pesos  $D$  para el espacio de las variables  $IR^n$ .

### El algoritmo

Para el método Statis en pseudocódigo el cual puede escribirse fácilmente en cualquier lenguaje de programación estructurado como Pascal o C. El algoritmo lo dividimos en: Entrada de datos, Infraestructura e Inter estructura.

En la métrica  $D$  consideramos todos los pesos iguales y  $M_k$  la definimos como la identidad correspondiente.

**Entrada:** Se reciben  $r$  tablas (matrices) de datos  $X_k$  individuos x variables de tamaño

$n \times p_k$ , centradas respecto a  $D$ , donde

$n$  es el número de individuos, el mismo en las  $r$  mediciones.

$p_k$  es el número de variables en la  $k$ -ésima medición.

$D = \text{diag}(1/n, 1/n, \dots, 1/n)$  es la matriz de pesos, invariable para las  $r$  mediciones.

**Inter estructura:** Como los individuos en las diferentes mediciones han sido los mismos, comparamos su distribución espacial a través de las matrices  $W_i = X_i X_i^t$ . Para ello usamos la métrica definida por el producto interno

$$\langle W_i, W_j \rangle_{\emptyset} = \text{traza}(W_i W_j) = \langle \vec{W}_i, \vec{W}_j \rangle$$

Donde  $\vec{W}_i \in IR^{n^2}$  es un vector formado por todas las filas de la matriz  $W_i$  para  $i = 1, 2, \dots, r$ .

Paso 1: Calcular las matrices  $W_i$   $n \times n$  dadas por  $W_i = X_i X_j^i$  para  $i = 1, 2, \dots, r$

Paso 2: Genere la matriz

$$X = [\vec{W}_1 \mid \vec{W}_2 \mid \dots \mid \vec{W}_r]_{n^2 \times r}$$

Paso 3: Efectué el Análisis en Componentes Principales del triplete  $(X, D_{\frac{1}{\sigma^2}}, \frac{1}{n^2} I n^2)$  (para efectuar este Análisis en Componentes Principales)

Guardar el primer vector propio de este ACP, que denotamos por  $u = (u_1, u_2, \dots, u_r)$ , y el primer valor propio  $\lambda_1$  para usarlos en el compromiso.

Paso 4: Graficar el círculo de correlaciones (en este aparecerán representados los vectores  $\vec{W}_i$ . Puntos cercanos significaría configuraciones de individuos parecidas y vectores paralelos representará configuraciones homotéticas.

Infraestructura: El estudio evolutivo de individuos y variables lo realizamos con la factorización del compromiso  $\sum_{k=1}^r \beta_k W_k$  obtenida del ACP:  $(\tilde{X}, I, D)$ , donde  $I = \sum_{k=1}^r p_k$ ,  $\beta_k$  está definido y  $\tilde{X}$  se define como sigue:

$$\tilde{X} = [\sqrt{\beta_1} X_1 \mid \sqrt{\beta_2} X_2 \mid \dots \mid \sqrt{\beta_r} X_r]_{n \times \sum_{k=1}^r p_k}$$

Dado que el operador WD del ACP de  $\tilde{X}$  es igual al compromiso, este nos da una representación simultánea de las  $\sum_{k=1}^r p_k$  variables en la base D-ortonormal de sus vectores propios de WD que denominamos ejes del compromiso.

Pero el ACP anterior no resuelve el problema de la representación simultánea de los individuos, para ello definimos las coordenadas de los individuos de la k-ésima tabla como la D-proyección de la matriz  $W_k$  sobre los ejes del compromiso. Para una justificación de esta definición.

Una trayectoria es la representación de un mismo individuo a través de las diferentes tablas, en los ejes del compromiso.

Evolución de las variables :

Paso 5: Calcular  $\beta = (\beta_1, \beta_2, \dots, \beta_r) = \frac{1}{\sqrt{\lambda_1}} u$

Paso 6: Calcular la matriz por bloques  $\tilde{X}$  como sigue:

$$\tilde{X} = [ \sqrt{\beta_1} X_1 \mid \sqrt{\beta_2} X_2 \mid \dots \mid \sqrt{\beta_r} X_r ]_{n \times \sum_{k=1}^r p_k}$$

Paso 7: Efectuar un ACP( $\tilde{X}$ ,  $I_I$ , D) (no reducido).

Paso 8: Guardar la matriz de componentes principales, denotada por  $C = [C_1, C_2, \dots, C_I]$

donde  $I = \sum_{k=1}^r p_k$

Paso 9: Graficar el círculo de correlación de este ACP, en el cual se podrá estudiar la evolución de las variables.

Paso 10: Graficar el plano principal de este ACP, en el que aparecen representados los  $n$  individuos promedio

Evolución de los individuos:

Paso 11: Calcular la matriz IND definida por bloques como sigue:

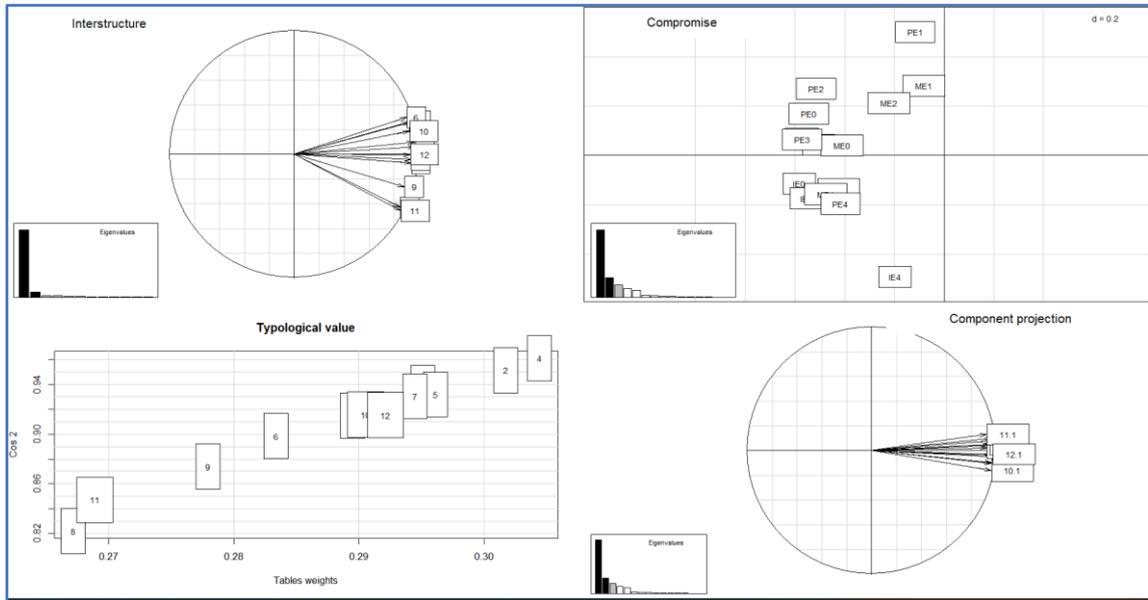
$$\text{IND} = \begin{bmatrix} W_1 \\ W_2 \\ \cdot \\ \cdot \\ W_r \end{bmatrix}_{(rn) \times r}$$

Paso 12: Calcular las coordenadas de los individuos a través de las diferentes tablas, denotada por  $\Omega$ , mediante el siguiente producto matricial:

$$\Omega = \text{IND} \times C$$

Paso 13: Graficar el plano principal, en éste se podrá estudiar la evolución de los individuos.

**Figura 1 Ejemplo método Statis**



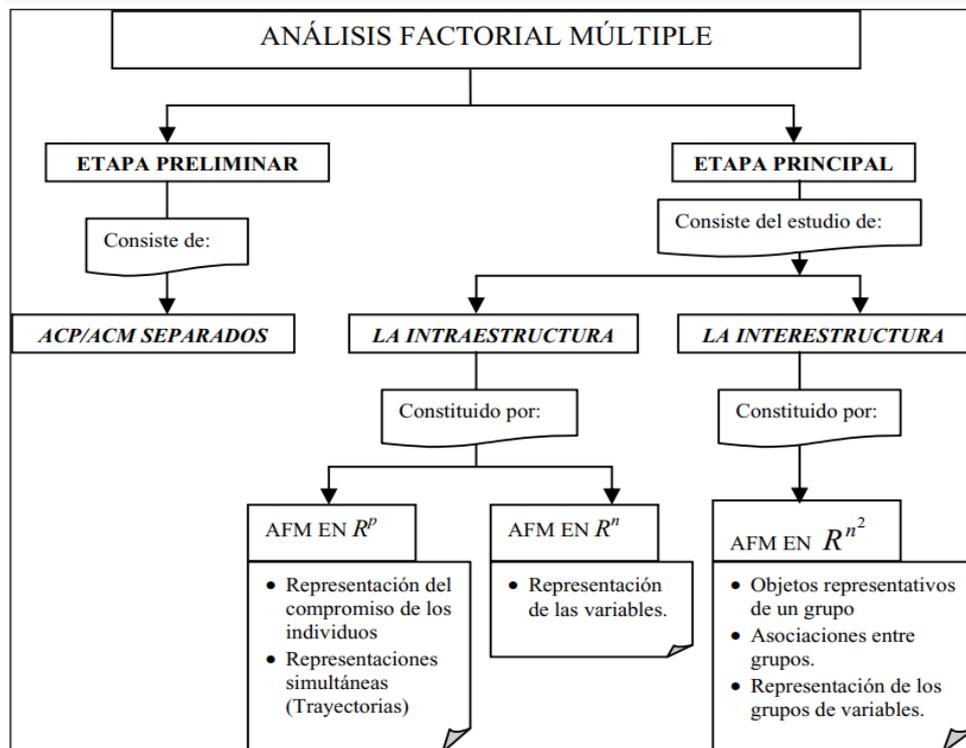
**Fuente: Propia**

## 2.5 ANALISIS FACTORIAL MULTIPLE (AFM)

De la academia francesa de análisis de datos, el Análisis Factorial Múltiple (ESCOFIER & PAGÉS, 1992) permite un análisis comparativo de un conjunto de tablas referidas a los mismos individuos. De esta manera, su objetivo es estudiar la estructura común de las distintas tablas de datos que constituyen el estudio, proporcionando información sobre los individuos homogéneos, así como también acerca de los elementos heterogéneos. El equilibrio de la influencia de los distintos grupos en el análisis se logra mediante una ponderación de las diversas tablas, y de esa manera se posibilita un estudio pertinente de:

- Comparaciones entre los diferentes conjuntos de variables, además de medir el grado de semejanza entre tales conjuntos.
- Cotejos entre las variables que conforman un determinado conjunto y las del resto de los conjuntos.
- Similitudes entre los individuos, observados estos desde las ópticas de los diferentes conjuntos de variables

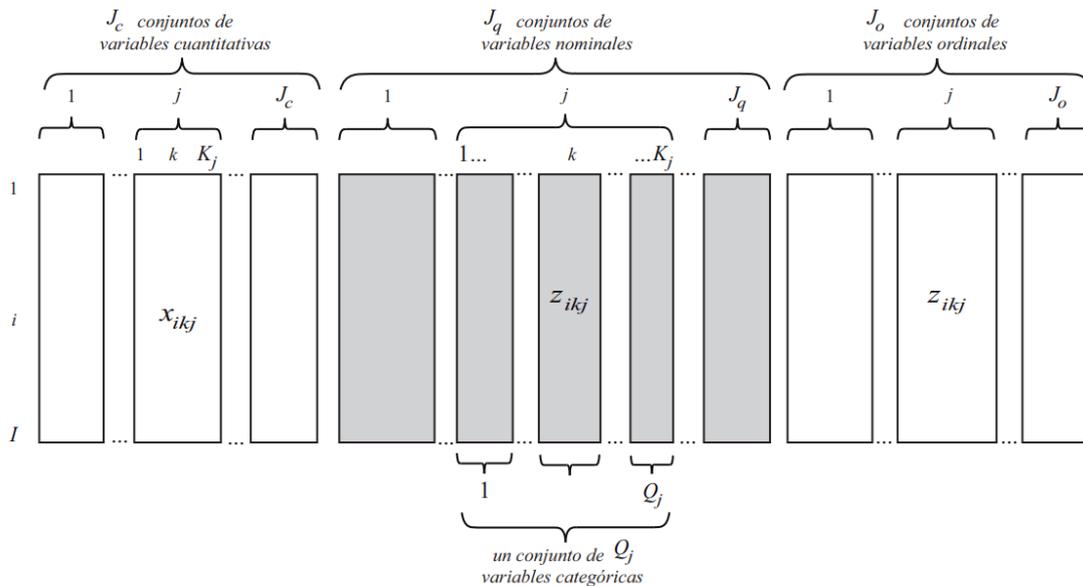
**Figura 2 Esquema general del funcionamiento del AFM**



Fuente: Herrera, J. M. (2007).

El AFM, es uno de los métodos estadísticos multivariados con el que se puede analizar tablas de datos la cual contiene,  $J_c$  grupos o tablas de variables cuantitativas,  $J_q$  grupos o tablas de variables cualitativas (nominales) y los  $J_o$  grupos o tablas de variables cualitativas (ordinales), como la tabla de la figura 3

**Figura 3 Matriz X conformada por la yuxtaposición de tablas**



**Fuente: Becerra Avella, M. O. (2010).**

- $I$  individuos o unidades estadísticas
- $J_c$  número y conjuntos de variables cuantitativas
- $J_q$  número y conjuntos de variables nominales
- $J_o$  número y conjuntos de variables ordinales

## **2.6 Paquetes de R**

R es un software donde se realizó un gran número de Biplots (R Core Team, 2021), El lenguaje de programación R son colecciones de funciones, datos y código compilado que se pueden utilizar para realizar tareas específicas y permiten acceder a herramientas especializadas para análisis de datos, visualización, modelado estadístico y mucho más.

