

**ANALISIS MULTIVARIANTE A LA MORTALIDAD
CAUSADA POR TUMORES MALIGNOS EN
ECUADOR PERÍODO 2005-2014**

Dra. Vilma Raffo Babici, Mgtr.
Compilador
PhD. Juan Tarquino Calderon Cisneros, Mgtr.
Dra. Vilma Raffo Babici, Mgtr.
PhD. José Luis Vicente-Villardón

**ANALISIS MULTIVARIANTE A LA MORTALIDAD
CAUSADA POR TUMORES MALIGNOS EN
ECUADOR PERÍODO 2005-2014**

© Dra. Vilma Raffo Babici, Mgtr.
Compilador
PhD. Juan Tarquino Calderon Cisneros, Mgtr.
Dra. Vilma Raffo Babici, Mgtr.
PhD. José Luis Vicente-Villardón

Título del libro

ANÁLISIS MULTIVARIANTE A LA MORTALIDAD CAUSADA POR
TUMORES MALIGNOS EN
ECUADOR PERÍODO 2005-2014

ISBN: 978-9942-33-571-5

Publicado 2022 por acuerdo con los autores.
© 2022, Editorial Grupo Compás
Guayaquil-Ecuador

Grupo Compás apoya la protección del copyright, cada uno de sus textos han sido sometido a un proceso de evaluación por pares externos con base en la normativa del editorial.

El copyright estimula la creatividad, defiende la diversidad en el ámbito de las ideas y el conocimiento, promueve la libre expresión y favorece una cultura viva. Quedan rigurosamente prohibidas, bajo las sanciones en las leyes, la producción o almacenamiento total o parcial de la presente publicación, incluyendo el diseño de la portada, así como la transmisión de la misma por cualquiera de sus medios, tanto si es electrónico, como químico, mecánico, óptico, de grabación o bien de fotocopia, sin la autorización de los titulares del copyright.

   @grupocompas.ec
compasacademico@icloud.com


Grupo de capacitación e investigación pedagógica


Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura


Cátedra UNESCO «Paz, Solidaridad y Diálogo Intercultural», Universitat Abat Oliba CEU

 Dialnet


RED HEMABIGDAT


Red Internacional Grupos de Investigación
Capítulo Ecuador: No. SENESCIT-2016-040REG-RED-18-0073


Grupo de Investigación Asociado Promoción de Espacios Científicos Editoriales de Acceso libre


INSTITUTO SUPERIOR UNIVERSITARIO DE FORMACIÓN


Ciesac
Centro de Investigaciones en Educación
Milton Rosenberg y Cecilia

Dedicatoria

A **DIOS** por todas sus bendiciones

A **Luis Tarquino Calderón Inca (+)** y **Olga Cisneros Cordero**, mis padres, que por sus enseñanzas y sabidurías impartidas estoy cada día logrando mis sueños.

A mis tesoros que son la razón de que cada día tengo que salir adelante por ellos y para ellos.

A **Luis Calderón Cisneros**, mi Hermano, por su lucha constante y superación diaria.

A mi gran hermano de crianza y compadre **Carlos Esteban Alcívar Trejo**. Por su apoyo, enseñanzas y su aguante en este proceso académico.

Tabla de contenido

Capítulo 1.....	7
EL Cáncer en cifras	9
Mortalidad	11
Neoplasias malignas como causa de mortalidad	12
Importancia económica y social para el sistema nacional de salud.....	13
DATOS DE MORTALIDAD DEL ECUADOR	15
MÉTODOS ESTADÍSTICOS APLICADOS EN ESTUDIOS DE MORTALIDAD	17
Capítulo 2	27
Capítulo 3	36
Pre-procesamiento de los datos	39
Depuración de la Base de datos.....	39
Categorización.	39
Etiquetado de variables.	40
Matriz de datos PARA ACM	44
Matriz de datos para AGP.....	47
Capítulo 4	52
análisis de correspondencias múltiple (acm)	52
Análisis Procrustes Generalizado (APG)	56
DISCUSIONES	63
Bibliografía.....	68

Capítulo 1

La mortalidad como concepto expresa la acción de muerte, es una de las componentes demográficas básicas, determinantes del tamaño, de la composición por género y edad de una población determinada. La Sociedad Americana del Cáncer (2014), señala al cáncer, la causa más importante de mortalidad en el mundo. Con un número de muertes que superan las de la enfermedad isquémica del corazón u otras enfermedades específicas. El riesgo de desarrollar cáncer antes de los 75 años es más aproximadamente 1 de cada 3 personas, alrededor del 35%. En 2012 se estima que hubo alrededor de 14,1 millones de nuevos casos de cáncer, con una prevalencia de más de 8 millones de muertes, siendo las regiones del mundo menos desarrolladas las más afectadas. El 50% de los casos nuevos y muertes por cáncer se presentan en Asia, y una cuarta parte de la carga mundial se produce en China. Para el año 2025, se estiman alrededor de 19,3 millones de nuevos casos. La mortalidad por cáncer es objeto de gran interés, lo que ha generado una gran cantidad de investigación sobre el tema. Igualmente, en las agendas públicas y en las poblaciones, la mortalidad por todo tipo de cáncer son temas de gran preocupación, que requiere atención inmediata a nivel mundial y regional (PAHO, 2008). En el Ecuador para el año 2014 se registraron 9670 de 62.981 muertes por cáncer que representan el 15,35% del total nacional, ocupando el primer lugar entre las principales causas de mortalidad. Las enfermedades isquémicas del corazón y la diabetes mellitus, representan la segunda y tercera causas de mortalidad general con un porcentaje de 7,03% y 6,99%, respectivamente (INEC, 2014). El Ecuador durante los últimos diez años ha tenido una transformación, en varios puntos estratégicos y vulnerables dentro de la

sociedad, el gobierno le ha dado prioridad a manejo de la información, mediante decretos ejecutivos y la creación de nuevas instituciones, con el objetivo principal de ofrecer un marco sociopolítico para vincular la dinámica demográfica, los procesos poblacionales y la formulación de políticas públicas (Plan del Buen Vivir, 2016). El presente trabajo estudia la mortalidad causada por las principales neoplasias malignas en las regiones del Ecuador para el periodo 2005-2014, utilizando la información de las bases de datos públicas del Instituto Nacional de Estadísticas y Censos. Los resultados se analizan utilizando Análisis de Correspondencia Múltiple (ACM) y el Análisis de Procrustes Generalizado (APG) y son contrastados con los descritos en la literatura mundial de acuerdo a la Clasificación de la OMS. El estudio de la mortalidad se basa en la observación de las defunciones que ocurren en una población durante un tiempo determinado y la evolución de las mismas. La baja originada en la mortalidad por medio de políticas públicas en el área de salud, las mejoras en las condiciones de vida, a pesar que la mortalidad es un hecho inevitable, su comportamiento presenta diferencias muy importantes entre países, regiones, clases sociales. (IARC, 2014)

Los datos de mortalidad constituyen uno de los elementos fundamentales para cuantificar los problemas de salud. A partir de ellos se han desarrollado numerosos métodos estadísticos para el análisis de este fenómeno (PAHO, 2012).

EL Cáncer en cifras

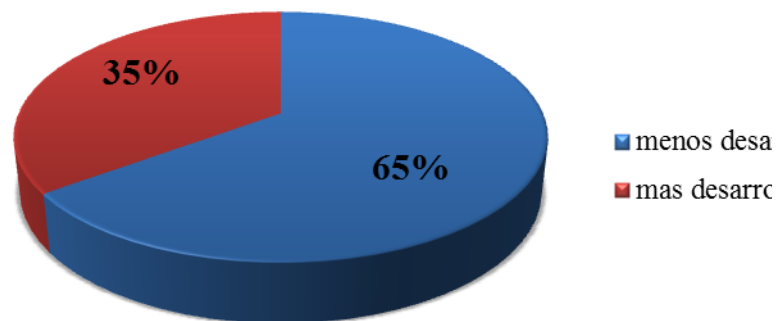
La Organización Mundial de la Salud (OMS, 2014), indican que el cáncer es una de las principales causas de morbilidad y mortalidad en todo el mundo; en el 2012 hubo unos 14 millones de nuevos casos y 8,2 millones de muertes relacionadas con el cáncer (IARC, 2014). Se prevé que el número de nuevos casos aumente en aproximadamente un 70% en los próximos 20 años. Para el 2012, los cánceres diagnosticados con más frecuencia en el hombre fueron los de pulmón, próstata, colon y recto, estómago e hígado. Indican que en la mujer fueron los de mama, colon y recto, pulmón, cuello uterino y estómago.

El cáncer es ya la principal causa de muerte en muchos países de altos ingresos, y se convertirá en una de las principales causas de morbilidad y mortalidad en las próximas décadas en todas las regiones del mundo, independientemente de sus niveles de recursos. (Bray *et al*, 2012).

Aproximadamente un 30% de las muertes por cáncer corresponden a cinco factores de riesgo conductuales y dietéticos: índice de masa corporal elevado, ingesta reducida de frutas y verduras, falta de actividad física, consumo de tabaco y consumo de alcohol. El consumo de tabaco es el factor de riesgo más importante, y es la causa más del 20% de las muertes mundiales por cáncer en general, y alrededor del 70% de las muertes mundiales son producidos por cáncer de pulmón. Los cánceres causados por infecciones víricas, tales como las infecciones por virus de las hepatitis B (VHB) y C (VHC) o por papiloma virus humanos (PVH), son responsables de hasta un 20% de las muertes por cáncer en los países de ingresos bajos y medios. Más del 60% de los

nuevos casos anuales totales del mundo se producen en África, Asia, América Central y Sudamérica. Estas regiones representan el 70% de las muertes por cáncer en el mundo, la Figura 1, muestra la mortalidad estimada a las regiones más desarrolladas y menos desarrolladas. (OMS, 2014).

Figura 1. Mortalidad estimada.



Fuente: Globocan, 2012

Se estima que nuevos casos de cáncer en el mundo aumentará de más de 14 millones de casos a casi 22 millones para el 2030, por al crecimiento demográfico y al envejecimiento exclusivamente. Los mayores incrementos del 70% se verán en África, Asia y América Latina, que incluyen a muchos países que carecen de recursos suficientes para hacer frente al creciente número de pacientes con cáncer. (Registro de Tumores-SOLCA Matriz Guayaquil, 2016).