Al tratarse de un software libre, se encuentran multitud de paquetes que permiten realizar este tipo de representaciones gráficas. Los autores de cada paquete incluyen los Biplots que son de interés con las técnicas que se han incluido en el mismo.

La mayor parte de los paquetes presentados a continuación es posible encontrarlos en el repositorio CRAN.

### **2.6.1 Stats**

El paquete de estadística básico de R contiene una función ("biplot") que permite realizar los biplots clásicos sobre los resultados de unas Componentes Principales.

### **2.6.2 MultibiplotGUI**

Este paquete contiene una interfaz gráfica que permite realizar un biplot asociado al Análisis Factorial Múltiple (Librero et al., 2022).

Es posible elegir el tipo de biplot utilizado entre los biplots clásicos y el HJ-biplot.

### **2.6.3 biplotbootGUI**

En el paquete biplotbootGUI (Librero et al., 2019) las autoras introducen métodos bootstrap a los biplots clásicos para obtener las regiones de confianza a través de una interfaz gráfica de R.

Con esta interfaz también es posible realizar los cálculos del Clustering Disjoint Biplot, así como el Análisis de Componentes Principales asociado a este tipo de biplot.

#### **2.6.4 GGEBiplot**

Este paquete contiene las funciones necesarias para construir y evaluar un biplot genotipo ambiente (Dumble, 2022). Es posible realizar la comparación de dos genotipos en todos los ambientes sometidos a estudio, así como establecer la relación entre los ambientes, examinar por separado genotipos y ambientes, o construir un orden de cada uno de ellos.

Recientemente ha sido eliminada de CRAN la interfaz gráfica de este paquete por compatibilidad con las versiones más recientes.

#### **2.6.5 ade4**

El paquete desarrollado por el Laboratorio de Biometría y Biología Evolutiva (UMR CNRS5558) de la Universidad de Lyon contiene un conjunto de funciones que permiten analizar datos Ecológicos y Ambientales utilizando métodos de tres vías (Dray et al., 2022).

Entre la multitud de funciones que se pueden encontrar en este paquete, encontramos diversas representaciones gráficas, una de ellas son los biplots presentados en este capítulo. Utilizando la función "scatter" haremos una representación básica. Existen funciones dentro del paquete para realizar biplots asociados a los mapas factoriales de diferentes Análisis de Correspondencias, a un PLS con el algoritmo NIPALS o a un Análisis de Coordenadas Principales.

#### **2.6.6 MultBiplotR**

Este paquete es una versión mejorada de la interfaz gráfica desarrollada en Matlab y contiene un mayor número de recursos que el paquete libre de partida (Vicente-Villardón, 2021).

Este paquete contiene varias técnicas multivariantes desde una perspectiva de biplot. Son muchos los análisis que se incluyen en este paquete, desde innovaciones hasta técnicas clásicas. Algunos de ellos son los Biplots clásicos, HJ-Biplot, Biplots canónicos, MANOVA Biplots, Análisis de Correspondencia, Análisis de Correspondencia Canónico, STATIS-ACT canónico, Biplots logísticos para datos binarios y ordinales, Despliegue multidimensional, Biplots

externos para el Análisis de Coordenadas Principales o el Escalado Multidimensional, entre muchos otros

### **2.6.7 Esquisse**

Complemento (o paquete) de RStudio diseñado para ayudar en la exploración y visualización inicial de conjuntos de datos. Es una herramienta útil, especialmente para aquellos que son nuevos en R o prefieren una interfaz gráfica para explorar datos de forma interactiva.

Con Esquisse, puedes cargar un conjunto de datos y realizar una variedad de acciones, como: Exploración de datos, Selección de Variable, Visualización de datos, Edición de gráficos, Exportación de gráficos, Código R generado

### **2.6.8 Factoextra**

Visualización y el análisis de resultados de técnicas de análisis factorial, como el análisis de componentes principales (PCA), el análisis de correspondencias (CA), entre otros. Proporciona una serie de funciones para generar gráficos informativos y realizar análisis adicionales sobre los resultados de estos métodos

### **2.6.9 SISTEMA DE INFORMACIÓN HOSPITALARIO (HIS)**

Hospital Information System (HIS), sistema de información hospitalario por sus siglas en inglés, también conocido como sistema de información de salud, es un sistema de información que permite, generar, procesar, modificar, recuperar y almacenar información así como también reinterpretar datos, emitir reportes e informes de tipo médico-administrativos de una institución hospitalaria, optimizando los recursos humanos y materiales, para ofrecer una atención de calidad a los pacientes sobre los servicios de salud.

### **2.6.10 ESTANDARIZACION**

Técnicas en el preprocesamiento de datos, especialmente en el ámbito del aprendizaje automático y la estadística. Tiene como objetivo principal hacer que las características de los datos sean más comparables o interpretables para los algoritmos de modelado como:

- La estandarización implica transformar los datos para que tengan una media de cero y una desviación estándar de uno.
- Se calcula restando la media y luego dividiendo por la desviación estándar de los datos:

$$x_{\text{estandarizado}} = \frac{x - \mu}{\sigma}$$

- Este método es útil cuando los datos tienen diferentes medias y variabilidades, lo que puede afectar al rendimiento de algunos algoritmos

La estandarización es útil cuando queremos que todas tengan la misma media y desviación estándar

### **2.6.11 Algoritmo de Ward**

Método de agrupamiento jerárquico aglomerativo utilizado en análisis de clústeres. El objetivo principal del algoritmo de Ward es minimizar la varianza total dentro de cada clúster, la función es de Inicialización, Cálculo de la distancia, Fusión de clústeres, Actualización de la matriz de distancia, Repetición.

La clave del algoritmo de Ward es la definición de la distancia entre clústeres. En este caso, la distancia de Ward mide la diferencia entre los clústeres antes y después de fusionarse, tomando en cuenta la varianza de los datos. Esto significa que, durante cada fusión, se selecciona la combinación de clústeres que minimiza el incremento en la varianza total dentro de los clústeres combinados

### **2.6.12 CLUSTER**

Agrupaciones de elementos que comparten características similares entre sí, y son útiles en análisis de datos para identificar patrones y estructuras subyacentes. Los algoritmos de agrupamiento (o clustering) se utilizan para agrupar datos en función de la similitud entre ellos.

La elección del algoritmo de agrupamiento depende del tipo de datos que estás tratando, la estructura de los clústeres que esperas

encontrar y otros factores como la escalabilidad y la interpretación de los resultados como:

K-Means, algoritmo de particionamiento que agrupa de los datos en  $k$  clústeres, donde  $k$  es un número predefinido. Cada clúster está representado por su centroide, que es la media de todos los puntos asignados a ese clúster.

# CAPÍTULO 3

## 3. METODOLOGIA

En el marco de esta investigación se realizó con matrices de datos aplicando Power Bi herramienta moderna de visualización y en el lenguaje de programación de R-Studio.

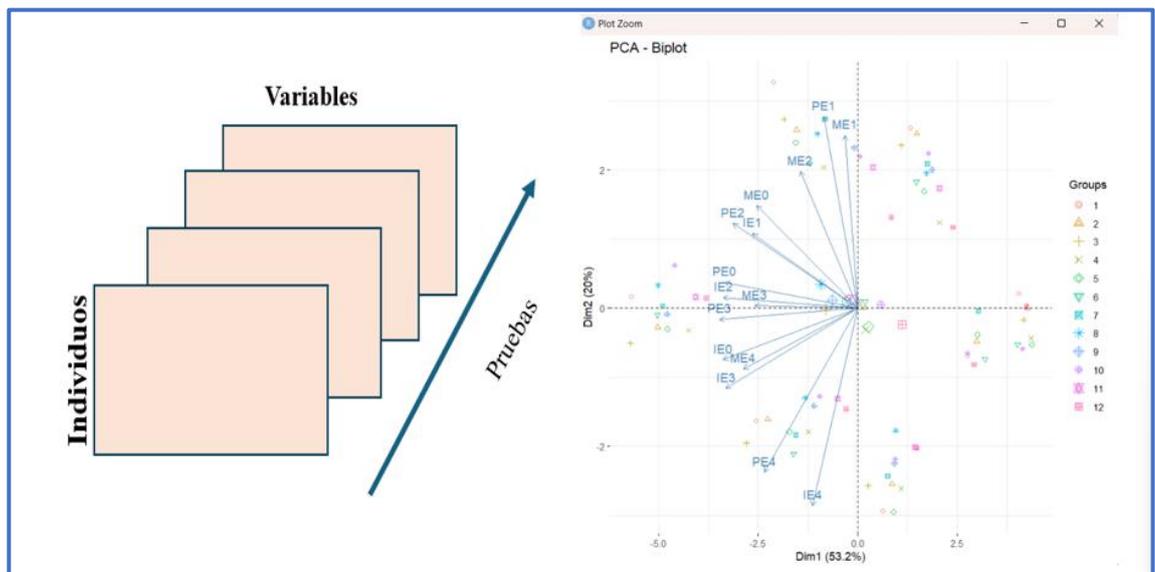
Los datos obtenidos son mediante del Sistema Hospitalario donde se encuentran las variables de los Estadios y una matriz de los pacientes provenientes de una institución de salud con  $n > 9.400$ , previa analítica de datos se determinó resumen estadístico que proporcionó información clave sobre el conjunto de datos. La analítica de datos nos permitió las medidas de tendencia central, dispersión y algunas características adicionales como Mínimo, 1er Cuartil, Mediana, Media, 3er Cuartil, Máximo.

De la matriz se realizó la imputación de datos proceso de estimar o completar valores faltantes en el conjunto de datos. La imputación de datos es importante porque los valores faltantes pueden afectar negativamente el rendimiento de los análisis estadísticos y los modelos predictivos si no se manejan adecuadamente. El método utilizado de imputación fue la mediana.

De los gráficos utilizada para un descriptiva preliminar un boxplot visualización de la distribución de un conjunto de datos e identificación la presencia de valores atípicos (outliers), Clusterización agrupación de datos en función de la similitud entre ellos.

Posteriormente se aplicó la técnica multivariante STATIS para analizar las variables codificadas para los estadios de cáncer en pacientes de la institución prestadora de servicios de salud, con el objetivo de comparar y analizar las relaciones entre los diferentes conjuntos de datos, en un promedio ponderado óptimo, que se lo conoce como matriz compromiso o consenso, que luego analiza a través de un análisis de componentes principales, la estructura común entre las observaciones y finalmente proyecta cada uno de los conjuntos de datos originales en el compromiso, para analizar comunalidades y discrepancias (Abdi et al., 2012)

**Figura 4 Estructura metodológica Statis**



**Fuente: Propia**

Este proceso se realizó con diferente grupo de datos de individuos, calculados en un mismo grupo de variables, en distintos tiempos.

STATIS consta de los siguientes pasos importantes:

- Preprocesamiento de cada tabla centralizando y normalizando cada columna, de tal manera que la suma de todos los valores al cuadrado de los elementos de cada tabla sea igual a uno (Gallardo, 2021).
- Se adquiere la matriz de covarianzas correlaciones para cada tabla, posteriormente se determina la matriz de correlaciones vectorial de Hilbert-Smith.
- Se realizará la descomposición de valores singulares, con la finalidad de poder obtener conjunto de pesos que servirán para ponderar cada K tabla, también se conseguirán unos  $\cos^2$ , ambos permitirán para la elaboración de la matriz compromiso, la cual evidencia un arreglo lineal ponderada de cada K tabla.
- Los gráficos que se lograrán del análisis multivariante de K vías Statis, permitirá fortificar el control, y estar al tanto que variables han causado la variabilidad en el proceso.

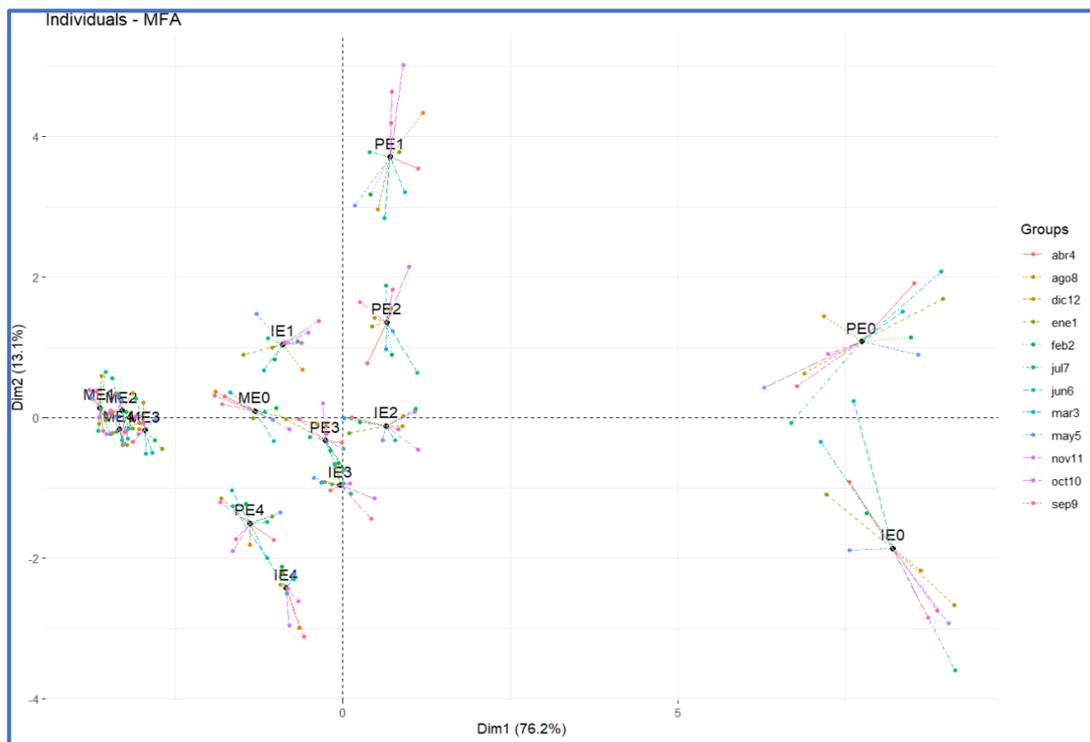
Posteriormente se realizó un Análisis Factorial Múltiple (AFM) que permite un análisis comparativo de un grupo de tablas referidas a los mismos individuos. De esta manera, su objetivo es estudiar la

estructura común de las distintas tablas de datos que constituyen el estudio, proporcionando información sobre los individuos homogéneos, así como también acerca de los elementos heterogéneos.

El AFM busca poner de realce los principales factores que dan cuenta de la variabilidad de los individuos, descritos estos, de manera objetiva, por los diferentes grupos de variables que reducen la estructura de partición de la información.

La finalidad, como en todos los métodos factoriales, es reducir la dimensión de la tabla para así poder proceder a su estudio. Permite, además de analizar las semejanzas entre individuos y las relaciones entre variables como lo hacen los métodos factoriales clásicos, comparar globalmente los grupos de variables.

**Figura 5 Análisis Factorial Múltiple**

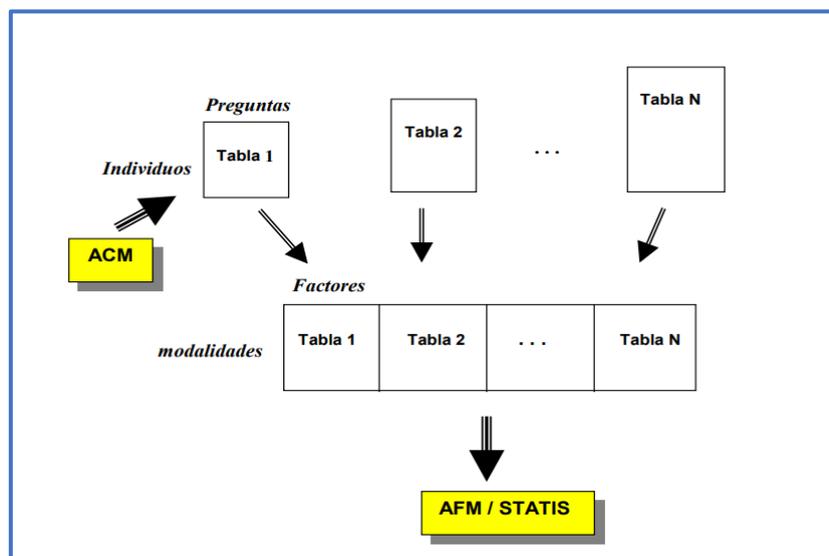


**Fuente: Propia**

El AFM es más exhaustivo que el STATIS, y da una descripción más detallada de las relaciones entre las variables de las diferentes tablas, no obstante, el STATIS es más clarificador en cuanto a las relaciones entre las tablas diferentes, dando menos importancia a la

infraestructura de las N tablas analizadas. De forma esquemática, el proceso antes descrito sería el siguiente:

**Figura 6 Esquema AFM / STATIS**



**Fuente: propia**

### 3.1 Descripción de Investigación.

El principal objetivo de este trabajo es poner de manifiesto, las metodologías más habituales en el análisis de tablas múltiples presentan entre sí. Los métodos seleccionados han sido: el método STATIS y el Análisis Factorial Múltiple. El estudio comparativo se realizó principalmente, desde el punto de vista de la interpretación.

### 3.2 Autorización Información

Previa autorización para acceder la información de la prestadora de servicios de salud oncológica de la ciudad de Guayaquil, recopilados durante el año 2023 de los cuales se consideró las variables a estudiar:

- Tipo de diagnósticos:
  - Tumores Mama
  - Tumores Cabeza y Cuello
  - Tumores Órganos Digestivos
  - Tumores Órganos Genitales Femeninos
  - Tumores Órganos Genitales Masculinos
  - Tumores Piel Sarcoma Y Partes Blandas
  - Tumores-Mielomas-Leucemias Otros

- Tipo de Estadios:
  - Estadio 0
  - Estadio 1
  - Estadio 2
  - Estadio 3
  - Estadio 4
- Tipo de Clientes:
  - I.E.S.S.
  - M.S.P.
  - Pacientes Institucionales (PCTE INST.)
- Periodo:
  - Enero 2023 a diciembre 2023

### 3.3 Obtención de la Data

La información fue obtenida en dos matrices de datos utilizando Power BI herramienta moderna de visualización de datos, se procedió a relacionar dos tablas donde contienen información de cantidad de los pacientes por cliente y otra tabla de tipo de estadios, relacionándola por medio de clave primaria que tiene los pacientes en cada tabla obtiene k matrices.

**Figura 7 Obtención de la Data en Power BI**

		CANTIDADES DE PACIENTES POR TIPO DE ESTADIO - PERIODO 2023																											
ESTADIO	TIPO DE CLIENTE	ESTADIO 0				ESTADIO 1				ESTADIO 2				ESTADIO 3				ESTADIO 4											
		GEN	FEB	MAR	ABR	GEN	FEB	MAR	ABR	GEN	FEB	MAR	ABR	GEN	FEB	MAR	ABR	GEN	FEB	MAR	ABR								
IES	ESTADIO 0	348	30	231	172	250	86	52	246	96	211	152	197	26	52	295	156	203	164	209	45	38	234	81	182	142	177	38	52
IES	ESTADIO 1	72	65	40	45	32	13	6	72	58	47	44	39	12	9	178	73	45	44	33	19	9	64	67	47	39	32	9	4
IES	ESTADIO 2	189	26	70	78	51	2	24	181	64	64	25	52	2	22	189	60	69	77	56	4	24	178	42	62	62	45	2	24
IES	ESTADIO 3	118	38	108	45	54	13	25	117	53	102	45	47	18	28	112	54	119	47	59	7	18	111	22	91	41	58	9	13
IES	ESTADIO 4	52	11	81	19	124	4	14	54	18	83	21	127	4	11	57	19	86	21	132	4	13	47	19	72	18	182	8	14
MSP	ESTADIO 0	26	26	43	51	16	15	22	26	27	46	41	16	17	24	25	21	40	46	18	15	21	29	29	32	45	14	12	21
MSP	ESTADIO 1	6	8	3	11	5	3	3	9	3	10	4	5	5	2	2	9	4	10	7	2	3	4	5	3	8	3	2	5
MSP	ESTADIO 2	9	4	12	34	6	4	6	8	5	9	21	5	4	6	7	4	7	21	6	4	6	6	2	8	24	5	3	7
MSP	ESTADIO 3	21	8	20	32	7	1	14	19	4	28	17	8	1	13	22	5	56	28	8	1	11	17	4	21	18	3	2	10
MSP	ESTADIO 4	14	5	14	11	7	2	9	11	5	12	12	8	1	7	12	7	15	12	9	2	7	18	6	12	11	11	2	7
PCTE INST.	ESTADIO 0	509	127	189	281	140	83	52	433	193	171	208	130	61	45	638	128	197	248	189	73	41	418	85	162	198	162	64	38
PCTE INST.	ESTADIO 1	97	176	95	89	35	25	12	91	149	25	75	34	19	11	84	151	35	79	26	32	12	17	112	25	78	28	17	10
PCTE INST.	ESTADIO 2	229	72	60	118	28	12	14	215	15	48	112	38	9	22	248	64	17	128	14	8	13	200	26	40	182	26	6	14
PCTE INST.	ESTADIO 3	146	28	129	11	31	7	21	146	18	115	64	23	8	18	148	29	122	73	31	7	28	144	26	96	88	22	5	18
PCTE INST.	ESTADIO 4	78	12	70	30	98	8	14	72	9	68	28	71	5	19	81	6	41	28	37	2	14	61	8	47	21	73	6	13

Fuente: Propia

Posteriormente se exportó a Excel la información dando forma según la estructura matricial para el método Statis en forma horizontal, al analizar la data registrar valores faltantes realizando la imputación del dato utilizando la mediana.

Se Seleccionó 7 diagnóstico de los cuales se determinó la cantidad de pacientes por cada estadio (5 estadios que corresponden grupo de cáncer) de los clientes IESS(I), MSP(M) PCTE INST. (P). A continuación, se procedió en los primeros 12 meses del 2023 para evaluar los diagnósticos.

Los Estadios se reemplazó por Estadio 0=E0, Estadio 1=E1, Estadio 2=E2, Estadio 3=E3, Estadio 4=E4

Con la información se analizó:

Los diagnósticos según la cantidad de pacientes, utilizando los estadios. Los datos brutos se presentan en la Figura 8. Los objetivos del análisis fueron dos: (1), descubrir en que mes existe mayor incidencia de estadio y (2) obtener que cantidad cliente del sistema de salud deriva más pacientes

**Figura 8 Estructura matriz Método Statis en R Studio**

mes	DIAGNOSTICO	IE0	IE1	IE2	IE3	IE4	ME0	ME1	ME2	ME3	ME4	PE0	PE1	PE2	PE3	PE4
1	T1	364	72	189	118	52	35	6	9	21	14	509	97	231	145	78
1	T2	92	65	46	30	11	26	8	4	6	5	127	176	73	26	13
1	T3	221	40	70	108	81	43	3	12	26	14	199	36	60	129	70
1	T4	173	45	78	43	19	51	11	34	32	11	241	89	118	71	30
1	T5	202	32	51	54	134	16	3	6	7	7	142	35	30	31	90
1	T6	46	13	3	10	4	15	5	4	1	2	80	35	12	7	8
1	T7	52	6	24	23	14	22	3	6	14	9	52	12	14	21	14
2	T1	366	72	181	117	56	36	3	9	19	11	433	91	215	146	72
2	T2	96	66	44	33	10	27	9	5	4	5	113	149	55	28	9
2	T3	211	47	64	102	83	46	3	9	28	13	171	35	48	115	68
2	T4	152	44	81	45	21	41	10	31	27	12	208	75	112	64	28
2	T5	197	39	52	47	127	16	4	5	6	8	130	24	30	23	71
2	T6	36	12	3	10	4	17	4	4	1	1	61	19	9	6	5
2	T7	52	8	22	20	11	24	3	5	13	7	45	11	22	16	10
3	T1	395	78	189	112	57	35	2	7	22	12	508	94	244	149	81
3	T2	106	73	50	34	10	31	9	4	5	7	120	131	58	26	9
3	T3	203	45	69	110	86	40	4	7	36	15	197	35	57	122	81
3	T4	166	44	77	47	21	46	10	31	26	12	240	79	120	73	28
3	T5	209	33	56	59	132	18	7	5	6	9	149	26	34	31	97
3	T6	45	16	4	7	4	15	2	4	1	2	73	32	8	7	2
4	T1	356	64	178	111	47	29	4	6	17	10	418	77	203	144	61
4	T2	81	67	43	22	10	20	5	2	4	6	85	112	36	26	8
4	T3	183	47	63	91	72	32	3	8	21	12	163	25	44	96	67
4	T4	142	36	62	41	18	45	8	24	18	11	199	78	102	68	31
4	T5	177	32	45	50	103	14	3	5	3	11	102	28	26	22	71
4	T6	38	9	2	9	4	12	2	3	2	2	64	27	6	5	5
5	T1	352	65	188	121	51	30	4	10	17	13	454	76	206	136	61
5	T2	92	69	45	23	10	24	7	3	7	6	82	115	54	23	13
5	T3	183	43	62	87	72	36	2	8	27	14	164	40	59	102	70
5	T4	132	40	64	37	18	51	8	29	30	11	216	73	117	71	35
5	T5	196	31	57	55	117	14	3	4	5	9	113	24	24	25	80
5	T6	37	9	4	7	4	13	1	4	2	1	79	20	9	5	5
5	T7	51	3	27	16	18	28	4	6	11	6	42	8	17	20	15

**Fuente: Propia**

K= 12 Periodos (meses)

J= diferentes variables en cada k tabla

I= mismos individuos 7 diagnósticos (3 Cliente, 5 por cada estadio)

# CAPÍTULO 4

## RESULTADOS

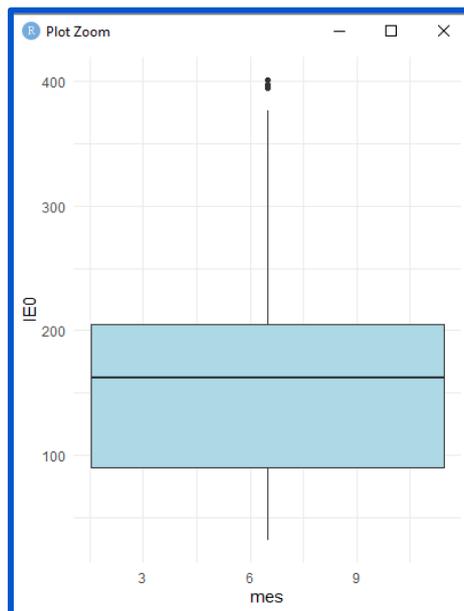
### 4.1 Análisis Explicativo

De cada variable mencionada anteriormente se obtuvo la media, mínimo, máximo, mediana y los cuartiles según Tabla 1.