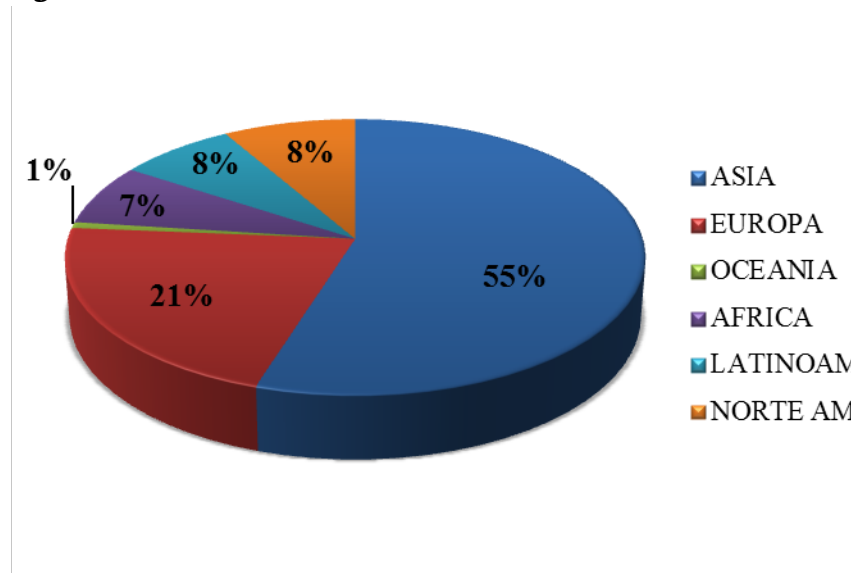
En 2011 casi 8 millones de muertes se produjeron por cáncer, superando a todas las otras causas específicas de muerte. Las 10 causas principales de muerte en el mundo para el 2011 son: neoplasias malignas, cardiopatía isquémica, enfermedad cerebro-vascular (derrames), infecciones respiratorias inferiores, enfermedad pulmonar obstructiva

crónica, enfermedades diarreicas, VIH/SIDA, diabetes mellitus, lesiones por accidentes de tránsito y complicaciones en el parto prematuro (Sociedad Americana del Cáncer, 2014).

Mortalidad

Una definición dada por la Organización mundial de la salud establece que según los datos de mortalidad indican el número de defunciones por lugar, intervalo de tiempo y causa. Los datos de mortalidad de la OMS reflejan las defunciones recogidas en los sistemas nacionales de registro civil, con las causas básicas de defunción codificadas por las autoridades nacionales, en la Figura 2, se muestra el porcentaje de mortalidad estimada por cáncer en cada continente. (OMS, 2014).

Figura 2. Mortalidad mundial estimada de Cáncer



Fuente: Globocan, 2012

Existen diferencias socio-geográficas en la mortalidad y la supervivencia del cáncer. Las tasas relativamente altas de cáncer son las de hígado, estómago y cuello uterino en algunos países de Asia, América del Sur y África. Estas poblaciones se distinguen además por su gran carga de cáncer de próstata y cáncer de pulmón, así como la notable alta tasa de melanoma de piel. La variación en la ocurrencia de diferentes tipos de cánceres dentro de las diferentes zonas del mundo da una indicación de la proporción de cánceres que podrían evitarse modificando estilos de vida nocivos específicos o factores medioambientales. (Sociedad Americana del Cáncer, 2014).

Neoplasias malignas como causa de mortalidad

La Organización Mundial de la Salud señala que el cáncer es un proceso de crecimiento y diseminación descontrolados de células, alteración de la proliferación y de la diferenciación celular anormal, que se manifiesta por la formación de una masa. Este crecimiento anormal de tejido, es producido por la alteración de los mecanismos que controlan la multiplicación celular en el organismo (IARC-OMS, 2016). Pueden aparecer prácticamente en cualquier lugar del cuerpo, suelen invadir el tejido circundante, provocan metástasis en puntos distantes del organismo, y, además, una vez originados, continúan creciendo, aunque dejen de actuar las causas que los provocan. Muchos tipos de cánceres se pueden prevenir evitando la exposición a factores de riesgo comunes como el humo de tabaco, malos hábitos alimentarios, exceso en consumo de alcohol, exposición a radiación ultravioleta, entre otros. Además, un porcentaje

importante de cánceres pueden curarse mediante cirugía, radioterapia o quimioterapia, especialmente si se detectan en una fase temprana. A nivel mundial, 1 de cada 5 hombres y 1 de cada 6 mujeres serán diagnosticados con cáncer antes de la edad de 75 años. Uno de cada 8 hombres y 1 de cada 12 mujeres morirán de la enfermedad (OMS, 2016).

Importancia económica y social para el sistema nacional de salud

El Estado, debe asegurar y fortalecer la seguridad social y le corresponde a éste, organizar, dirigir y reglamentar la prestación de servicios de salud a los habitantes, no sólo desde el diagnóstico y tratamiento de las enfermedades, sino desde la implementación de políticas de largo alcance hacia el fomento de la salud que beneficie a amplios grupos poblacionales, con políticas públicas saludables, intervenciones dirigidas a personas y comunidades. (Martínez, 2016).

Las funciones del sistema de salud en el país según el artículo 32 de la Constitución Política de la República del Ecuador 2008 consagra a la salud como un derecho garantizado por el Estado mediante políticas económicas, sociales, culturales, educativas y ambientales; y el acceso permanente, oportuno y sin exclusión a programas, acciones y servicios de promoción y atención integral de salud, salud sexual y salud reproductiva. La prestación de los servicios de salud se regirá por los principios de equidad, universalidad, solidaridad, interculturalidad, calidad, eficiencia, eficacia, precaución y bioética, con enfoque de género y generacional. Además, el artículo 358 menciona que “el sistema nacional de salud tendrá por finalidad el desarrollo, protección y recuperación

de las capacidades y potencialidades para una vida saludable e integral”; el artículo 359 enuncia que “El sistema nacional de salud comprenderá las instituciones, programas, políticas, recursos, acciones y actores en salud...” y, en el artículo 360 que “el sistema garantizará, a través de las instituciones que lo conforman, la promoción, prevención y atención integral, familiar y comunitaria, con base en la atención primaria de salud (Constitución de la República del Ecuador, 2008.).

Las políticas del estado ecuatoriano giran en torno al Plan Nacional del Buen Vivir y dentro de sus objetivos en referente a salud el objetivo 2, propone auspiciar la igualdad, la cohesión, la inclusión y la equidad social y territorial, en la diversidad, dentro de sus políticas, con el cual se pretende garantizar la igualdad real con el acceso a servicios de salud y educación de calidad a personas y grupos que requieren especial consideración, por la persistencia de desigualdades, exclusión y discriminación. Para lograr estos objetivos se han implementado mecanismos y procesos en los servicios de salud pública, que garanticen la gratuidad dentro de la red pública integral de salud en todo el territorio nacional, con base en la capacidad de acogida de los territorios y la densidad poblacional y la generación e implementación de mecanismos y acciones afirmativas que además de garantizar la gratuidad y eliminen las barreras de acceso de los servicios de salud, con énfasis en el cierre de brechas de desigualdad (INEC, 2016).

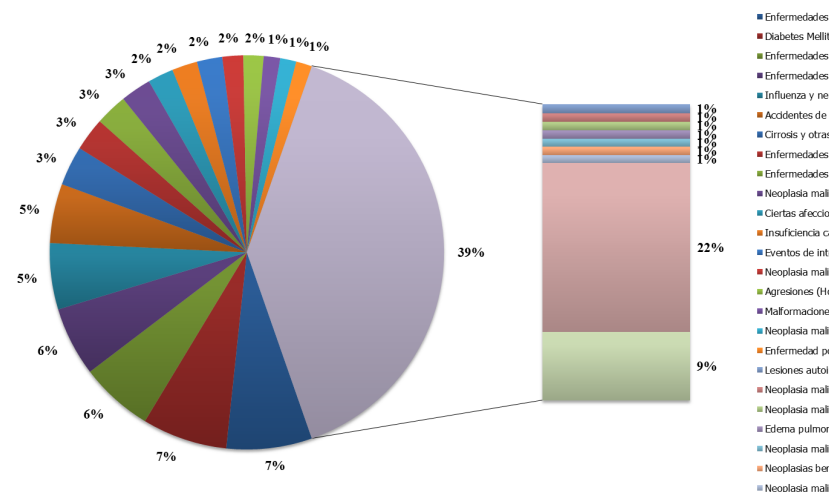
PNBV, además propone mejorar la calidad de vida de la población a través de la ampliación de los servicios de prevención y promoción de la salud, diseñando mecanismos integrales de promoción de la salud para prevenir riesgos durante todo el ciclo de vida, con énfasis sobre los determinantes sociales de salud. Hace énfasis en el

conocimiento del perfil epidemiológico y sanitario del país, como principal herramienta para la planificación de la oferta de servicios de promoción y prevención, y el fortalecimiento sistema de vigilancia y control epidemiológico, con corresponsabilidad comunitaria, ante posibles riesgos que causen morbilidad y mortalidad evitable o que sean de notificación obligatoria. Se destaca el papel que se le da a las tecnologías de la información, como herramienta de gestión en la consolidación del registro sanitario único que favorezca la respuesta inmediata y fortalezca la red pública integral de salud de manera coordinada e integrada, para optimizar el uso de recursos (Secretaría Nacional de Planificación y Desarrollo, 2013).

DATOS DE MORTALIDAD DEL ECUADOR

Según el Banco Mundial para Ecuador desde 1960 a 2013, la mortalidad media fue de 8.29 muertes por cada 1.000 personas, con un mínimo de 4.91 en 2013 y un máximo de 15.62 en 1960. (OMS, 2014), en la Figura 3, se muestra las 25 principales causas de mortalidad general del Año 2014 que el INEC reportó en su boletín oficial de ese año.

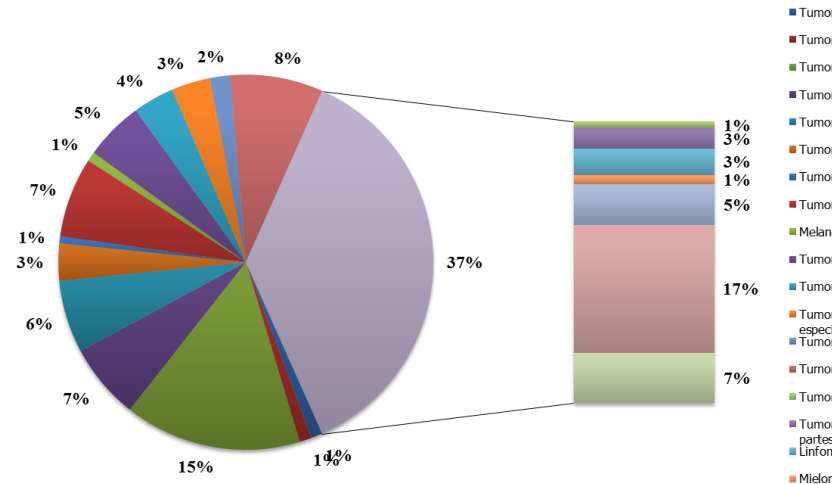
Figura 3. Principales causas de mortalidad general Año 2014.



Fuente: Anuario de Nacimiento y Defunciones INEC2014.

La CIE-10 constituye uno de los estándares internacionales más usados para elaborar estadísticas de morbilidad y mortalidad en el mundo, en la Figura 4, se muestra el patrón de distribución de Neoplasias malignas en el Ecuador.