**Tabla 1 Resumen de las variables.**

<b>IE0</b> Min. : 32.00 1st Qu.: 89.75 Median : 162.50 Mean : 170.05 3rd Qu.: 205.25 Max. : 401.00	<b>IE1</b> Min. : 3.00 1st Qu.: 32.00 Median : 44.00 Mean : 42.58 3rd Qu.: 59.75 Max. : 78.00	<b>IE2</b> Min. : 2.00 1st Qu.: 43.00 Median : 56.50 Mean : 69.58 3rd Qu.: 71.00 Max. : 195.00	<b>IE3</b> Min. : 7.00 1st Qu.: 27.25 Median : 42.50 Mean : 56.15 3rd Qu.: 90.75 Max. : 127.00	<b>IE4</b> Min. : 4.00 1st Qu.: 11.00 Median : 20.50 Mean : 42.68 3rd Qu.: 71.25 Max. : 134.00
<b>ME0</b> Min. : 12.00 1st Qu.: 18.25 Median : 28.50 Mean : 29.31 3rd Qu.: 36.75 Max. : 51.00	<b>ME1</b> Min.: 1.000 1st Qu.: 3.000 Median : 5.000 Mean : 5.095 3rd Qu.: 7.000 Max. : 11.000	<b>ME2</b> Min.: 2.000 1st Qu.: 4.000 Median : 6.000 Mean : 8.959 3rd Qu.: 9.750 Max. : 34.000	<b>ME3</b> Min.: 1.00 1st Qu.: 5.25 Median : 10.50 Mean : 13.49 3rd Qu.: 20.75 Max. : 36.00	<b>ME4</b> Min.: 1.000 1st Qu.: 6.000 Median : 10.000 Mean : 8.568 3rd Qu.: 11.000 Max. : 15.000
<b>PE0</b> Min. : 23.0 1st Qu.: 80.5 Median : 128.5 Mean : 168.0 3rd Qu.: 199.0 Max. : 509.0	<b>PE1</b> Min. : 8.00 1st Qu.: 26.00 Median : 43.00 Mean : 59.65 3rd Qu.: 81.50 Max. : 176.00	<b>PE2</b> Min. : 6.00 1st Qu.: 22.00 Median : 48.00 Mean : 71.55 3rd Qu.: 110.75 Max. : 244.00	<b>PE3</b> Min. : 4.00 1st Qu.: 22.00 Median : 31.00 Mean : 55.92 3rd Qu.: 89.25 Max. : 149.00	<b>PE4</b> Min. : 2.00 1st Qu.: 9.00 Median : 30.50 Mean : 37.77 3rd Qu.: 65.25 Max. : 97.00

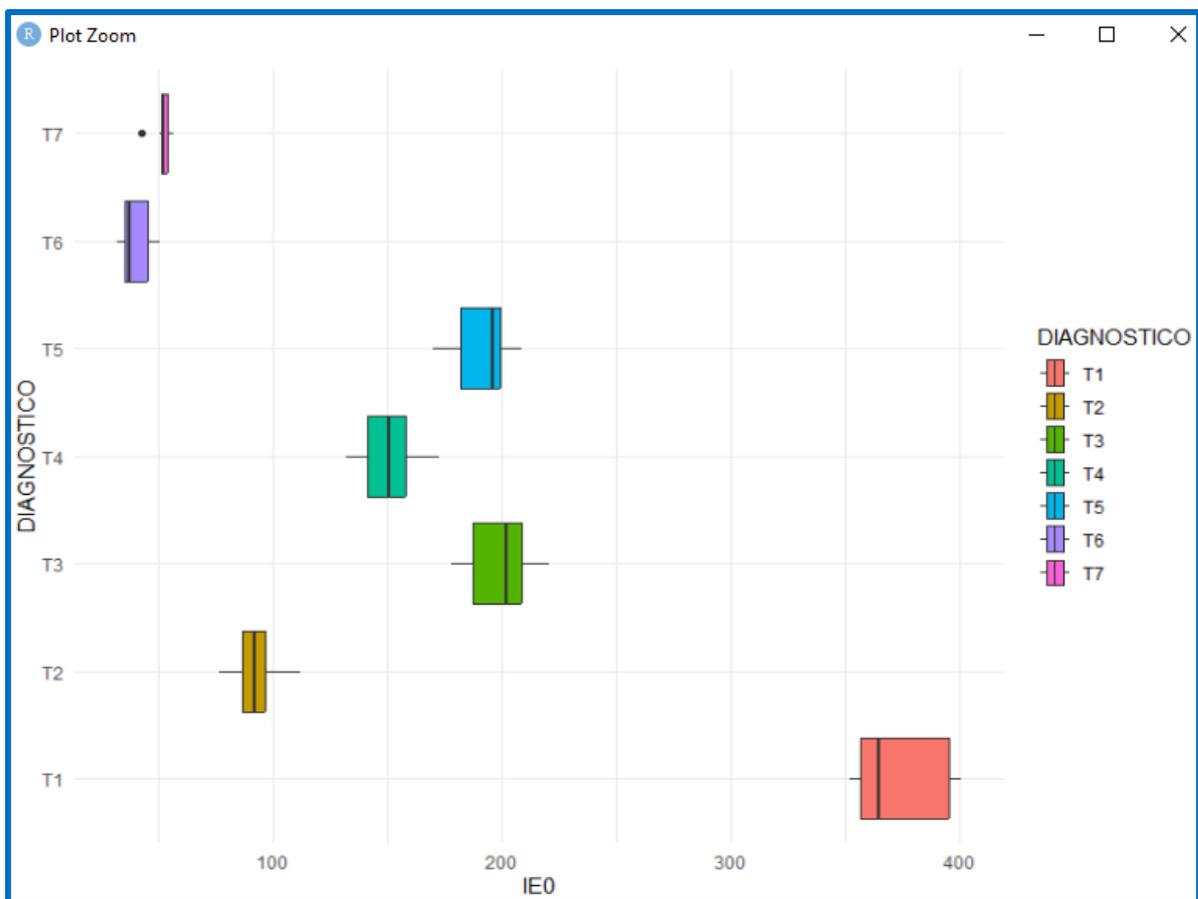
**Figura 9 Resumen Boxplot variable IE0**



**Fuente: Propia**

Además, mediante un Boxplot (Figura 9), se observó los tipos de diagnóstico con relación a la variable IE0 el cual tiene la tendencia central del conjunto de datos superior comparado con las otras variables, se visualiza el tumor mama (T1) con gran cantidad de pacientes teniendo el cuartil 3 con el 75% de los datos, seguidos por el tumor órgano digestivo (T3) y tumor órganos genitales masculino (T5) y con menor cantidad de pacientes tumor piel, sarcoma y partes blandas (T6).

**Figura 10 Boxplot diagnóstico con relación IE0**

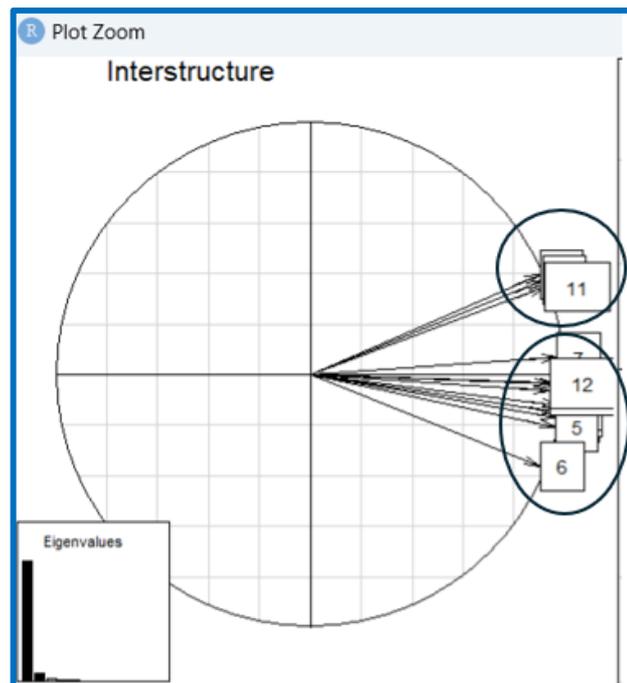


**Fuente: Propia**

## 4.2 Análisis Método Statis

**Inter-estructura:** comparaciones de la tablas donde se ha reflejado las k tablas presentan estructuras similares (periodos-mes) se ha determinado dos grupos: el grupo 1 han calificado en la misma forma, tiene un valor parejo (mes junto al 11 tiene ángulos muy pequeños tiene estadios similares) y el grupo 2 (meses 12-7-5 tiene ángulos muy pequeños en estadios similares), también se puede apreciar si comparamos el mes 6 con el mes 11 tiene un distanciamiento con un ángulo grandes es decir que tiene diferentes situaciones de estadios entre los dos periodos. Entre más cortito están los vectores mayor correlación, entre más separado se encuentre los vectores correlación bajas, figura 10

**Figura 11 Inter-estructura Statis**



**Fuente: Propia**

**Coefficientes de correlación vectorial RV** Coeficientes de correlación altos, indican estructura similar entre los periodos de los meses (1 al 12), esto se ve reflejado en el gráfico de la Inter-estructura

Se demuestra en la tabla 2, lo bordeado en azul tiene correlación vectorial alta y lo resaltado con rojo tiene correlación vectorial baja no hay similitud

**Tabla 2 Coeficiente de correlación vectorial**

```
> stasis1$RV#coeficientes de correlación vectorial RV
```

	1	2	3	4	5	6	7	8	9	10	11	12
1	1.0000000	0.9801105	0.9517747	0.9680336	0.9811768	0.9202896	0.9194913	0.8006094	0.8248538	0.9026259	0.8424667	0.8786390
2	0.9801105	1.0000000	0.9682847	0.9601046	0.9719536	0.9164593	0.9356504	0.8141143	0.8392719	0.9019438	0.8450015	0.9085336
3	0.9517747	0.9682847	1.0000000	0.9554544	0.9510060	0.8660573	0.9297516	0.8562556	0.8688976	0.8581715	0.8776056	0.8872026
4	0.9680336	0.9601046	0.9554544	1.0000000	0.9712763	0.9214834	0.9322361	0.8381643	0.8452553	0.9080510	0.8565889	0.9029223
5	0.9811768	0.9719536	0.9510060	0.9712763	1.0000000	0.9324394	0.9078274	0.7912933	0.7990588	0.8870324	0.8253783	0.8832041
6	0.9202896	0.9164593	0.8660573	0.9214834	0.9324394	1.0000000	0.8569747	0.6766082	0.7021060	0.8810227	0.7184879	0.8658055
7	0.9194913	0.9356504	0.9297516	0.9322361	0.9078274	0.8569747	1.0000000	0.8907909	0.9188218	0.8966105	0.9030384	0.9241309
8	0.8006094	0.8141143	0.8562556	0.8381643	0.7912933	0.6766082	0.8907909	1.0000000	0.9562354	0.8268383	0.9630641	0.8343007
9	0.8248538	0.8392719	0.8688976	0.8452553	0.7990588	0.7021060	0.9188218	0.9562354	1.0000000	0.8264248	0.9691921	0.8279677
10	0.9026259	0.9019438	0.8581715	0.9080510	0.8870324	0.8810227	0.8966105	0.8268383	0.8264248	1.0000000	0.8366670	0.9331013
11	0.8424667	0.8450015	0.8776056	0.8565889	0.8253783	0.7184879	0.9030384	0.9630641	0.9691921	0.8366670	1.0000000	0.8455611
12	0.8786390	0.9085336	0.8872026	0.9029223	0.8832041	0.8658055	0.9241309	0.8343007	0.8279677	0.9331013	0.8455611	1.0000000

**Coseno<sup>2</sup> – Calculo de los pesos en la primera componente con la descomposición de los valores singulares.**

K-tablas que más aportan con información para construir la matriz consenso (pesos altos) y a su vez son las tablas que son mejor representadas por la matriz consenso (cos<sup>2</sup> altos).

Los meses 5 y 7 tienen pesos altos el cual contribuyen con más información para la construcción de la matriz consenso como se demuestra en la tabla 3 donde el mes 5 con 0.9540873 y el mes 7 con 0.9621205 en cambio el mes 8 con 0.8871437 tiene coseno cuadrado bajo no contribuye en la matriz consenso. Figura 10

**Tabla 3 Coseno2**

**statis1\$cos2#coseno2**

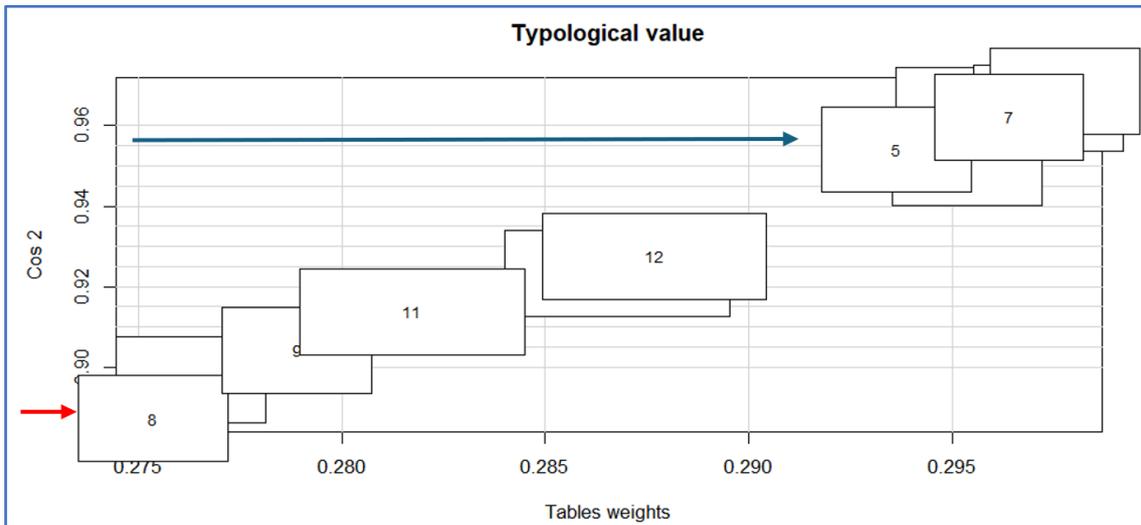
1	2	3	4	5	6	7	8	9	10	11	12
0.9639456	0.9644160	0.9507390	0.9686503	0.9540873	0.8967121	0.9621205	0.8871437	0.9042165	0.9233014	0.9137306	0.9275000

**Tabla 4 Pesos k-tablas**

**statis1\$RV.tabw#pesos para las k-tablas**

1	2	3	4	5	6	7	8	9	10	11	12
0.2954512	0.2973554	0.2953723	0.2977774	0.2936291	0.2762863	0.2963948	0.2753503	0.2788973	0.2867685	0.2817297	0.2876818

**Figura 12 Coseno2**



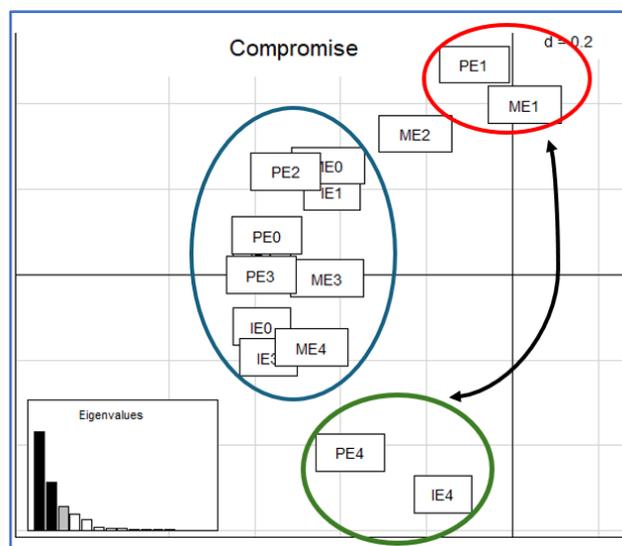
**Fuente: Propia**

### Matriz Compromiso

Proyección de un modelo lineal ponderado con cada k tablas donde refleja la estructura promedio de las puntuaciones de los clientes según los tipos de estadios.