Figura 4. Distribución de Neoplasias malignas

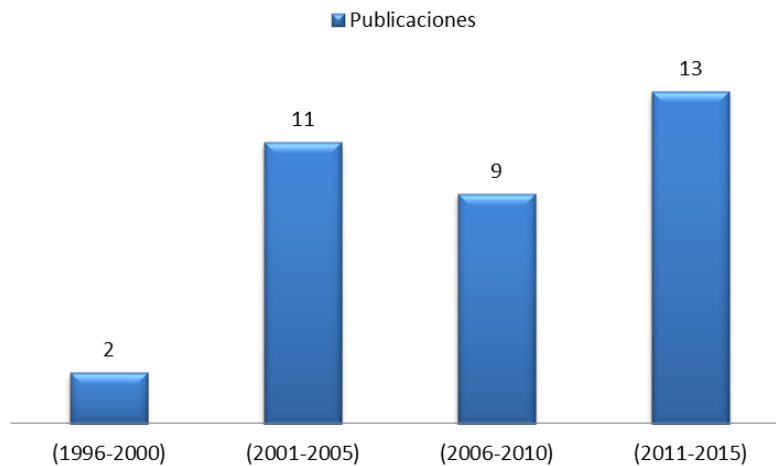


. **Fuente:** Anuario de Nacimiento y Defunciones INEC2014

MÉTODOS ESTADÍSTICOS APLICADOS EN ESTUDIOS DE MORTALIDAD

Se realizó una revisión bibliográfica de los estudios publicados que abordaban las diferentes técnicas estadísticas y el cáncer. Se consultó como referencia de partida las publicaciones en el Ecuador para denotar los tipos de trabajos que se han publicado en la Web of Science, (consulta junio 2016), cifras que ofrecen una idea de la información del Ecuador y del tipo de estudios en el contexto de la investigación mundial, en la Figura 5, se muestra el informe de artículos relacionados con epidemiología, cáncer y Ecuador en los últimos 20 años.

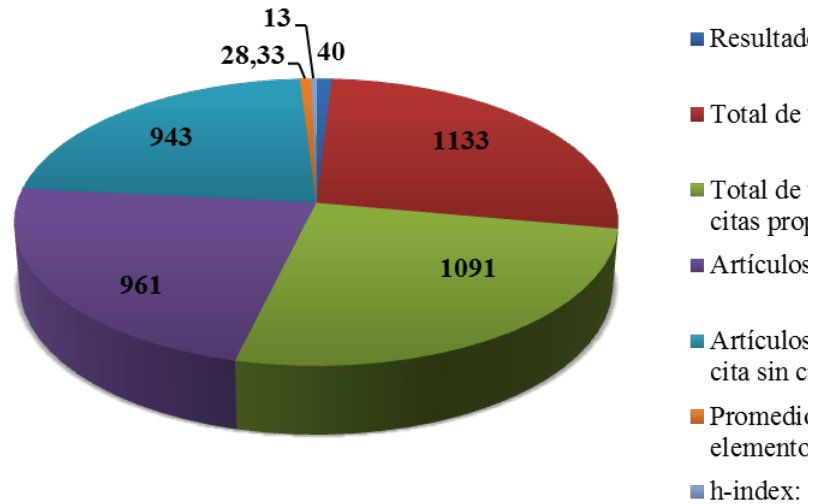
Figura 5. Informe de citas Ecuador.



Fuente: Web of Science 2016

En la Web of Science se filtró por país: se utilizó operador "AND" fue utilizado para recuperar toda la literatura existente de tal manera; Tema: (epidemiology) AND Tema: (cáncer) AND Tema: (ecuador), en la Figura 6, se muestra el informe de citas de los artículos publicados con referencia a Ecuador.

Figura 6. Informe de citas Ecuador.



Fuente: Web of Science 2016

Pacheco-Ojeda y colaboradores (1997) en su análisis sobre el cáncer de laringe en Ecuador, se basaron en tasas de supervivencia general de 5 a 10 años, de 75 casos de cáncer de laringe evaluadas y tratadas en el Hospital de la Seguridad Social de Quito (Ecuador).

San Sebastián y colaboradores (2001) en su publicación las exposiciones y la incidencia de cáncer cerca de los campos de petróleo en la cuenca del Amazonas de Ecuador, en su análisis estadístico usado se basó en la comparación de los números observados y esperados de los casos de cáncer, estadísticas descriptivas.

Hurtig y colaboradores (2002), estudiaron las diferencias geográficas y la incidencia del cáncer en la cuenca del Amazonas de Ecuador, su relación con la residencia cerca de los campos de petróleo, su análisis estadístico se basó en series de tiempo midiendo el riesgo de todos los cánceres y

cánceres específicos para la contaminación por hidrocarburos, de la región amazónica.

Zucca y colaboradores estudiaron la leucemia en niños y jóvenes de la provincia de Azuay, Ecuador: 2000–2010 (2013), mediante un modelo bayesiano mapearon la enfermedad para su respectivo análisis.

La reducción de la mortalidad por cáncer cérvico uterino en Chile, 1990-2003, los autores Donoso y colaboradores (2006), para su análisis tuvieron la finalidad de comparar la mortalidad general por cáncer en la mujer chilena entre los años 1997 y 2003, establecer los sitios más frecuentes de mortalidad y la variación de las tasas entre los dos años estudiados, efectuaron comparación de proporciones, se calculó el Chi cuadrado con corrección de Yates.

Cayuela y colaboradores (2008) en la mortalidad por cáncer de pulmón en España, su evolución en 25 años (1980-2005), analizaron las tendencias de las tasas en la investigación, calcularon las tasas específicas globales y truncadas (35-64 años) por grupos de edad y género, que se articulan como tasas por 100.000 personas al año, el análisis de tendencias de las tasas se utilizaron modelos de regresión joinpoint

Cecchini y colaboradores (2009) para los programas de detección del cáncer de cuello uterino en las comunidades de bajos ingresos en Ecuador. Detección de bajo costo de infección por VPH en un país en desarrollo. En su análisis se basaron en estadísticas descriptivas de muestras de laboratorio por un periodo de tiempo de tres años consecutivamente.

Para Sansó Soberats y colaboradores (2010), realizaron un estudio descriptivo y retrospectivo de la mortalidad por cáncer en Cuba.

En las muertes causadas por cirrosis de Isabel Aguilar y colaboradores en su artículo desigualdades y mortalidad por cirrosis en varones (Zaragoza, 1996-2003). (2011), para cada unidad de análisis se calculó la razón de mortalidad estandarizada cruda y suavizada mediante un modelo lineal generalizado mixto bayesiano.

Sobre la tendencia de la mortalidad por cáncer de cuello uterino en Chile los autores Vidal y colaboradores (2013) usaron para evaluar los cambios en la tendencia el modelo joinpoint estratificado por grupos de edad, en ambos modelos se empleó el porcentaje de cambio anual (PCA) como medida resumen.

Martínez y colaboradores (2013), el valor de la estadística para la Salud Pública, señalan que los métodos estadísticos que se emplean en diversos campos de la salud pública, que son el diagnóstico de enfermedades, predicción del resultado probable de un programa de intervención, elección apropiada de intervención en paciente o en una región, administración sanitaria pública, realización y análisis en las investigaciones en salud.

Los autores Restrepo y colaboradores en su trabajo Análisis comparativo de la esperanza de vida en Sudamérica, 1980 - 2010 (2014), utilizaron el análisis multivariado de la varianza (Manova) con un contraste canónico de tipo ortogonal, un análisis descriptivo exploratorio de tipo unidimensional, con el uso de diferentes bases de datos provenientes de la Organización Mundial de la Salud.

Metodologías propuestas para el análisis de información de mortalidad, Jones y colaboradores (2011), en su artículo clase y el estilo de vida entre los hombres de mediana edad y mayores: un análisis de correspondencias múltiples de la British Regional Heart Study, sugieren que el tabaquismo y consumo de alcohol afectan a un grupo fuerte de edad en los hombres.

La mortalidad por cáncer Moolgavkar y colaboradores, sobre la mortalidad por cáncer y la producción cuantitativa de petróleo en la región amazónica de Ecuador, 1990-2010, (2014), se utilizó la regresión de Poisson para estimar las razones de tasas de mortalidad para estudiar el impacto de las acciones de exploración y producción de petróleo en las tasas de mortandad por cáncer en el Oriente.

Domínguez en su artículo morbilidad y mortalidad de la ileostomía derivativa temporal en la cirugía por cáncer de recto, utilizaron el método de análisis de componentes principales (PCA). (Domínguez *et al*, 2014).

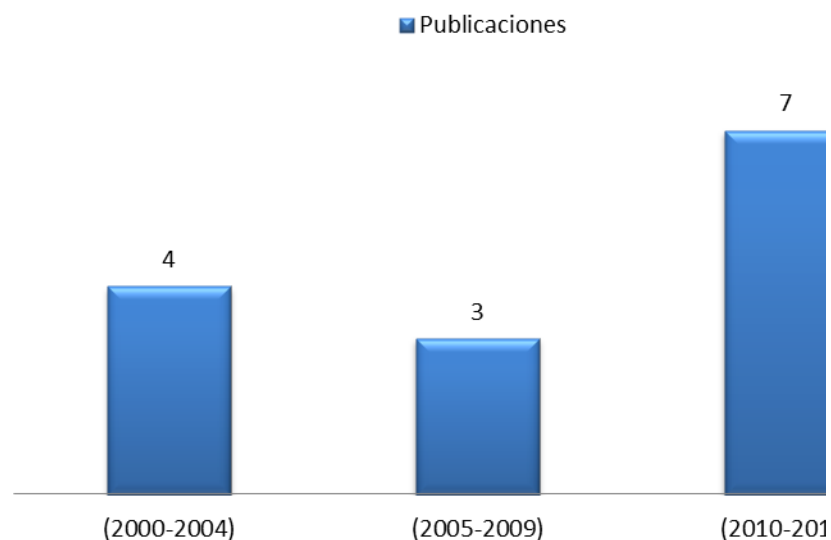
De Luna Ortega y colaboradores, (2015) la detección Automatizada de Cáncer de Mama Basada en Espectroscopia Raman y el Clasificador de Mahalanobis, utilizaron el método de análisis de Clúster y el Análisis de Componentes Principales (PCA).

Para estudiar la mortalidad causada por cáncer que produce el humo de tabaco ambiental de Aura Palencia (Palencia, 2015), utilizo el método de tabla de contingencia de doble entrada, para determinar los valores de sensibilidad, de especificidad, y valores predictivos, en el estudio sobre cáncer de mama (Natal, 2016), realizo un análisis de la varianza de un factor (ANOVA).

Wang y colaboradores (2016) la clasificación automática para la retro dispersión en microespectro a escala de una sola célula usando análisis de componentes principales. Señalan que el diagnóstico precoz y el tratamiento para el cáncer es la clave para reducir la mortalidad de cáncer, recogieron micro espectros de las células epiteliales del estómago normales y células cancerosas. Con el fin de segregar los dos tipos de células, usaron análisis de componentes principales en los que se utilizaron las dos primeras puntuaciones de los componentes principales.

En la Web of Science se filtró por operador "AND" fue utilizado para recuperar toda la literatura existente de tal manera; TOPIC: (epidemiology) AND TOPIC: (cáncer) AND TOPIC: (Multiple correspondence analysis), consulta junio 2016 que se muestra en la Figura 7, se muestra el informe de artículos relacionados con epidemiología, cáncer y análisis de correspondencias múltiples en los últimos 20 años.

Figura 7. Informe de citas epidemiología, cáncer y ACM.



Fuente: Web of Science 2016

Alaba y colaboradores (2014), en su artículo las desigualdades socioeconómicas en la prevalencia de obesidad en adultos en Sudáfrica: un análisis de descomposición. En su análisis indica que, para medir las desigualdades socioeconómicas entre los adultos de Sudáfrica, se basó en los datos de la encuesta nacional de dinámica de Ingresos de Sudáfrica del 2008 y utilizando análisis de correspondencias múltiples (ACM) como un indicador de estatus socioeconómico y medir las contribuciones de la obesidad por género.

Souza y colaboradores en su artículo (2014), Perfil de los pacientes con cáncer de pulmón del Instituto Nacional del Cáncer, por consumo de tabaco, 2000-2007, como método estadístico utilizaron el análisis de correspondencias múltiples para identificar las características predominantes entre los fumadores y los no fumadores.

Para Carter y colaboradores, (2014) en su publicación Crowdsourcing awareness: exploration of the ovarian cancer knowledge gap through Amazon Mechanical Turk. Se utilizaron tablas de contingencia para evaluar las relaciones entre los datos, análisis de regresión logística y el análisis de correspondencia para examinar las diferencias residuales entre el elemento de medidas repetidas de los datos.

La mayoría de trabajos consultados en las bases de datos académicas, estos presentan sus trabajos bajos técnicas estadísticas básicas realizando solamente cálculo de tasas de incidencia, mortalidad, análisis de series de tiempo, y los más actuales indican el análisis de varianza, uso de técnicas de regresión., modelo joinpoint, modelo lineal generalizado mixto bayesiano

Se realizó una revisión bibliográfica de los artículos publicados que se enfocaban en el cáncer y sobre su mortalidad. Para ello, se consultaron las bases de datos, Web of Science. Entonces los métodos más utilizados en el campo de la salud, tenemos las Componentes Principales (PCA), Análisis Factorial (AF), Análisis de Correspondencia Múltiples (ACM) entre otros, expuesto en su investigación por los autores (Tenenhaus, M. y Young, F. , 1985) el ACM ha tenido varios enfoques en muchas ocasiones, por diversos autores, y bajo nombres o enfoques diferentes. No obstante, la tradición más conocida fue establecida por la denominada Escuela Francesa (Benzécri J. , 1973), en la cual se adopta propiamente la expresión ‘ACM’ (Ledesma, 2008).

Artículos publicados sobre cáncer, epidemiología, mortalidad en la cual utilicen técnicas de análisis multivariantes, como el Análisis de Correspondencia Múltiples (ACM), para este método tenemos el trabajo que mediante un estudio descriptivo transversal analizaron pacientes con diagnóstico de cáncer pulmonar durante el periodo 2007-2010 del Instituto Nacional de Cancerología (INC) de Bogotá (Colombia), donde se identificó los datos de tipo laboral registrada. Se realizaron análisis descriptivos de la información reportada y análisis de correspondencias múltiples para explorar posibles asociaciones entre variables y el reporte de información laboral (Gil *et al*, 2015).

Se consultó sobre Análisis de Procrustes Generalizado (APG) en cáncer, epidemiología y mortalidad, como la siguiente propuesta de análisis multivariantes no se obtienen publicaciones usando esta técnica se tomó como referencia un estudio que tiene como objetivo de evaluar la importancia del análisis conjunto de los datos en la cual se calcularon las correlaciones entre las matrices de distancia resultante de

ambas la matriz de consenso. Luego se realizó un Análisis de Procrustes Generalizado (APG), para describir, a través de una ordenación consenso, aspectos morfológicos y moleculares en simultáneo, del conjunto de accesiones de quínoa y su agrupamiento, considerando los ambientes de origen de las accesiones. (Rodríguez *et al*, 2011).

Capítulo 2

ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLE (ACM)

Dentro del análisis de datos, los métodos multivariantes que nos permite analizar los patrones sistemáticos de variaciones con datos categóricos, se procede usar el ACM, es una técnica descriptiva que, a diferencia del ACP, estudia las relaciones entre varias variables categóricas y más precisamente, entre sus categorías (Escofier, 1990). Su dominio de aplicación, por tanto, son tablas rectangulares que contienen por filas individuos y por columnas contienen las categorías posibles de las variables. Se parte de una matriz de datos Z con n filas y tantas columnas como la suma de las categorías del conjunto de variables a analizar (p). Entre estos se puede citar los trabajos de Benzecri (1978), Greenacre (1984) y el análisis de homogeneidad de Gifi (1981).

Para el análisis de correspondencias múltiples (ACM) que contiene la información correspondiente a más de una variable categórica y es una tabla rectangular donde las filas son individuos u observaciones y las columnas son normalmente variables en codificación condensada con categorías o modalidades representadas por números consecutivos (Quaglino, Pagura, 1998), indicamos R a la matriz $I \times J$ de variables en su codificación condensada y r_{ij} a la categoría de la variable j escogida por el individuo i . Si k_j es el número de modalidades posibles de la variable j , entonces puede construirse una matriz Z de termino general Z_{ik} tal que: se ordenan los datos en una tabla $Z=(Z_1, Z_2, \dots, Z_s)$, la cual se denomina tabla disyuntiva completa que consta de

un conjunto de individuos $I=1,\dots,i,\dots,n$ (en filas), un conjunto de variables o caracteres cualitativos J_1,\dots,J_k,\dots, J_q (en columnas) y un conjunto excluyente de modalidades $1,\dots,m_k$ para cada carácter cualitativo.

$$J = \sum_{k=1}^Q m_k$$

Siendo Z_i una matriz $(n.p_i)$

de forma que $Z_{ij} = \begin{cases} 1 & \text{si el individuo } i\text{-ésimo elige modalidad } j \\ 0 & \text{si el individuo } i\text{-ésimo no elige modalidad } j \end{cases}$

La proposición siguiente indica la forma del estadístico ji-cuadrado para la matriz binaria Z en caso esta fuera tratada como una tabla de contingencia. El estadístico ji-cuadrado asociado a la prueba de independencia sobre la matriz binaria Z viene dado por

$X^2 = n(J-Q)$ el estadístico ji-cuadrado asociado a $Z = [Z_{ij}]$

Inercia total $X^2(Zq) = ntr(Q^{-T}Q)$

TABLA DE BURT.

La matriz de Burt, simétrica y compuesta de las modalidades de todas las variables usadas, la cual al ser dividida en bloques de matrices muestra cómo se comportan a través de su frecuencia absoluta, los individuos respecto a las modalidades de cada variable y, al mismo tiempo, respecto de las modalidades cruzadas de otras variables (López-Roldán, 2013). Es importante tener en cuenta que ambas

tablas conducen a similares resultados. Ahora bien, la escala difiere ya que los valores propios de \mathbf{B} son, el cuadrado de los valores hallados según la tabla \mathbf{Z} , se representa con $\mathbf{B}=\mathbf{Z}'\mathbf{Z}$, donde las columnas son variables binarias codificadas con 1 ó 0. En el ACM se analiza la llamada tabla de Burt, en las que las observaciones están descritas por un conjunto de variables cualitativas (es decir, categóricas), se observó en varios artículos basados en el tema (Varon, 2011), señalan que la verdadera estructura y ventaja de la ACM se basa en sus propiedades de reducción de dimensiones que nos permiten visualizar nuestros datos, entre otras cosas. Además de los valores propios, la fila de las coordenadas proporciona información sobre la estructura de las filas en la tabla de análisis.

A su vez, la columna muestra información sobre la estructura de las variables analizadas y sus correspondientes categorías formadas por Q_2 bloques, de modo que sus bloques de la diagonal $\mathbf{Z}'_k\mathbf{Z}_k$ cuyos elementos son matrices diagonales que cruzan una variable con ella misma, siendo los elementos de la diagonal los efectivos de cada modalidad k_j , conteniendo las frecuencias marginales de cada una de las variables analizadas (Lebart *et al*, 1984).

Figura 8. Tabla Burt ACM.

	J_1	J_2	...	J_Q
J_1	0 0
J_2	.	0 0
.
.
.
J_Q	0 0

B es una yuxtaposición de tablas de contingencia y está formada de s^2 bloques de la forma de la Figura 8. Cada bloque es una submatriz formada por tablas de contingencia de las variables dos a dos, salvo los bloques que se están en la diagonal que son las tablas de contingencia de cada variable consigo misma.

Las inercias principales del análisis de Burt son los cuadrados de los de la matriz binaria. Dado que las inercias principales son menores que 1, sus cuadrados darán valores más pequeños. En consecuencia, los porcentajes de inercia serán siempre mayores en el análisis de Burt (Aguirre *et al*, 2013). La inercia de B será la media de las inercias de las subtablas B_{qs} , que lo componen. Los valores altos de las inercias de la diagonal de la matriz de Burt explican su inercia total y por lo tanto también explica los bajos porcentajes de la inercia de los ejes.

El centro de gravedad de la nube de puntos de cada variable $N(j)$ en análisis múltiple de correspondencias (ACM) es $\sqrt{f_i}$, que en este caso puede equipararse a una distribución uniforme $1/\sqrt{n}$, ya que:

$$K_i = \sum_j K_{ij} = Q \Rightarrow \sum_j K_i = nQ \Rightarrow f_i = \frac{1}{n}$$

Entonces para la inercia en el centro de gravedad de la nube de modalidades

$$G = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)$$

$$d^2(j, G) = n \sum_{i=1}^n \left(\frac{Z_{ij}}{Z_{.j}} - \frac{1}{n} \right)^2 = n \sum_{i=1}^n \left(\frac{Z_{ij}}{Z^2_{.j}} - \frac{2Z_{ij}}{nZ_{.j}} + \frac{1}{n^2} \right) = \frac{n}{Z_{.j}} - 1$$

La distancia es mayor si el efectivo es pequeño.

Inercia de la modalidad j:

$$I(j) = \frac{Z_{.j}}{nq} d^2(j, G) = \frac{Z_{.j}}{nq} d^2\left(\frac{n}{Z_{.j}} - 1\right) = \frac{1}{q} \left(1 - \frac{Z_{.j}}{n}\right)$$

Hay mayor inercia si el efectivo es pequeño

Inercia de la variable k:

$$I(K) = \sum_{j=1}^{P_k} i(j) = \sum_{j=1}^{P_k} \frac{1}{q} \left(1 - \frac{Z_{.j}}{n}\right) = \frac{1}{q} (P_k - 1)$$

La inercia crece con el número de modalidades. Si $P_k = 2$ entonces el mínimo $1/p$

$$\text{Inercia total: } I = \sum_k i(k) = \sum_k \frac{1}{q} (P_k - 1) = \frac{1}{q} (p - q) = \frac{p}{q} = 1$$

Entonces no tiene significado estadístico, para lo cual la interpretación del ACM se da

La proximidad entre los individuos en términos de ser semejantes o parecidos, entonces dos individuos se parecen si tienen casi las mismas características. Es decir, dos individuos están próximos si han elegido global las mismas características.