Lo que se encuentran dentro del eclipse de color azul tiene similares características bajo el punto de las variables analizadas, mientras que los encuentra en los eclipse rojo y verde tiene características opuestas.

**Figura 13 Matriz Compromiso**

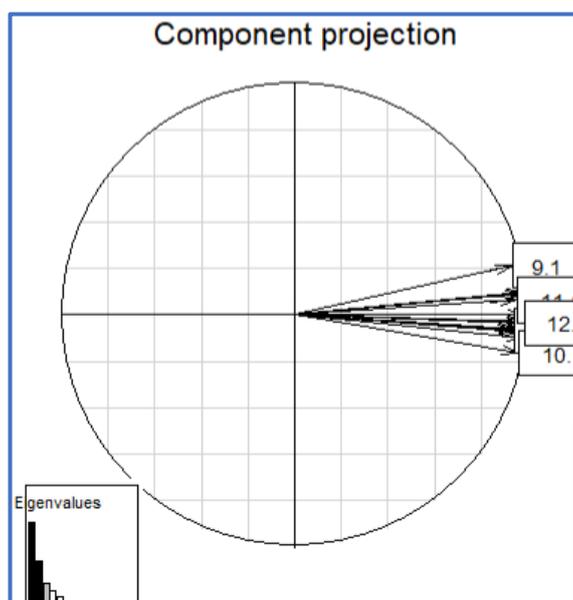


**Fuente: Propia**

## Proyección de Componentes

Por cada periodo K- Tabla se proyecta en el primer eje, valor máximo es 1 donde retiene importante información, por lo tanto, el STATIS es idóneo para este análisis

**Figura 14 Proyección de Componentes**



**Fuente: Propia**

## Valores propios para la inter-estructura

El análisis estadístico con el método Statist ha reflejado en los 12 meses el primer eje - componente representa el valor de la inter-estructura el 90 % de la varianza acumulada y el segundo eje componente explican un 95% de la varianza acumulada, cabe indicar que se ha trabajado en los dos primeros ejes en este modelo.

**Tabla 5 Valores propios para la inter-estructura**

statis1\$RV,eig#valores propios para la inter-estructura													sumatoria
Valores Propios	10,7315	0,66761	0,22254	0,0954	0,07673	0,0599	0,04494	0,0344	0,02777	0,01691	0,01346	0,00883	12
% Varianza	0,89722	0,05582	0,01861	0,00798	0,00642	0,00501	0,00376	0,00288	0,00232	0,00141	0,00113	0,00074	
% Varianza acumulada	90%	95%	97%	98%	99%	99%	99%	100%	100%	100%	100%	100%	

## Valores propios para el compromiso

El análisis estadístico con el método Statist ha reflejado en las variables de cliente según los tipo de estadios el primer eje - componente representa el valor de compromiso el 53 % de la varianza acumulada y el segundo eje - componente explican un 74% de la varianza acumulada, cabe indicar que se ha trabajado en los dos primeros ejes en este modelo.

**Tabla 6 Valores propios para el compromiso**

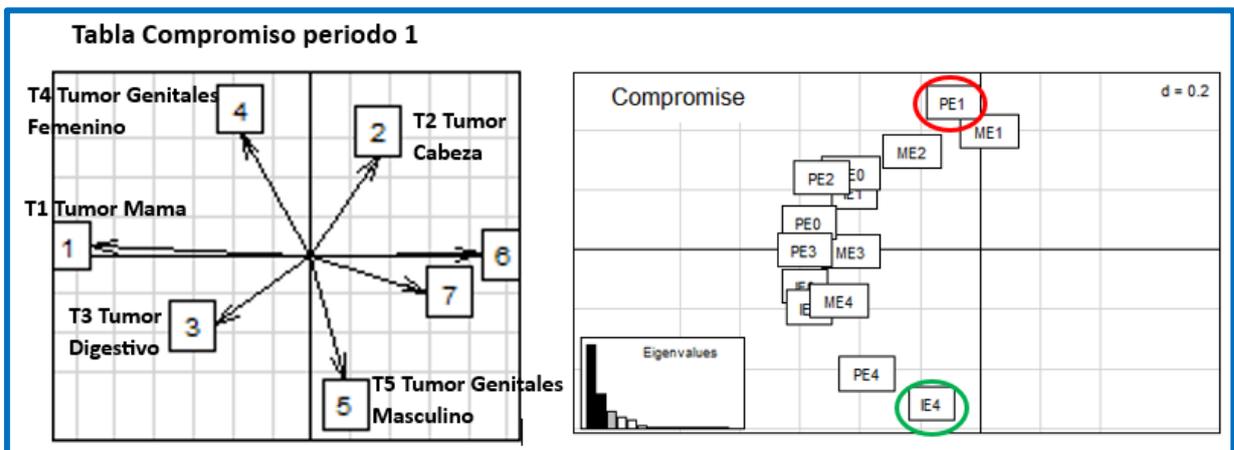
statis1\$C.eig#valores propios para el compromiso																sumatori
Valores Propios	2,94771	1,19965	0,60452	0,40023	0,26362	0,06702	0,04559	0,03181	0,02006	0,01131	0,0081	0,00391	0,00315	0,00171	0,00137	5,610
% Varianza	0,52546	0,21385	0,10776	0,07135	0,04699	0,01195	0,00813	0,00567	0,00358	0,00202	0,00144	0,0007	0,000562	0,00031	0,00024	
% Varianza acumulada	53%	74%	85%	92%	97%	98%	99%	99%	99%	100%	100%	100%	100%	100%	100%	

## Intra-estructura

Para la interpretación, nos basamos en el grafico del compromiso con el **Periodo 1**: la variable Pcte Inst. estadio1 - PE1 obtuvo puntuaciones por encima del valor promedio en diagnóstico de T2-Tumor de Cabeza y por debajo del valor promedio en diagnóstico de T3 Tumor Digestivo.

La variable IESS estadio 4 -IE4 obtuvieron puntuaciones por encima del valor promedio en de T5-Tumor Genitales Masculino y por debajo del valor promedio en diagnóstico de T4 Tumor Genitales Femenino

**Figura 15 Intra-estructura**



Fuente: Propia

### 4.3 Análisis Factorial Múltiple

Para el análisis factorial múltiple se utilizaron las matrices en forma horizontal, donde se tienen k tablas según los periodos de enero a diciembre 2023

**Tabla 7 K tablas Análisis factorial múltiple**

CLIENTESTADIO	T1	T2	T3	T4	T5	T6	T7	T1	T2	T3	T4	T5	T6	T7	T1	T2	T3	T4	T5	T6	T1	T2	T3	T4	T5	T6	T1	T2	T3	T4	T5	T6	T7
IE0	364	92	221	173	202	46	52	366	96	211	152	197	36	52	395	106	203	166	209	45	356	81	183	142	177	38	352	92	183	132	196	37	51
IE1	72	65	40	45	32	13	6	72	66	47	44	39	12	8	78	73	45	44	33	16	64	67	47	36	32	9	65	69	43	40	31	9	3
IE2	189	46	70	78	51	3	24	181	44	64	81	52	3	22	189	50	69	77	56	4	178	43	63	62	45	2	188	45	62	64	57	4	27
IE3	118	30	108	43	54	10	23	117	33	102	45	47	10	20	112	34	110	47	59	7	111	22	91	41	50	9	121	23	87	37	55	7	16
IE4	52	11	81	19	134	4	14	56	10	83	21	127	4	11	57	10	86	21	132	4	47	10	72	18	103	4	51	10	72	18	117	4	18
ME0	35	26	43	51	16	15	22	36	27	46	41	16	17	24	35	31	40	46	18	15	29	20	32	45	14	12	30	24	36	51	14	13	28
ME1	6	8	3	11	3	5	3	3	9	3	10	4	4	3	2	9	4	10	7	2	4	5	3	8	3	2	4	7	2	8	3	1	4
ME2	9	4	12	34	6	4	6	9	5	9	31	5	4	5	7	4	7	31	5	4	6	2	8	24	5	3	10	3	8	29	4	4	6
ME3	21	6	26	32	7	1	14	19	4	28	27	6	1	13	22	5	36	26	6	1	17	4	21	18	3	2	17	7	27	30	5	2	11
ME4	14	5	14	11	7	2	9	11	5	13	12	8	1	7	12	7	15	12	9	2	10	6	12	11	11	2	13	6	14	11	9	1	6
PE0	509	127	199	241	142	80	52	433	113	171	208	130	61	45	508	120	197	240	149	73	418	85	163	199	102	64	454	82	164	216	113	79	42
PE1	97	176	36	89	35	35	12	91	149	35	75	24	19	11	94	131	35	79	26	32	77	112	25	78	28	27	76	115	40	73	24	20	8
PE2	231	73	60	118	30	12	14	215	55	48	112	30	9	22	244	58	57	120	34	8	203	36	44	102	26	6	206	54	59	117	24	9	17
PE3	145	26	129	71	31	7	21	146	28	115	64	23	6	16	149	26	122	73	31	7	144	26	96	68	22	5	136	23	102	71	25	5	20
PE4	78	13	70	30	90	8	14	72	9	68	28	71	5	10	81	9	81	28	97	2	61	8	67	31	71	5	61	13	70	35	80	5	15

Se realiza el análisis factorial múltiple con la función MFA donde permite separar por cada periodo (1 al 12 mes), luego por tipo se selecciona "s" escalado es decir reducir variabilidad a 1 por columna, luego se registró el nombre de grupo, falso tabla de contingencia – cuando se tienen frecuencias.

### Resumen de los resultados de MFA

#### Valores propios del ACP global

Se muestra los valores propios del Análisis Componentes Principales (ACP) global, que se obtuvo a partir del ACP de la Matriz Total  $x = (Dijk)$  Se presenta el resultado del ACP para cada K tabla (periodo) de forma independiente, Valores propios, Porcentaje de varianza y Porcentaje Acumulado de Varianza de cada componente (dimensión)

La primera componente representa el valor el 76.2 % de la varianza acumulada y la segunda componente explican un 89.3% de la varianza acumulada, cabe indicar que se ha trabajado en los dos primeros ejes en este modelo.

**Tabla 8 Valores propios del ACP global**

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	11.830175397	76.22106759	76.22107
comp 2	2.040929327	13.14957783	89.37065
comp 3	0.848429412	5.46637673	94.83702
comp 4	0.313577778	2.02036168	96.85738
comp 5	0.217111997	1.39883879	98.25622
comp 6	0.161811911	1.04254385	99.29877
comp 7	0.043058518	0.27742329	99.57619
comp 8	0.026715582	0.17212680	99.74832
comp 9	0.012052667	0.07765457	99.82597
comp 10	0.009985290	0.06433459	99.89031
comp 11	0.005649744	0.03640094	99.92671
comp 12	0.005387649	0.03471228	99.96142
comp 13	0.003321065	0.02139741	99.98282
comp 14	0.002667054	0.01718366	100.00000

**Análisis por Separado por cada periodo en forma individual**

**Muestra los valores propios del ACP del periodo 1- mes1, con relación a los diagnostico**

- Se obtiene a partir del ACP de la Matriz Total  $x = (Dijk)$  Se presenta el resultado del ACP para cada K tabla (periodo) de forma independiente, Valores propios, Porcentaje de varianza y Porcentaje Acumulado de Varianza de cada componente (dimensión)

**Tabla 9 Valores propios del ACP del periodo 1- ene1**

```
> res.mfa2$separate.analyses$ene1$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	5.49664620	78.5235171	78.52352
comp 2	0.90293656	12.8990938	91.42261
comp 3	0.31499215	4.4998878	95.92250
comp 4	0.13019191	1.8598844	97.78238
comp 5	0.09543576	1.3633681	99.14575
comp 6	0.04847848	0.6925498	99.83830
comp 7	0.01131894	0.1616991	100.00000

Se obtiene información detallada ACP del periodo 1- mes1 como relación a los diagnósticos:

- Coordenadas
- Correlación
- Calidad de Representación cos2
- Contribución

**Tabla 10 Coordenadas-Correlación -Cos2-Contribucion del ACP del periodo 1- ene1 con relación a los diagnósticos**

```
> res.mfa2$separate.analyses$ene1$var
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T1	0.9619453	0.006542816	-0.18595449	0.11133844	-0.152518639
T2	0.6876832	0.663859005	0.24624065	0.13131385	0.088950029
T3	0.9111305	-0.346192438	-0.01452133	0.11285618	0.125845098
T4	0.9559761	0.179892354	-0.18900570	0.03471801	-0.096936141
T5	0.8024295	-0.422979547	0.40439844	-0.01980007	-0.106519846
T6	0.9158292	0.268948252	0.01254317	-0.28790784	-0.006753144
T7	0.9336563	-0.242342017	-0.14194774	-0.05768456	0.166240896

```
$coord
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T1	0.9619453	0.006542816	-0.18595449	0.11133844	-0.152518639
T2	0.6876832	0.663859005	0.24624065	0.13131385	0.088950029
T3	0.9111305	-0.346192438	-0.01452133	0.11285618	0.125845098
T4	0.9559761	0.179892354	-0.18900570	0.03471801	-0.096936141
T5	0.8024295	-0.422979547	0.40439844	-0.01980007	-0.106519846
T6	0.9158292	0.268948252	0.01254317	-0.28790784	-0.006753144
T7	0.9336563	-0.242342017	-0.14194774	-0.05768456	0.166240896

```
$cos2
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T1	0.9253387	4.280844e-05	0.0345790739	0.0123962487	2.326194e-02
T2	0.4729081	4.407088e-01	0.0606344595	0.0172433263	7.912108e-03
T3	0.8301589	1.198492e-01	0.0002108691	0.0127365184	1.583699e-02
T4	0.9138903	3.236126e-02	0.0357231542	0.0012053403	9.396615e-03
T5	0.6438931	1.789117e-01	0.1635380984	0.0003920426	1.134648e-02
T6	0.8387431	7.233316e-02	0.0001573311	0.0828909241	4.560496e-05
T7	0.8717141	5.872965e-02	0.0201491601	0.0033275088	2.763604e-02

```
$contrib
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
T1	16.834605	0.004741024	10.97775749	9.5215200	24.37444194
T2	8.603576	48.808387713	19.24951471	13.2445452	8.29050584
T3	15.103007	13.273269611	0.06694423	9.7828801	16.59439589
T4	16.626325	3.584001395	11.34096661	0.9258181	9.84601051
T5	11.714291	19.814426010	51.91815108	0.3011267	11.88912511
T6	15.259179	8.010879727	0.04994763	63.6682606	0.04778602
T7	15.859017	6.504294520	6.39671825	2.5558492	28.95773469

Además, se obtiene información detallada ACP del periodo 1- mes1 como relación al cliente – estadios como:

- Coordenadas
- Calidad de Representación cos2
- Contribución
- Distancia

**Tabla 11 Coordenadas-Correlación -Cos2-Contribucion del ACP del periodo 1- mes1 con relación a los cliente-estadios**

```
> res.mfa2$separate.analyses$enel$ind
```

\$coord	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
IE0	4.87244867	-1.19648920	0.49279948	0.09849435	0.188676620
IE1	-1.00286380	0.74328760	0.37035648	0.09302259	-0.132036285
IE2	0.05618813	-0.19488078	-0.35962836	0.57423516	-0.057886215
IE3	-0.12342526	-0.66077024	-0.09648998	0.17370133	0.458972874
IE4	-0.64227953	-1.38819640	1.18346659	-0.03009956	-0.280264464
ME0	-0.89867558	0.05551424	-0.40216762	-0.49244303	0.455596163
ME1	-2.39112075	0.21126090	-0.09781986	-0.34309906	-0.228572623
ME2	-2.11765731	0.08274230	-0.28519991	-0.28412580	-0.224602851
ME3	-1.82504677	-0.16584598	-0.42296084	-0.15244482	0.142524801
ME4	-2.18322241	-0.07248288	-0.20733888	-0.24944084	0.001097585
PE0	6.09776789	0.56238531	-0.61245466	-0.56539690	-0.283287719
PE1	0.58992162	2.64097650	1.01669698	0.10638185	0.364261624
PE2	0.30654930	0.83054224	-0.50753447	0.66722390	-0.575151431
PE3	-0.00324825	-0.57956958	-0.61125892	0.50982055	0.399240873
PE4	-0.73533594	-0.86847404	0.53953398	-0.10582971	-0.228568953

```
$cos2
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
IE0	9.307833e-01	0.056126969	0.009521262	0.0003803441	1.395694e-03
IE1	5.650287e-01	0.310384748	0.077059555	0.0048614232	9.794295e-03
IE2	4.270851e-03	0.051376354	0.174957732	0.4460718557	4.532894e-03
IE3	1.985334e-02	0.569018498	0.012133607	0.0393216494	2.745361e-01
IE4	1.079697e-01	0.504377181	0.366577335	0.0002371231	2.055842e-02
ME0	5.371904e-01	0.002049893	0.107581240	0.1612999908	1.380646e-01
ME1	9.582326e-01	0.007480079	0.001603696	0.0197290874	8.756214e-03
ME2	9.461466e-01	0.001444449	0.017161128	0.0170321077	1.064333e-02
ME3	9.168010e-01	0.007570716	0.049241060	0.0063966476	5.591238e-03
ME4	9.734809e-01	0.001073007	0.008779969	0.0127076896	2.460413e-07
PE0	9.707703e-01	0.008257387	0.009793154	0.0083460608	2.095223e-03
PE1	4.093125e-02	0.820342984	0.121576452	0.0013310717	1.560606e-02
PE2	5.137040e-02	0.377081911	0.140813192	0.2433634975	1.808326e-01
PE3	7.788576e-06	0.247953610	0.275809799	0.1918641974	1.176601e-01
PE4	3.259732e-01	0.454698969	0.175488190	0.0067518918	3.149523e-02

\$contrib					
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
IE0	2.879423e+01	10.56985593	5.1398388	0.49676092	2.486759e+00
IE1	1.219818e+00	4.07910869	2.9030126	0.44309983	1.217823e+00
IE2	3.829130e-03	0.28040743	2.7372653	16.88512595	2.340711e-01
IE3	1.847647e-02	3.22368387	0.1970487	1.54500906	1.471538e+01
IE4	5.003330e-01	14.22831029	29.6429227	0.04639221	5.486983e+00
ME0	9.795280e-01	0.02275414	3.4231286	12.41757357	1.449966e+01
ME1	6.934481e+00	0.32952605	0.2025177	6.02786905	3.649606e+00
ME2	5.439041e+00	0.05054831	1.7215030	4.13376787	3.523937e+00
ME3	4.039792e+00	0.20307730	3.7862504	1.19000858	1.418987e+00
ME4	5.781053e+00	0.03879023	0.9098515	3.18610341	8.415389e-05
PE0	4.509753e+01	2.33517501	7.9388373	16.36935617	5.606000e+00
PE1	4.220847e-01	51.49683937	21.8772175	0.57950848	9.268819e+00
PE2	1.139755e-01	5.09301497	5.4518025	22.79648749	2.310798e+01
PE3	1.279705e-05	2.48006275	7.9078685	13.30942883	1.113442e+01
PE4	6.558168e-01	5.56884565	6.1609349	0.57350859	3.649489e+00

\$dist						
	IE0	IE1	IE2	IE3	IE4	ME0
5.0503676	1.3341563	0.8597801	0.8759661	1.9546689	1.2261365	
	ME1	ME2	ME3	ME4	PE0	PE1
2.4426769	2.1770904	1.9060595	2.2127598	6.1888884	2.9158609	
	PE2	PE3	PE4			
1.3525206	1.1639126	1.2879377				

También se obtiene Tabla resumen del Promedio, desviación estándar, puntuación mínima y puntuación máxima de cada  $J_k$  variable

**Tabla 12 Tabla resumen de las variables**

```
> res.mfa2$summary.quantile
```

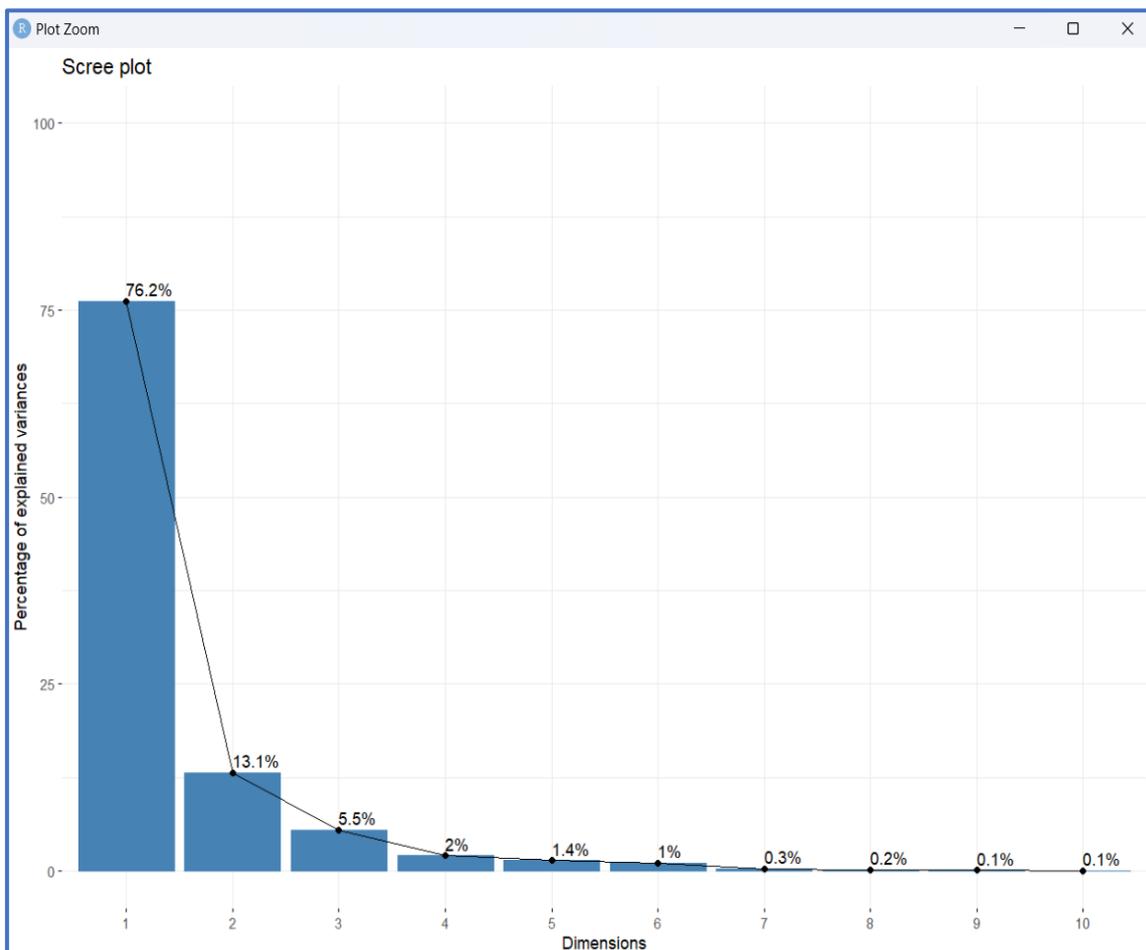
	group	variable	moyenne	ecart.type	minimum	maximum
1	1	T1	129.33	139.06	6	509
2	1	T2	47.20	49.33	4	176
3	1	T3	74.13	63.27	3	221
4	1	T4	69.73	62.29	11	241
5	1	T5	56.00	57.74	3	202
6	1	T6	16.33	20.91	1	80
7	1	T7	19.07	14.32	3	52
8	2	T1.1	121.80	125.58	3	433
9	2	T2.1	43.53	43.21	4	149
10	2	T3.1	69.53	57.37	3	211
11	2	T4.1	63.40	53.98	10	208
12	2	T5.1	51.93	54.93	4	197
13	2	T6.1	12.80	15.61	1	61
14	2	T7.1	17.93	13.57	3	52
15	3	T1.2	132.33	143.31	2	508
16	3	T2.2	44.87	42.38	4	131
17	3	T3.2	73.80	59.84	4	203
18	3	T4.2	68.00	61.68	10	240
19	3	T5.2	58.07	59.53	5	209
20	3	T6.2	14.80	19.66	1	73
21	4	T1.3	115.00	122.82	4	418
22	4	T2.3	35.13	33.92	2	112
23	4	T3.3	61.80	51.87	3	183
24	4	T4.3	58.87	51.71	8	199
25	4	T5.3	46.13	47.13	3	177
26	4	T6.3	12.67	16.92	2	64
27	5	T1.4	118.93	128.21	4	454
28	5	T2.4	38.20	34.90	3	115
29	5	T3.4	64.60	50.93	2	183
30	5	T4.4	62.13	53.77	8	216

## Gráficos Análisis Factorial Múltiple

### Cree plot

Se obtiene un Cree plot (grafica del codo) donde se visualiza los resultados de los valores obtenidos en el análisis factorial, indican la varianza explicada por cada componente principal el cual la primera componente está cerca del 76.2% y la segunda cerca del 13.1%