La proximidad entre características de variables diferentes en términos de asociación, se dicen que son cercanos puesto que en el global están presentes en los mismos individuos, entonces dos características están próximas si han sido elegidas globalmente por el mismo conjunto de individuos.

Si las modalidades que están próximas pertenecen a una misma variable, dado que dichas modalidades son mutuamente excluyentes y no pueden estar asociadas, eso

indica semejanza entre los grupos de individuos que las han elegido.

ANÁLISIS DE PROCRUSTES GENERALIZADO

Estos métodos tienen como objetivo principal la búsqueda de una estructura común para las diferentes condiciones el cual difiere en la manera en que dicha estructura se construye (configuración consenso). (Demey *et al*, 2008). Existen dos tipos análisis de Procrustes el tipo I, que se mantengan las distancias entre los individuos de todas las configuraciones individuales, y el tipo II, que se minimice la suma de cuadrados entre los puntos análogos, en otras palabras los correspondientes al mismo elemento y su centroides, que es sobre los ejes del Análisis de Correspondencias. El Análisis de Correspondencias Múltiples (ACM) trabaja sobre la tabla X que se obtiene colocando todas las tablas originales previamente ponderadas una junto a la otra y así obteniendo la configuración consenso, tal que Y_1 y Y_2 las matrices formadas por los k ejes principales para las filas de las matrices X_1 y X_2 .

El Análisis Procrustes (AP) es una técnica enteramente matemática basado en geometría que se concentra en la búsqueda de una configuración consenso óptima Y_1 y Y_2 , que represente la verdadera estructura subyacente a los datos, aplicando transformaciones o movimientos sobre la matriz Y_2 , manteniendo Y_1 como matriz de referencia, en el sentido de aproximar lo máximo posible las distintas configuraciones asociadas a cada matriz individuos-variables de modo que se preserven las inter distancias entre los puntos hasta obtener el mejor ajuste entre ambas matrices.

Una de las características principales de esta técnica frente a las tradicionales es que los resultados que proporcionan son de gran interés desde el punto de vista de la interpretación, los movimientos que preservan las distancias entre puntos o que minimizan la función de pérdida son: la traslación y rotación de los ejes, y las reflexiones y dilataciones de la configuración completa (Tucker, 1958; Gower, 1975; Gower y Dijksterhuis, 2004).

Por un lado, en el Análisis Procrustes (AP) a los grupos presentes su estructura se puede identificar, y la variabilidad resultante se podría determinar si de la falta de integración se debe a las diferencias propias que aporta cada configuración o a la que se están comparando espacios no equivalentes, sobre el mismo plano factorial, pudiéndose observar las proximidades entre los puntos individuos de las diferentes tablas.

TRANSFORMACIÓN PROCRUSTES

Por un lado, en el Análisis Procrustes (AP) su transformación se realiza en las coordenadas de Y_2 , manteniendo Y_1 como referencia, hasta hallar una nueva matriz $Z = \rho Y_2 T$, donde ρ es un factor de escala y T una matriz de transformación ortogonal, donde las desigualdades entre Y_1 y Z sean mínimas.

Es así puesto que, deberán elegirse valores de ρ y T de forma que L se minimice, habiendo sido L la suma de cuadrados de la interdistancias dada entre las nubes de puntos de las Y_1 y Y_2 configuraciones, tal que su modelo es:

$$L = tr \left[(Y_1 - \rho Y_2 T)' (Y_1 - \rho Y_2 T) \right]$$

La correlación entre Y_1 y Y_2 T fuese igual a 1, no necesariamente sus escalas tienen que coincidir, entonces es necesario dilatar o comprimir la matriz, para esto se utiliza el escalar ρ , donde las escalas de Y_1 y Y_2 sean comparables, siendo ρ de la forma:

$$\rho = \frac{\text{tr}(\mathbf{T}'\mathbf{Y}_2'\mathbf{Z}\mathbf{Y}_1)}{\text{tr}(\mathbf{Y}_2'\mathbf{Z}\mathbf{Y}_1)}$$

Gower (1975) indica que para dos configuraciones ha sido generalizada para M matrices, si todas las configuraciones son comparadas a una configuración de referencia, y así obteniéndose una única configuración de consenso. Entonces la configuración de referencia es la media de las configuraciones individuales, esta transformación para M configuraciones es lo que se denomina Análisis de Procrustes Generalizado (APG).

Configuración consenso Z a través de la rotación

$$\mathbf{Z}_m = \rho_m \mathbf{Y}_m \mathbf{T}_m = \mathbf{Z} + \mathbf{E}_m$$

donde \mathbf{Z}_m , ρ_m , \mathbf{Y}_m y \mathbf{T}_m son los definidos previamente y \mathbf{E}_m representa el error; todos para $m=1, \dots, M$ configuraciones.

El Análisis de la Varianza clásico, propuesto por Gower (1975), plantea completar el Análisis de Procrustes Generalizados (APG) realizando un análisis de la varianza, trastornando la variabilidad total en una componente compensada a la traslación y otra a la orientación que se divide en un término de consenso y uno residual:

$$V_{\text{Total}} = V_{\text{Consenso}} + V_{\text{Residual}}$$

Tabla 1. Descomposición de la suma de cuadrados en el Análisis de Procrustes Generalizado (APG)

Fuentes de variación	Sumas de Cuadrados		
	Orientación	Dimensión	Individuos
		n	
Consenso	$M \sum_{i=1}^n \sum_{j=1}^k (\bar{z}_{ij} - \bar{z}_j)$	$M \sum_{i=1}^n (\bar{z}_{ij} - \bar{z}_j)$	$M \sum_{j=1}^k (\bar{z}_{ij} - \bar{z}_j)$
Residual	$\sum_{m=1}^M \sum_{i=1}^n \sum_{j=1}^k (z_{ij(m)} - \bar{z}_{ij})$	$\sum_{m=1}^M \sum_{i=1}^n (z_{ij(m)} - \bar{z}_i)$	$\sum_{m=1}^M \sum_{j=1}^k (z_{ij(m)} - \bar{z}_j)$
Total	$\sum_{m=1}^M \sum_{i=1}^n \sum_{j=1}^k (z_{ij(m)} - \bar{z}_{ij})$	$\sum_{m=1}^M \sum_{i=1}^n (z_{ij(m)} - \bar{z}_i)$	$\sum_{m=1}^M \sum_{j=1}^k (z_{ij(m)} - \bar{z}_j)$

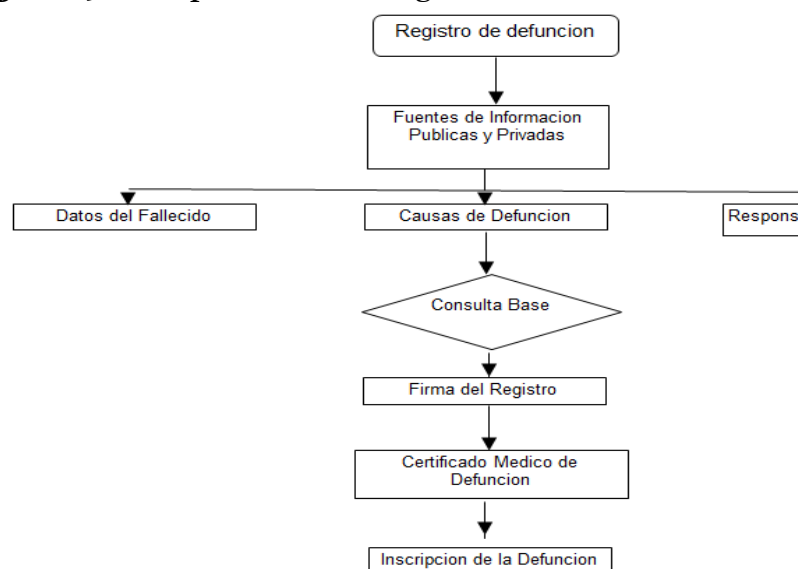
Fuente: Demey, J. R. 1998

Capítulo 3

FUENTE DE LA INFORMACIÓN

La base de datos que se utilizó para el análisis de este documento, es el resultado de la labor conjunta realizada entre instituciones estatales y de publicación pública que son efectuadas por el INEC, en conjunto con otras oficinas como la Dirección Nacional de Registro Civil, Identificación y Cedulación y del Ministerio de Salud Pública, integrantes del Sistema Estadístico Nacional que constituyen la fuente de información y está disponible en el sitio web de Ecuador en cifras (INEC, 2014). Sobre mortalidad, existen dos instancias: la primera que se refiere al aspecto médico-estadístico que se encarga de certificar el acontecimiento del hecho y la segunda que se refiere a la institución pública, al de Registro Civil, Identificación y Cedulación que se encarga de la inscripción y legalización del hecho vital, en la Figura 9, se muestra el diagrama.

Figura 9. Componentes del Registro de Defunción.



Fuente: Estadísticas Vitales en el Ecuador noviembre 2015 INEC

Actualmente en el INEC se puede tener acceso a las Estadísticas Vitales (Nacimientos y Defunciones) esta información es sistemática y continúa, se refiere a la inscripción de nacimientos y defunciones. Se revisó las bases de datos de las defunciones generales desde el año 1990 hasta el 2014, de todas las provincias del Ecuador, la cantidad de variables presentes en cada una de las bases es 48, basados en nuestro objetivo general se verifico los datos contenidos en la matriz de mortalidad por NEOPLASIAS (tumores cancerígenos) tomando los 10 últimos años (2005-2014), las causas de fallecimiento consideradas son solo las relacionadas a nuestro estudio, con esta premisa la base global de todos los 10 años, se redujo a 89 961 registros de defunción con sólo cinco variables de interés como se muestra en la tabla 2.

Tabla 2.- Variables del Estudio

Ítem	Variables	Tipo	Descripción
1	Año	Cuantitativa	Se registra el año del diagnóstico de cáncer y el deceso del individuo
2	Sexo	Cualitativa	Consignado en el documento de identidad del fallecido (femenino, masculino)
3	Edad	Cuantitativa	Edad del individuo al momento de fallecer
4	Región	Cualitativa	Región donde falleció el individuo, de acuerdo a la ubicación de las provincias en el territorio ecuatoriano
5	Neoplasias	Cualitativa	Neoplasias por las que falleció el individuo categorizada según CIE-10, número asignado automáticamente por el Sistema de Registro de Tumores

Pre-procesamiento de los datos

Depuración de la Base de datos

Después de la revisión de cada una de las base de datos por cada año de estudio, se procedió a definir las variables de interés, en la cual como primer paso, antes de realizar el análisis de los datos, es observar la existencia de datos faltantes, no se tomó en cuenta la variable “Etnia o Raza” ya que en los registros de las bases de datos no se tomaron en cuenta desde el 2005 hasta el 2009 este último con muchos casos desconocidos, en la variable “Provincia”, a partir del 2008 hasta el 2012 presento cambios en la parte territorial algunos cantones pasaron a ser provincia entonces la discontinuidad de esta variable podría efectuar algún tipo de ruido en el análisis, en la variable “Edad” se encontró un 0.04% de datos desconocidos que si fue tomada en cuenta, en la variable “Región” se tomó en cuenta solo tres de las cuatro regiones del país solo las continentales, efectuado estas revisiones se pudo observar que la matriz inicial fue reducida a 89 926 registros de defunciones siendo esta nuestra data.