**Figura 16 Scree plot Análisis factorial múltiple**

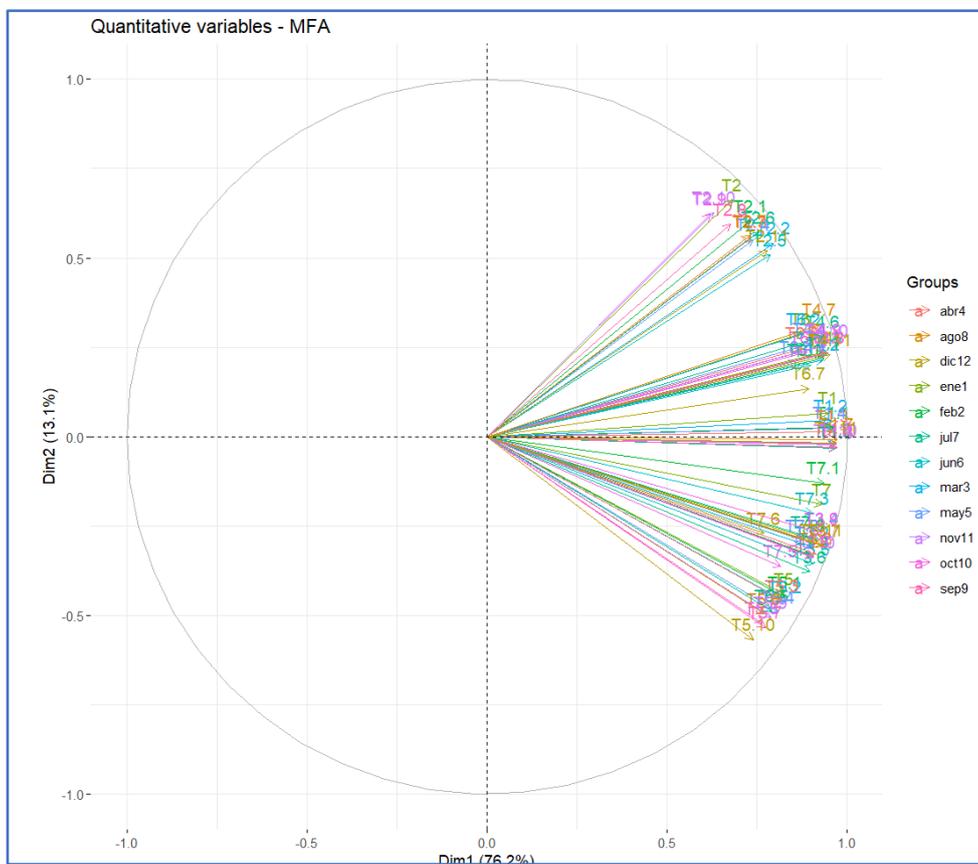


**Fuente: Propia**

## Visualiza los resultados de las variables

Esta visualización no indica cómo se distribuyen las variables en el espacio definido por las dimensiones principales del MFA. Esto nos ayudará a identificar las variables que más contribuyen a cada dimensión y a comprender mejor la estructura de tus datos en relación con los diferentes grupos o bloques especificados en el análisis MFA, las variables se encuentran al externo tienes puntuaciones altas, mientras el lejos del centro tendrá valores altos.

**Figura 17 Visualiza los resultados de las variables**



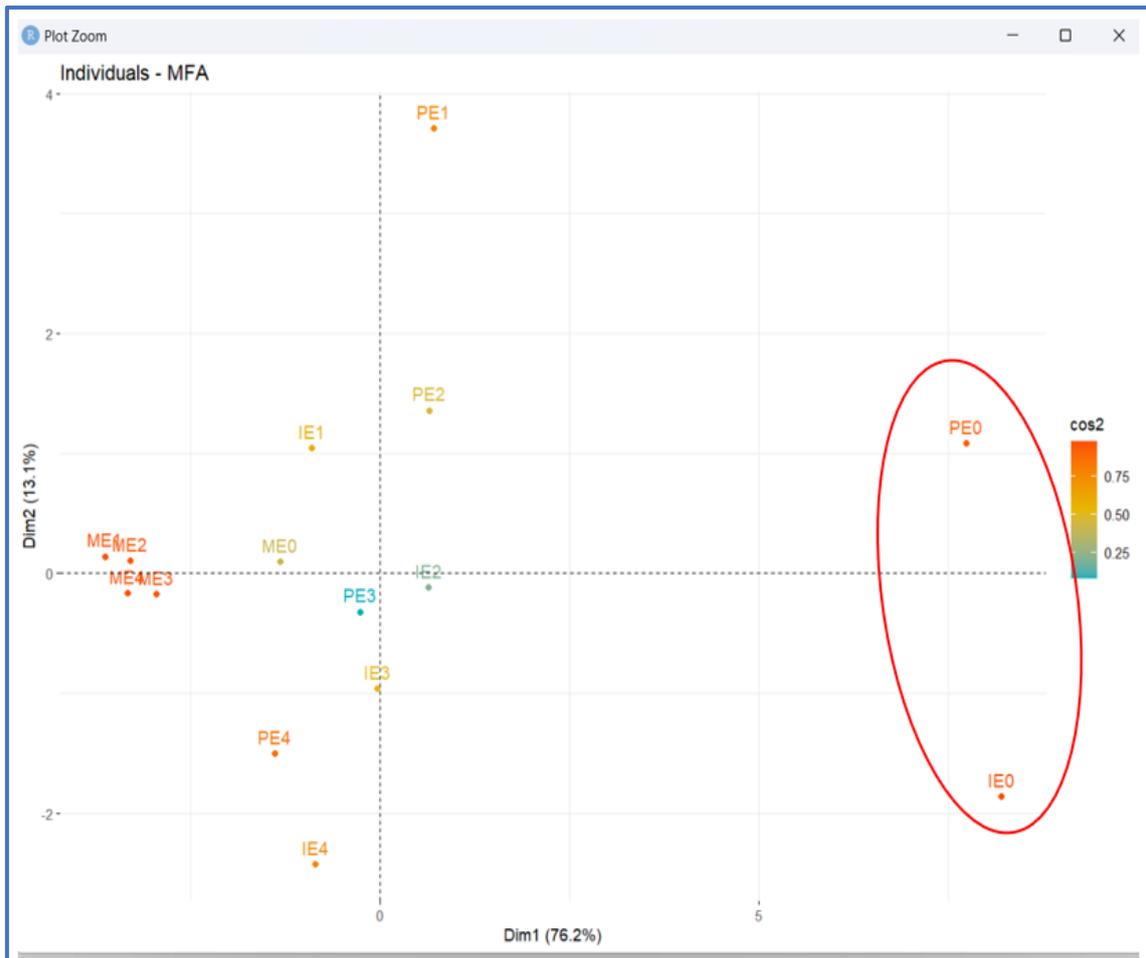
**Fuente: Propia**

## Visualiza los resultados de los individuos

Esta visualización del Análisis Factorial Múltiple (MFA) nos permite visualizar las observaciones (individuos) en el espacio de las primeras dimensiones, es útil para explorar cómo se distribuyen las observaciones en las dimensiones principales obtenidas. Permite identificar patrones en tus datos y a comprender la relación entre las observaciones y las diferentes variables o grupos especificados.

Los clientes por tipo de estadio donde se encuentra ubicados, donde IE0 y PE0 se encuentra distante del grupo de individuos. Es decir que tienes mayor cantidad de pacientes en estas variables.

**Figura 18 Visualiza los resultados de los individuos**



**Fuente: Propia**

## Longitud de los vectores

Permite visualizar las observaciones (individuos) en un gráfico tipo "star plot", donde cada punto representa una observación y los radios del gráfico indican la contribución y la calidad de representación de cada observación en las dimensiones del MFA seleccionadas.

Longitud de los vectores presentan meses que otorgaron una mayor puntuación a los individuos, es decir de los clientes por tipo de estadios, mientras mayor es la longitud del vector le dio alta calificación.

**Figura 19 Longitud de los vectores**



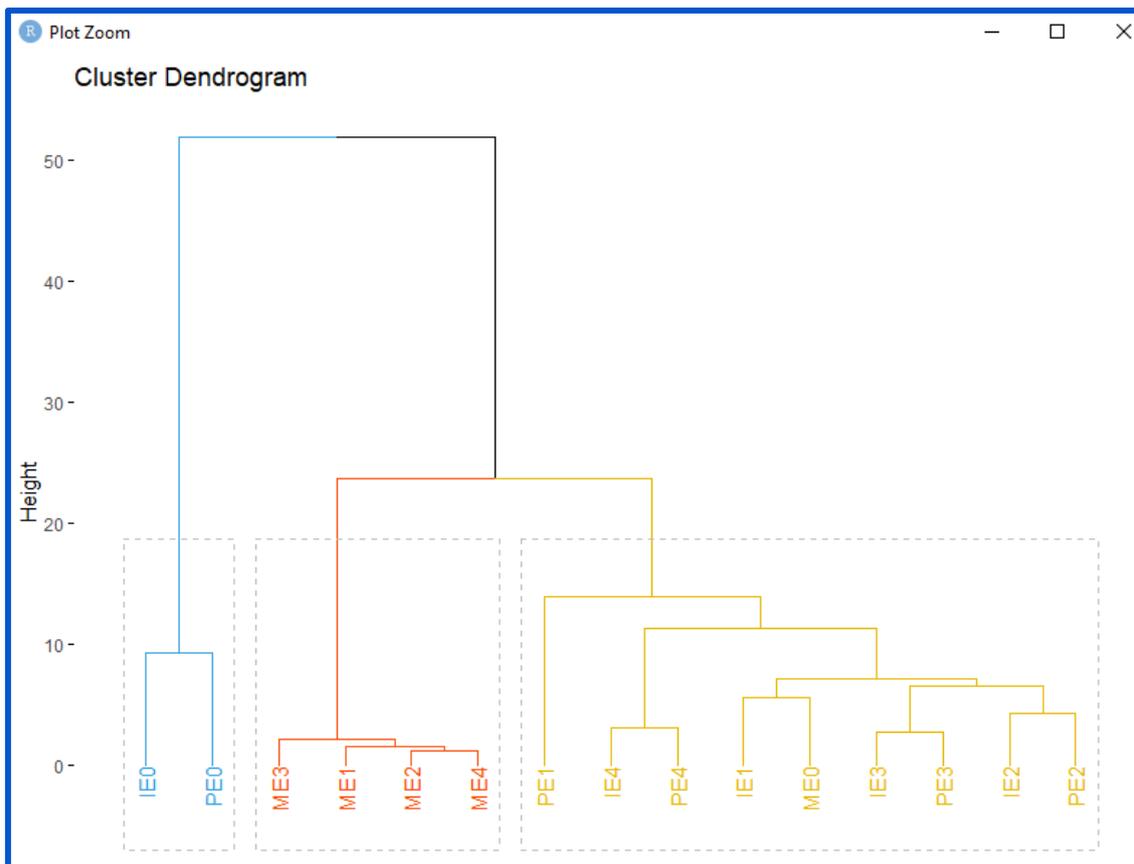
**Fuente: Propia**

## Cluster Usando el algoritmo de Ward

Se convirtió en un dataframe la data y realizando una estandarización de las variables se procede a realiza un cluster de agrupamiento jerárquico utilizando el algoritmo de Ward y luego visualiza el dendrograma resultante con los datos divididos en cuatro grupos. Los colores y otros aspectos visuales del dendrograma se controlan mediante los parámetros proporcionados a la función.

Donde se visualizan 3 agrupaciones tales como los clientes IE0 y PE0 reflejan el primer grupo de individuos. Es decir que tienen gran cantidad de pacientes en estas variables

**Figura 20 Clúster agrupamiento jerárquico**



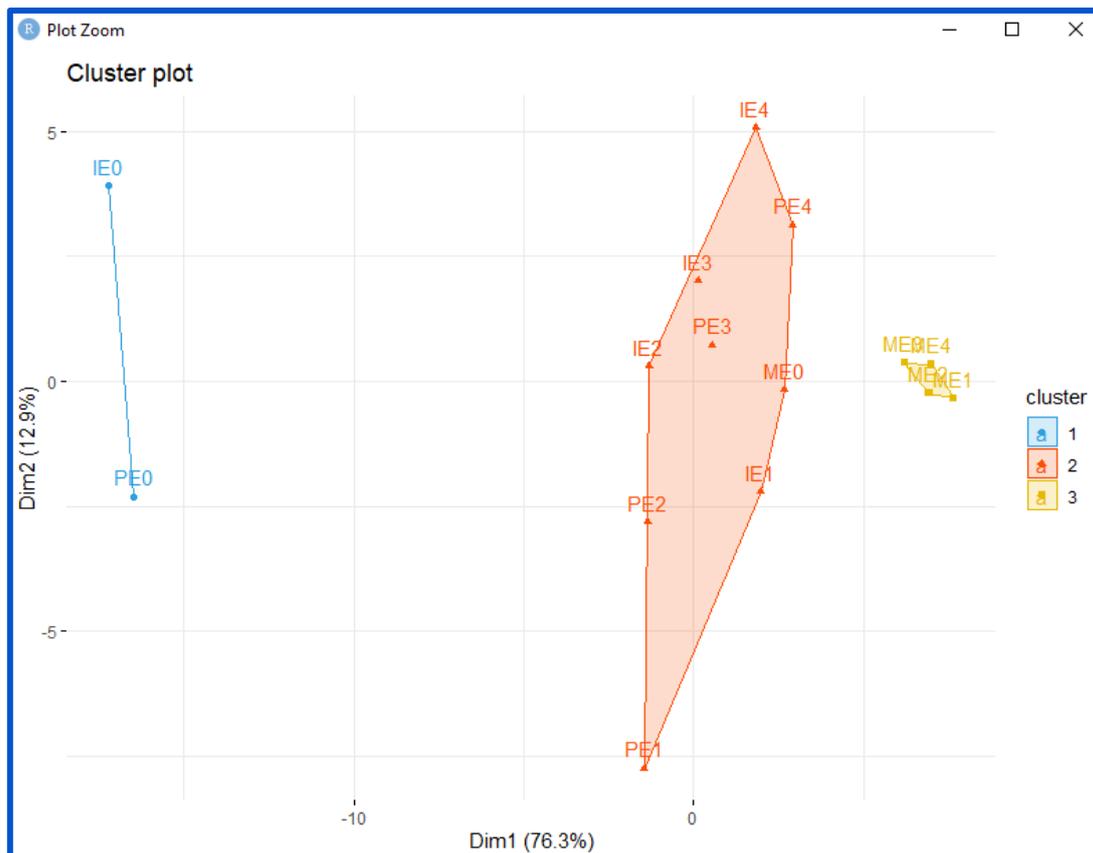
**Fuente: Propia**

## Cluster nube de puntos

Utilizando otro código en R la Clusterización con la nube de puntos se corta el dendrograma en tres clústeres y visualiza los datos con los clústeres resaltados, utilizando colores específicos para cada clúster y representando la concentración de datos en cada clúster mediante elipses.

Se visualizan 3 agrupaciones tales como los clientes IE0 y PE0 reflejan el primer grupo de individuos de color celeste, segundo grupo con otras variables de color roja y el tercer grupo con las variables de color naranja.

**Figura 21 Clúster nube de puntos**



**Fuente: Propia**

# CAPÍTULO 5

## 5. CONCLUSIONES Y RECOMENDACIONES

### 5.1 Conclusiones Base de datos

- Previa gestión e integración de los matrices se obtuvo una data estructurada que permitió realizar el análisis multivariante cumpliendo con uno de los objetivos propuestos.
- Se imputaron datos faltantes en la data, realizando mediante la técnica de imputación de la mediana, necesario para el Método Statis.

### 5.2 Análisis Método Statis

- Se consiguió el análisis de los pacientes de cáncer relacionada con los tipos de estadios con la tendencia en relación con los diferentes periodos evaluados, con gráficos se pudo interpretar y tener un análisis adecuado.
- Por cada periodo K- Tabla se proyecta en el primer eje de las coordenadas, obteniendo el valor máximo 1 donde retiene importante información, por lo tanto, el Statis es idóneo para este análisis.
- Se logró una eficacia para integrar y comparar múltiples tablas de datos, logrando patrones y estructuras permitiendo analizar información relevante de los diferentes Estadios de Cáncer.
- Nos ofrece una visión clara de la tendencia de los pacientes derivados del sector de la salud, que ayudará en la labor de prevención y detección precoz del cáncer.

### 5.3 Análisis Factores Múltiples

- Se obtuvo un análisis integrado de múltiples tablas de datos como cliente, estadio y periodos, permitiendo identificar estructuras comunes y específicas, teniendo una visualización e interpretación de los datos.
- La Dimensión 1 (eje1), tiene una mayor variabilidad en los meses donde explica la información, por eso que con la data de Statis se puede realizar para un AFM por recoger la información en el primer eje.

- La elaboración de gráficos de Clúster permite una clara visualización de las agrupaciones de los diferentes Cliente y Estadios.
- Nos facilitó la visualización e interpretación de los resultados, proporcionando una visión integrada y coherente de los datos.

#### **5.4 Conclusión final**

Este trabajo permite determinar en qué proporción están derivando los pacientes del sistema de salud por las diferentes patologías y estadio de cáncer, para mejorar la gestión administrativa y financiera y con ello brindar una atención de calidad y calidez.

#### **5.5 Recomendaciones**

- Los siguientes desafíos es necesario conseguir métodos con nuevas herramientas tecnológicas, con el propósito de ayudar la implementación de proyectos.
- Probar varias técnicas estadísticas y comparar sus resultados para seleccionar la que mejor se adapte de acuerdo la base de datos de un periodo actualizado en vista que la población incrementa constantemente, entre la técnica multivariante HJ-Biplot, Análisis de Correlaciones Canónicas (CCA)

# CAPÍTULO 6

## 6. REFERENCIAS

### Bibliografía

Abdi, H. (2012). STATIS and DISTATIS: optimum multitable principal component analysis and three-way metric multidimensional scaling. The University of Texas at Dallas, Richardson, TX, USA, 12-34.

Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(2), 149-179. FactoMineR: An R Package for Multivariate Analysis F. Husson, S. Le y J. 2008. *Multivariate Exploratory Data Analysis and Data Mining* Francois Husson, Julie Josse, Sebastien Le, Jeremy Mazet 2009

Aguirre-Loaiza, H. H., Núñez, C., Navarro, A. M., & Sánchez, S. C. (2017). Calidad de vida según el estadio del cáncer de seno en mujeres: análisis desde el Fact-B y SF-36. *Psychologia. Avances de la disciplina*, 11(1), 109-120.

Aranda Aguilar, E. (2004). Tratamiento del cáncer de colon estadios II, III y IV. *Oncología (Barcelona)*, 27(4), 130-133.

Becerra Avella, M. O. (2010). Comparación del análisis factorial múltiple (AFM) y del análisis en componentes principales para datos cualitativos (Prinqual), en la construcción de índices (Doctoral dissertation).

CUADRAS, C.M. (1996). *Métodos de Análisis Multivariante*, EUB, Barcelona

ESCOFIER, B. & PAGÈS, J. 1992. *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación. (Trad.)*. Servicio Editorial Universidad del País Vasco.

- Gabriel, K.R. (1971). "The biplot-graphic display of matrices with application to principal component analysis." *Biometrika*, 58: 453-467.
- García, Y. P., Ramos, N. T., Torres, Y. D. L. A. H., & Miclin, O. T. (2022) Modelo predictivo del riesgo de padecer cáncer cervicouterino en la población femenina de Puerto Padre.
- González, J., & Rodríguez, O. (1995). Algoritmo e implementación del método Statis. In IX Simposio Métodos Matemáticos Aplicados a las Ciencias. UCR-ITCR, Turrialba.
- Gutiérrez, C., & Alarcón, E. (2008, December). Nivel de pobreza asociado al estadio de gravedad del cáncer ginecológico. In *Anales de la Facultad de Medicina* (Vol. 69, No. 4, pp. 239-243). UNMSM. Facultad de Medicina.
- HAIR, J.F., ANDERSON, R.E., TATHAM, R.L. and BLACK, W.C. (1998). *Multivariate Data Analysis*, Prentice Hall, New Jersey
- Herrera, J. M. (2007). El Análisis Factorial Múltiple: Un Paso más en la Superación de la Dicotomía Cualitativo-Cuantitativo. In VI Congreso Chileno de Antropología. Colegio de Antropólogos de Chile AG.
- Hidalgo, M., & VILLARDÓN, D. D. J. L. V. (2011). HJ-Biplot aumentado. Departamento de Estadística, Universidad de Salamanca.
- Kutner, M., Nachtsheim, C., & Neter, J. (2004). *Applied Linear Regression Models*. McGraw-Hill Education.
- Martín-Payo, R., Ferreras-Losilla, L., González Méndez, X., Leirós-Díaz, C., Martínez-Urquijo, A., & del Mar Fernández-Álvarez, M. (2021). Apps for individuals diagnosed with breast cancer: a preliminary assessment of the content and quality of commercially available apps in Spanish. *Mhealth*, 7.

N del Castillo Predicción en el diagnóstico de tumores de cáncer de mama empleando métodos de clasificación (2020)  
de:<http://portal.amelica.org/ameli/jatsRepo/368/3681483009/index.html>

Núñez, C., & Sánchez, S. C. (2017). calidad de Vida según El Estadio del Cáncer de seno en mujeres: análisis desde El Fact-B y sF-361. *Psychologia. Avances de la Disciplina*.

Peña, D. (2002). Análisis de datos multivariantes.

Sánchez, C., Maldonado, D., Jans, J., Domínguez, F., Galindo, H., Camus, M., ... & Acevedo, F. (2018). Aplicación de tres modelos pronósticos en cáncer de mama precoz. *Revista chilena de cirugía*, 70(3), 212-217.

Santana, E. (19 de abril de 2015). Apuntes R. Obtenido de Machine Learning con R: <http://apuntes-r.blogspot.com/2015/04/supuestos-en-regresionlineal.htm>

Seidl EM, Zannon CM. Calidad de vida y salud: aspectos conceituais e metodológicos. *Cad Saude Publica*. marzo a abril de 2004; 20(2):580-8. Portugués. doi: 10.1590/s0102-311x2004000200027. Epub 6 de abril de 2004. PMID: 15073639.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias*.

Wikipedia. (14 de diciembre de 2020b). Wikipedia. Obtenido de Bayesian information criterion: [https://en.wikipedia.org/wiki/Bayesian\\_information\\_criterion](https://en.wikipedia.org/wiki/Bayesian_information_criterion)

Wikipedia. (2020a). Wikipedia. Obtenido de Análisis de componentes principales: [https://es.wikipedia.org/wiki/An%C3%A1lisis\\_de\\_componentes\\_principales](https://es.wikipedia.org/wiki/An%C3%A1lisis_de_componentes_principales)

# CAPÍTULO 7

## 7. APENDICE Y ANEXOS

### CODIGO R

```
library(ade4)
```

```
library(openxlsx)
```

```
library(esquisse)
```

```
library(ggplot2)
```

```
library("factoextra")
```

```
library("FactoClass")
```

```
library("FactoMineR")
```

```
library("mosaic")
```

```
library(openxlsx)
```

```
library(readxl)
```

```
estadio=read.xlsx("C:/Users/hp/OneDrive - SOCIEDAD DE  
LUCHA CONTRA EL CANCER DEL ECUADOR  
SOLCA/Escritorio/MAESTRIA/PROYECTO TITULACION/DATOS  
TESIS/TESIS.xlsx",sheet="TESIS")
```

```
summary(estadio[,3:17])
```

```
ggplot(estadio) +
```

```
aes(x = IE0, y = DIAGNOSTICO, fill = DIAGNOSTICO) +  
geom_boxplot() +  
scale_fill_hue(direction = 1) +  
theme_minimal()
```

```
ggplot(estadio) +  
  aes(x = mes, y = IE0, fill = IE0) +  
  geom_boxplot(fill="lightblue") +  
  scale_fill_hue(direction = 1) +  
  theme_minimal()
```

```
pca_estadio <- prcomp(estadio[,3:17],  
                      scale = TRUE)
```

```
bipl1 <- fviz_pca_biplot(pca_estadio,  
                        axes = c(1, 2),  
                        label="var",  
                        habillage = estadio$mes)
```

```
bipl1
```

```

#####STATIS

kta1 <- ktab.within(withinpca(estadio[,3:17], factor(estadio$mes),
                        scaling="total",scannf = FALSE))

statis1 <- statis(kta1, scann = FALSE)

plot(statis1)

statis1$RV#coeficientes de correlación vectorial RV

statis1$cos2#coseno2

statis1$RV.tabw#pesos para las k-tablas

statis1$RV.eig#valores propios para la inter-estructura

statis1$C.eig#valores propios para el compromiso

kplot(statis1)

# METODO AFM

#Ejemplo ABDI

dato=read.xlsx("C:/Users/hp/OneDrive - SOCIEDAD DE LUCHA
CONTRA EL CANCER DEL ECUADOR
SOLCA/Escritorio/MAESTRIA/PROYECTO TITULACION/DATOS
TESIS/TESIS.xlsx",sheet="TESIS1")

dato2=dato[,2:75]

row.names(dato2)=dato$CLIENTESTADIO

```

```

head(dato2)

res.mfa2 <-
MFA(dato2,group=c(7,7,6,6,7,6,6,5,5,7,5,7),type=c("s","s","s","s",
"s","s","s","s","s","s","s","s"),name.group=c("ene1","feb2","mar3","
abr4","may5","jun6","jul7","ago8","sep9","oct10","nov11","dic12"),
graph = FALSE)

res.mfa2

summary(res.mfa2) # Resumen de los resultados de MFA

print(res.mfa2) # Información detallada sobre los resultados de
MFA

res.mfa2$eig ## Muestra los valores propios del ACP global

res.mfa2$separate.analysises

res.mfa2$separate.analysises$ene1$eig
res.mfa2$separate.analysises$ene1$var
res.mfa2$separate.analysises$ene1$ind
res.mfa2$separate.analysises$ene1$svd
res.mfa2$separate.analysises$ene1$call

res.mfa2$group

res.mfa2$inertia.ratio

res.mfa2$ind

res.mfa2$summary.quantit

res.mfa2$quantit.var

res.mfa2$partial.axes

res.mfa2$call

```

```
res.mfa2$global.pca$eig
```

```
fviz_eig(res.mfa2,addlabels = TRUE, ylim = c(0, 100))
```

```
fviz_mfa_group(res.mfa2)
```

```
fviz_mfa_var(res.mfa2) #visualiza los resultados de las variables
```

```
fviz_mfa_var(res.mfa2,geom = c("point", "text"))
```

```
fviz_mfa_ind(res.mfa2,col.ind = "cos2",
```

```
      gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07") ,
```

```
      repel = FALSE, select.var = list(cos2 = 0.6)) #visualiza los  
resultados de los individuos
```

```
fviz_mfa_ind_starplot(res.mfa2)
```

```
fviz_mfa_ind(res.mfa2, partial = "All")
```

```
fviz_mfa_axes(res.mfa2)
```

```
datos<-dato
```

```
#Convertir a Dataframe
```

```
d<-as.data.frame(datos)
```

```
#Cambiar el nombre a las filas
```

```
row.names(d)=d[,1]
```

```

#Escalado o Estandarizacion de las variables

datosC = d[,c(-1)]

#summary(datosC)

datosC = scale(datosC)

#Cluster Usando el algoritmo de Ward

res.hcw <- hclust(dist(datosC),method = "ward.D")

fviz_dend(res.hcw, k = 3, # Cut in four groups
          cex = 0.8, # label size
          k_colors = c("#2E9FDF", "#FC4E07", "#E7B800",
                      "#89C86E", "#8835DC"),
          color_labels_by_k = TRUE, # color labels by groups
          rect = TRUE # Add rectangle around groups
        )

#Graficando los cluster en la nube de puntos

grp <- cutree(res.hcw, k = 3)

fviz_cluster(list(data = datosC, cluster = grp, cex = 0.8),
             palette = c("#2E9FDF", "#FC4E07", "#E7B800"),
             ellipse.type = "convex", # Concentration ellipse
             repel = FALSE, # Avoid label overplotting (slow)
             show.clust.cent = FALSE, ggtheme = theme_minimal())

```