Categorización.

Como segundo paso después de la depuración de las base de datos, fue el de categorizar las variables por concepto de construcción de tablas para el proceso de análisis, la variable “Año”, inicialmente es cuantitativa, pero se procedió a etiquetarla para hacerla cualitativa, la variable “Sexo”, inicialmente es numérica pero se procedió a categorizarla [1=H]; [2=M] y paso a ser cualitativa, la variable “Edad”,

inicialmente es numérica pero se procedió a categorizarla en seis grupos etarios [0-14]; [15-29]; [30-49]; [50-64]; [65-74]; [75+] y paso a ser cualitativa.

Etiquetado de variables.

El tercer paso para facilitar el análisis de este trabajo de fin de master es que para cada característica que forma parte de este análisis fue etiquetarla como se presenta en la Tabla 3 a continuación.

Tabla 3. *Etiquetas de la base de Datos*

Descripción	Etiqueta	Descripción	
Año	A2005	2005	
	A2006	2006	
	A2007	2007	
	A2008	2008	
	A2009	2009	
	A2010	2010	
	A2011	2011	
	A2012	2012	
	A2013	2013	
	A2014	2014	
	Sexo	H	Hombre
		M	Mujer
	Edad	G1	[0-14]
		G2	[15-29]
G3		[30-49]	
G4		[50-64]	
G5		[65-74]	
G6		[75+]	

Región	A	Amazónica
	C	Costa
	S	Sierra

Para la variable “Neoplasias” esta presentaba una codificación de 87 tipos de causas de muerte, se procede a agruparlas y así facilitar el análisis posterior, realizando la agrupación tal como se indica en Tabla 4.

Tabla 4.- Grupo de Neoplasias

CATEGORIA	CODIGO	DESCRIPCION	VARIABLE
1	C00- C14	Labio, cavidad oral y faringe.	Orl
2	C15-C26	Neoplasias digestivas.	Diges
3	C30- C39	Neoplasias de órganos respiratorios e intratorácicos.	Resp
4	C40-C41	Neoplasias malignas de hueso y cartílago articular	Oseo
5	C43- C44	Neoplasias de piel.	Piel
6	C45- C49	Neoplasias malignos de tejidos conectivos y blandos	Tblando
7	C50	Neoplasias malignas de mama	Mama
8	C51-C58	Neoplasias órganos genitales femeninos.	Fem
9	C60- C63	Neoplasias de órganos genitales masculinos.	Masc
10	C64- C68	Neoplasias de órganos urinarios	Uro
11	C69	Neoplasia maligna de ojo y sus anexos	Ojo

12	C70-C72	Neoplasias malignas del sistema nervioso	Snc
13	C73-C75	Neoplasias malignas glándulas endócrinas	Gendoc
14	C76-C80	Neoplasias malignas secundarias	Metast
15	C81	Enfermedad de Hodgkin	Ehodgkin
16	C82-C85	Linfoma no Hodgkin	Lnh
17	C91-C96	Leucemia	Leucemias
18	C88, C90 y C97	Otras enfermedades malignas	Otros

Fuente: Estadísticas Vitales en Ecuador noviembre 2015 INEC

Según los datos oficiales del Censo 2010, el Ecuador tiene 14'483.499 habitantes, 120 habitantes por km², para el presente TFM se trabajó las bases de datos de mortalidad por neoplasias de 10 años (2005-2014), a nivel de todo el Ecuador teniendo un universo de estudio de 89926 personas fallecidas por 18 grupos, se indican las frecuencias en los grupos de neoplasias que se conformaron para el presente TFM, como vemos en la tabla 5.

Tabla 5.- Frecuencias por tipos de Neoplasias

Descripción	Frecuencia	Porcentaje
Labio, Cavidad Oral Y Faringe.	951	1,1
Neoplasias Digestivas.	35103	39
Neoplasias De Órganos	7470	8,3

Respiratorios e Intratorácicos.		
Neoplasias Malignas De Hueso y Cartílago Articular	970	1,1
Neoplasias De Piel.	1507	1,7
Neoplasias Malignos De Tejidos Conectivos y Blandos	1024	1,1
Neoplasias Malignas De Mama	4402	4,9
Neoplasias Órganos Genitales Femeninos.	8867	9,9
Neoplasias De Órganos Genitales Masculinos.	8342	9,3
Neoplasias De Órganos Urinarios	2105	2,3
Neoplasia Maligna De Ojo y Sus Anexos	153	0,2
Neoplasias Malignas Del Sistema Nervioso	2990	3,3

Neoplasias Malignas Glándulas Endócrinas	1222	1,4
Neoplasias Malignas Secundarias	5080	5,6
Enfermedad De Hodgkin	451	0,5
Linfoma No Hodgkin	2970	3,3
Leucemia	4966	5,5
Otras Enfermedades Malignas	1353	1,5

Fuente: Estadísticas Vitales en Ecuador noviembre 2015
INEC

Matriz de datos PARA ACM

En la matriz de datos de defunciones por neoplasias se aplicaría los métodos enunciados en el marco teórico, con la matriz original de 89926 registros, estarían conformadas por las variables SEXO, AÑO, GETAREO, REGIÓN Y NEOPLASIAS, en las columnas **J** estarían las combinaciones de los individuos de la fila **i**, adicionalmente se analizaría en **Q** años determinados por el período de estudio 2005-2014. Cada celda X_{ij} representada en la Figura 10.

Figura 10. Matriz de datos

	VARIABLES				J
	SEXO	AÑO	GETAREO	REGION	NEOCODIF
I INDIVIDUOS	X_{ij} $X (89926 \times 5)$				

Las características fueron consideradas y etiquetadas, de la siguiente manera ya que para el ACM aplica a tablas de contingencias en donde por filas hay (n) individuos y por columnas (s) variables categóricas con (pi =1,2,L,s) mutuamente excluyentes y exhaustivas, El ACM se basa en realizar un análisis de correspondencias sobre la llamada matriz de Burt: $B = Z'Z$, esta se construye partiendo de la tabla disyuntiva, que se convierte a una tabla de Burt que contiene todas las tablas de contingencia simples entre las variables (cruzadas dos a dos).

A partir de la tabla disyuntiva completa se puede construir la tabla de contingencia de Burt (B), que es una tabla simétrica de orden (p, p): $B = Z' \cdot Z$, donde B es una yuxtaposición de tablas de contingencia y está formada de s2 bloques de la forma ver la figura 11.

Figura 11. Tabla Disyuntiva Neoplasias.

SEXO		AÑO												GETAREO				REGION			NEOCODIF																	
		H	M	A2005	A2006	A2007	A2008	A2009	A2010	A2011	A2012	A2013	A2014	G1	G2	G3	G4	G5	G6	A	S	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		Z_1		Z_2												Z_3				Z_4			Z_5															
																																					Z	
λ_z																																						

La matriz de Burt se construye por superposición de cajas. En los bloques diagonales aparecen matrices diagonales conteniendo las frecuencias marginales de cada una de las variables analizadas, fuera de la diagonal aparecen las tablas de frecuencias cruzadas correspondientes a todas las combinaciones 2 a 2 de las variables analizadas. Se toman como dimensiones aquellas cuya distribución a la inercia supera $(1/p)$. Como se observa en la figura 12.

Figura 12. Matriz de Burt Neoplasias.

SEXO		AÑO												GETAREO				REGION			NEOCODIF																	
		H	M	A2005	A2006	A2007	A2008	A2009	A2010	A2011	A2012	A2013	A2014	G1	G2	G3	G4	G5	G6	A	S	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0		0												0				0			0															
1																																						
P_1	$(B = Z'Z)$																																					
P_2	$\lambda_B = \lambda_z^2$																																					

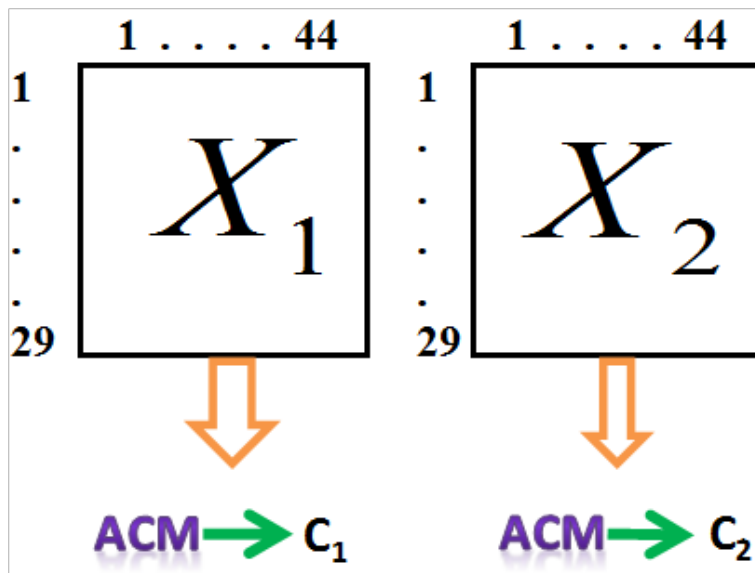
Tal como se describe y se observa en la figura 11 y 12, en la figura 16 se explica todo el proceso del ACM.

Matriz de datos para AGP.

Las configuraciones geométricas obtenidas mediante las coordenadas principales, para representar la estructura y su relación empírica en un conjunto de elementos o individuos, éste requiere de un ordenamiento, por el cual se utiliza el Procrustes, este propone una armonización de todas las configuraciones individuales a través de una serie de pasos algebraicos para las transformaciones de estas, la metodología algebraica como normalización, rotación, reflexión y escalamiento de los datos bajo dos criterios, el primero es que se mantengan las distancias entre los individuos de todas las configuraciones individuales, dos es que se minimice la suma de cuadrados entre los puntos análogos, en otras palabras los correspondientes al mismo elemento y su centroides. La configuración de consenso se obtiene como la media de todas esas configuraciones individuales transformadas.

El AGP de la información que resulta del Análisis de Correspondencias Múltiples (ACM) ya que se trabaja sobre la tabla X que se obtiene colocando todas las tablas originales previamente ponderadas una junto a la otra y así obteniendo la configuración consenso, tal que Y_1 y Y_2 las matrices formadas por los k ejes principales para las filas de las matrices X_1 y X_2 , como se observa en la figura 13.

Figura 13. Extracción de Componentes principales para el AGP



Las transformaciones o movimientos sobre la matriz Y_2 , manteniendo Y_1 como matriz de referencia, en el sentido de aproximar lo máximo posible las distintas configuraciones asociadas a cada matriz individuos-variables de modo que se preserven las inter distancias entre los puntos hasta obtener el mejor ajuste entre ambas matrices.

Figura 14. Extracción del matriz consenso para el AGP

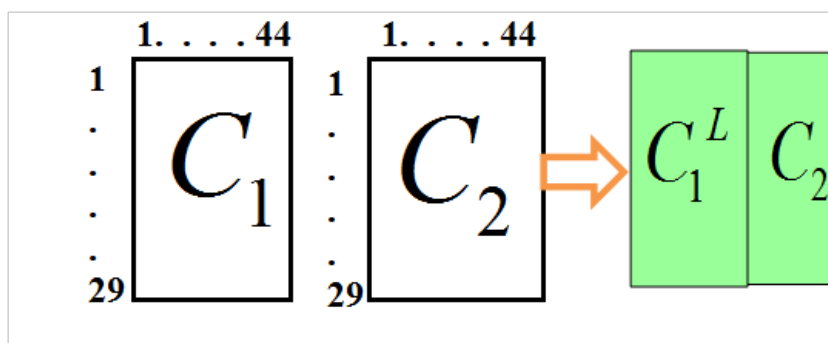
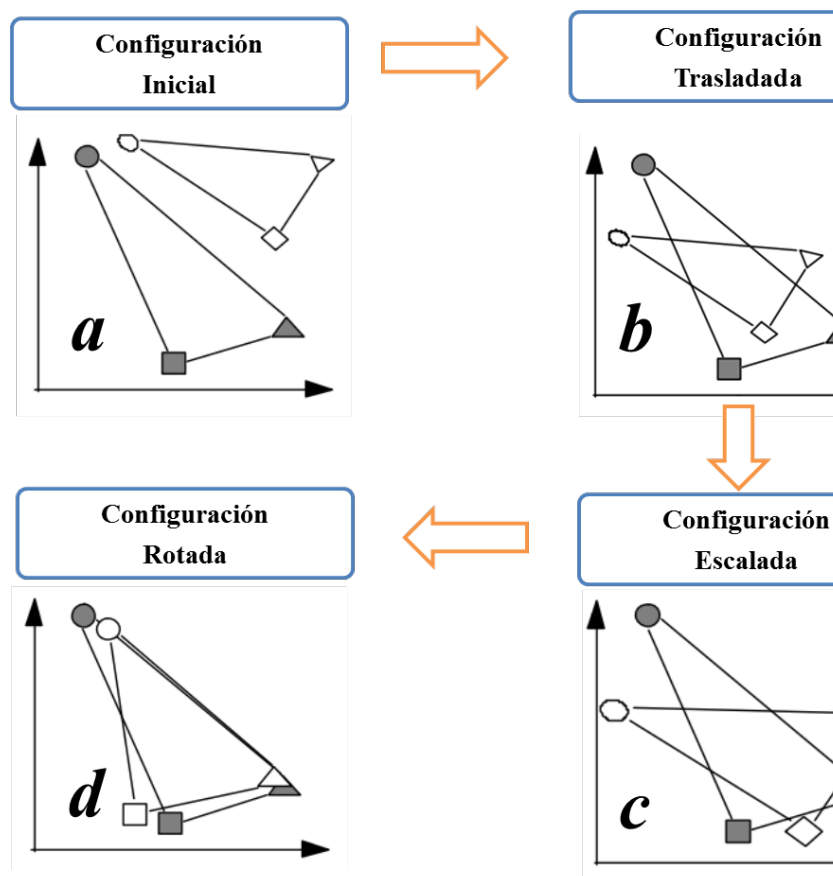


Figura 15. Configuraciones del AGP



Fuente : IRTA

Tal como se describe y se observa en la figura 13,14 y 15, en la figura 17 se explica todo el proceso del AGP.

Todos los análisis y representaciones gráficas, fueron realizados utilizando la librería 'ca' de R-project (Nenadic, O., Greenacre, M. , 2016), Infostat Versión 2016 (Di Rienzo *et al*, 2016).

Figura 16. Diagrama de Flujo ACM

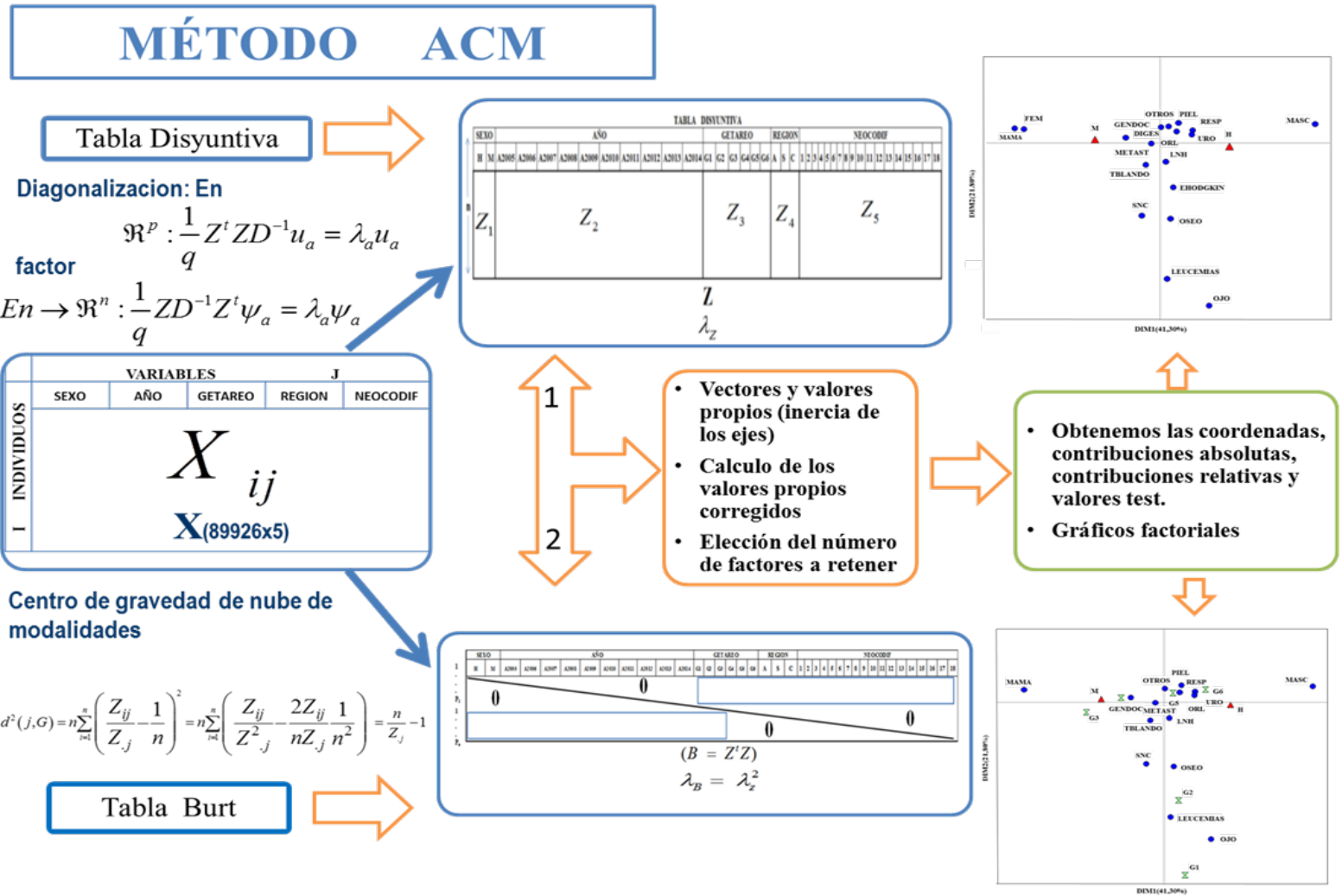
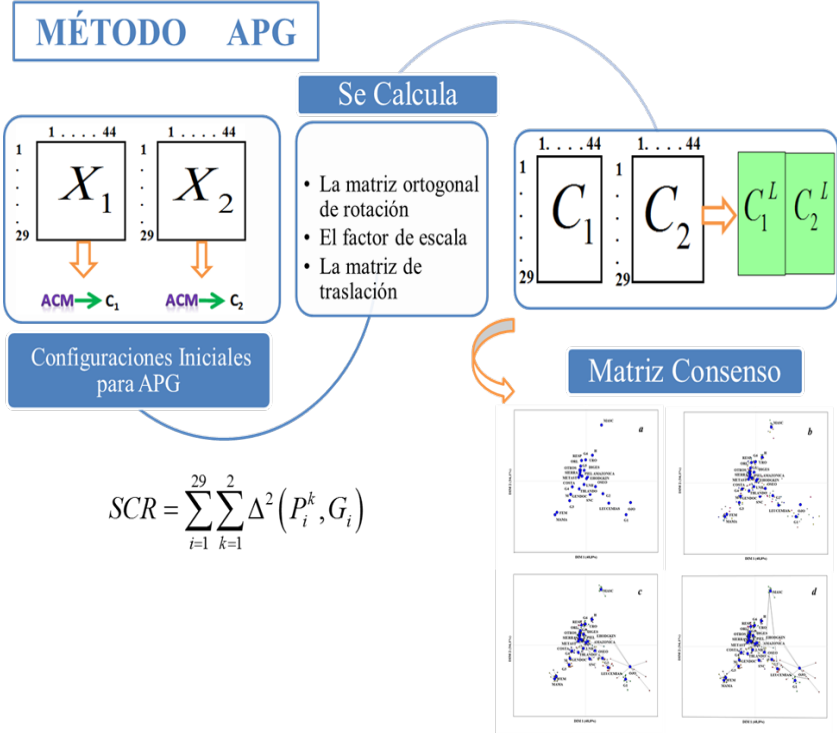


Figura 17. Diagrama de Flujo AGP



Capítulo 4

análisis de correspondencias múltiple (acm)

El Análisis de Correspondencias Múltiples como primer método de análisis multivariante para obtener un perfil completo de la mortalidad por cáncer en el Ecuador considerando cinco variables, en los años de estudio 2005 hasta el 2014, se representó la información bajo una estructura de asociaciones entre un grupo de variables categóricas (genero, año de fallecimiento, región de fallecimiento, edad de fallecimiento y neoplasias), los mismos que han permitido mejorar la interpretación.

Los siguientes puntos representan las modalidades de una misma variable respuesta, conectados con las dimensiones para cada variable de estudio, se obtuvo la tabla que se muestra a continuación representada en la Figura 18, que explica sobre su Inercia total, una medida similar a la variación total del grado de dependencia existente entre las variables X e Y, se calcularon las proporciones de inercia explicada por cada una de las dimensiones i que nos ayudan a medir la importancia de cada una de las dimensiones a la hora de explicar las dependencias observadas así como las proporciones de la inercia acumulada explicada por las 2 primeras dimensiones, para la primera (41.3% de la inercia). La segunda con (21.8% de la inercia) esto nos indicó el número mínimo de dimensiones necesario para explicar dichas dependencias a las contribuciones totales para medir la importancia de cada una de las modalidades de las variables analizadas en la construcción de los ejes o factores

construidos por el ACM donde los dos primeros ejes o dimensiones que son las que mejor representan. Estos 2 ejes acumulan 63.1% de la inercia total,

La Figura 19, muestra el espacio de las dos dimensiones generadas por la descomposición de la matriz X_{ij} , medida por la matriz de Burt para la generación de las tres primeras dimensiones y la comparación de las diferentes estructuras sobre las regiones por cada tabla (año), es decir, las regiones son a la vez similares desde el punto de vista de un grupo de neoplasias o de los grupos etarios, los valores propios de los análisis en el ACM, que se proyectan en la tabla bidimensional, con un primer eje muy dominante. Las estructuras sobre las regiones como una combinación por las diez tablas de mortalidad.

Las diez tablas contribuyen la construcción del primer eje con un 41.3% de la inercia, respectivamente. La contribución al segundo eje de los grupos de mortalidad es igual a 21.8%.

En el primer eje vemos al grupo etario G1 (0 a 14 años), indica también una de las causas de mortalidad frecuentes en este grupo etario la interpretación de las NEOPLASIAS y se denota los puntos de fila (los círculos azules) y están son las “NEOPLASIAS OJO” en la dimensión uno, no tan alejado también tenemos una fuerte presencia sobre este eje a las “NEOPLASIAS LEUCEMIAS” en la dimensión uno, con una alta presencia de mortalidad sobre el grupo etario G2 (15 a 29 años) en esta dimensión, no muy alejado de este grupo de edad tiene presencia sobre la misma las “NEOPLASIAS OSEO” en la primera dimensión, “NEOPLASIAS SNC” en la primera dimensión, “NEOPLASIAS TBLANDO” en la primera dimensión, “NEOPLASIAS METAST” en la primera dimensión, que se empieza a alejar para la izquierda del eje uno, “NEOPLASIAS LNH” en la primera dimensión, para

esta dimensión se muestra el comportamiento individual de cada grupo etario, se distribuyen en el eje 2; existe dispersión también en la dimensión 2; sin embargo, existen casos que se proyectan sobre el eje 1, “NEOPLASIAS ORL” con en la primera dimensión, “NEOPLASIAS URO” en la primera dimensión, “NEOPLASIAS PIEL” en la primera dimensión, “NEOPLASIAS RESP” en la primera dimensión, con presencia de mortalidad sobre el grupo etario G5 (65 a 74 años) y G6 (75+ años) en ambos grupos etarios, los puntos más dispersos, “NEOPLASIAS MASC” en la primera dimensión, para la izquierda de la dimensión 2 tenemos “NEOPLASIAS OTROS” para la segunda dimensión, “NEOPLASIAS GENDOC” en la segunda dimensión, “NEOPLASIAS DIGES” para la segunda dimensión, con presencia de mortalidad sobre el grupo etario G4 (50 a 64 años) y G3 (30 a 49 años), los puntos más dispersos, están las “NEOPLASIAS FEM en la segunda dimensión, “NEOPLASIAS MAMA” en la segunda dimensión.

La inercia total del grafico en un 63.1% de la dimensión de la matriz original, reduce aproximadamente en un 90%, con lo que se simplifica considerablemente la interpretación y los puntos de fila (las figuras como relojes de arena de color verde) se representan a continuación utilizando las dos primeras dimensiones en la cual se obtienen una mejor compresión en la proyección de la dimensión 1 para los grupos etarios 1,2 y 3, personas desde 0 hasta 49 años de edad, existiendo separación marcada en el grupo etario 1 (0 a 14 años) y 2 (15 a 29 años), los puntos que representan al género (los triángulos de color rojo) donde en el lado positivo del eje se encuentran aquellos pacientes de todas las regiones que son de género masculino y de la dimensión 2 para los grupos etarios 4, 5 y 6, personas desde 50 hasta mayores de 75 años de edad, existiendo separación marcada en el grupo etario 4 (50 a 64 años) del lado negativo del eje

se encuentran aquellos pacientes de todas las regiones que son de género femenino.

En ambos grupos etarios se caracterizan por pertenecer a las regiones: amazónica, costa y la sierra, su interpretación se indican con los puntos de fila (los triángulos hacia abajo de color amarillo).

Figura 18. Inercia de las Neoplasias

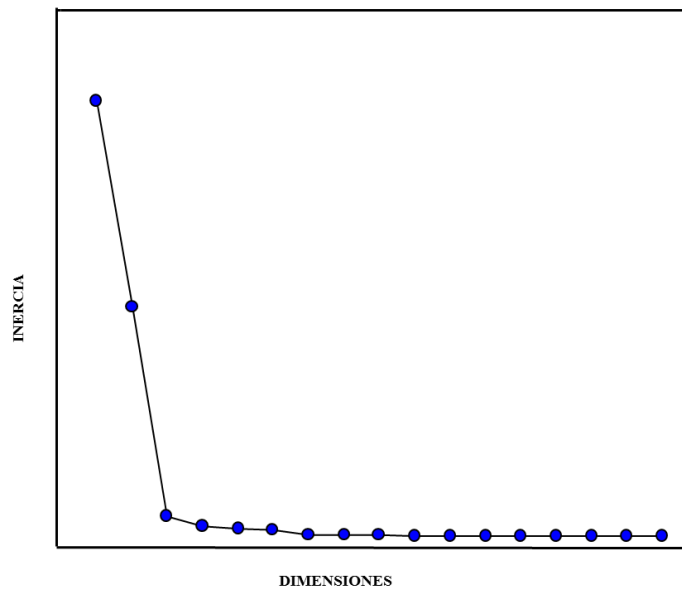
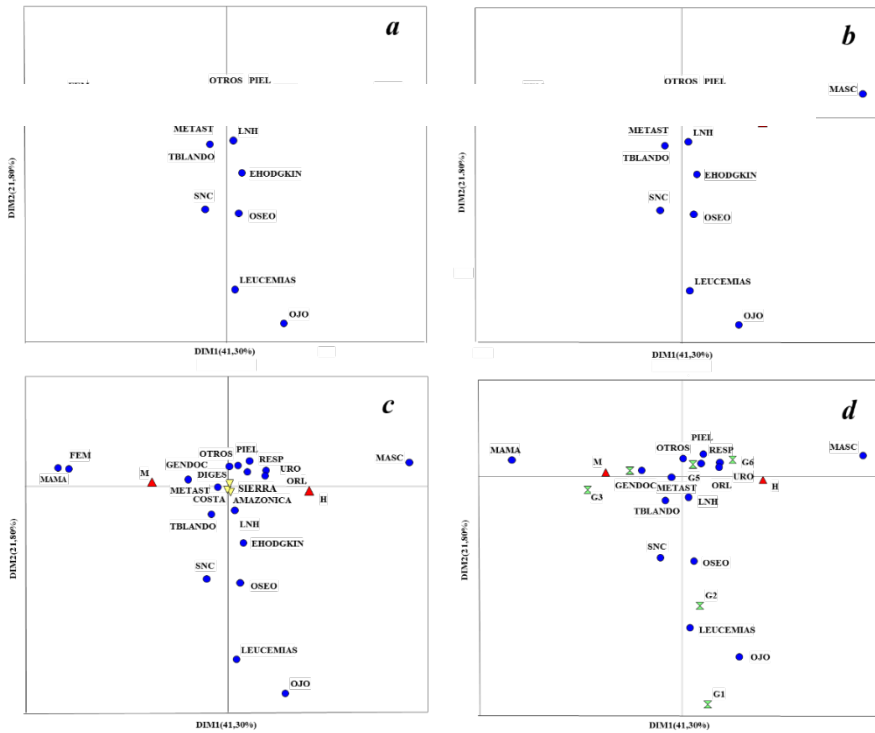


Figura 19. Neoplasias por Género, Grupo etario, Regiones



Análisis Procrustes Generalizado (APG)

Se utilizó el software InfoStat para el análisis de Procrustes generalizado provee un método para consensuar ordenaciones que involucran tres acciones, la traslación (centrado de las ordenaciones), la rotación (de las ordenaciones para minimizar diferencias entre ellas) y el de escalamiento (que es la multiplicación por factores para minimizar diferencias de tamaño), mediante el ACM se generó un matriz de 29 categorías por cada año con tres

dimensiones cada tabla, mediante el AGP se buscó cuantificar el consenso entre la ordenación de los individuos obtenida por el análisis de coordenadas principales de la matriz de distancia de las categorías y la ordenación de los mismos individuos mediante los datos por años.

Se puede observar en la Tabla 6 que las categorías del estudio neoplasias se denota los puntos de fila (los círculos azules), neoplasia OJO tiene la mayor suma de cuadrados consenso, marcando su carácter diferencial con el resto. Entonces, a partir de la suma de cuadrados residuales, se observa que las categorías del estudio 4(G1),12(FEM),16(MAMA) y 17(MASC) son las que poseen mayores diferencias entre las tablas (mayor efecto en las tablas). No pasa lo mismo con las categorías del estudio 1(AMAZONIA), 26(TBLANDO), 15(LNH) y 18(METAST) no presentan diferencias importantes entre ellas; es decir, serían las poblaciones con un comportamiento más homogéneo, siendo en consecuencia las más estables (menor efecto en las tablas). Por lo tanto, la Figura.20 permite corroborar las interpretaciones anteriores.

Se observa en la Figura.21, la representación bidimensional de la configuración consenso estimada con la aplicación del Análisis Procrustes sobre las dos configuraciones obtenidas, el plano principal representa el 48.8% de la variabilidad total de los datos, se observa en el componente uno que las neoplasias a la derecha del gráfico son TBLANDO, LNH y la más alejada OJO. En cambio, más a la izquierda en el mismo componente uno están las neoplasias, GENDOC, PIEL, MASC, DIGES y OTROS, con la presencia de la región SIERRA, el grupo etario G6 y el género femenino. En relación al segundo componente, se presentan en la parte derecha del gráfico las neoplasias LEUCEMIAS, URO, OSEO, SNC, EHODGKIN, con la presencia de los grupos etarios G2, G1 y

la región AMAZONICA, hacia la izquierda del grafico está presente el género masculino con los grupos etarios G3, G4, G5, las neoplasias RESPIRA, ORAL, MAMA, FEN, debemos observar que los géneros están casi en el centro por ese motivo algunas de sus neoplasias están cercanas unas de otras y está presente la región COSTA.

Se puede observar en la Tabla 7 En el grupo 1 tiene el valor más alto de los demás grupos, en relación al grupo 4 estaría con el siguiente mejor valor. Para los grupos 5, 7, 9 y 10 tienen valores por debajo de los demás grupos.

En la Figura 22 la representación bidimensional de la configuración consenso estimada con la aplicación del Análisis Procrustes sobre el grupo de características principales del componente 1 presenta rendimientos por debajo de la media, es importante destacar que esta caracterización representa un comportamiento medio de ambas tablas lo cual hace que si existe interacción de las características del estudio con los ambientes no pueda visualizarse en este análisis. se cuantificó el consenso entre la configuración espaciales del grupo etarios, neoplasias, regiones y genero obtenidas del ACM, en los datos agrupados podemos observar las trayectorias y así, luego de encontrada la rotación Procrustes correspondiente, rotar las categorías no comunes, de manera de preservar las posiciones relativas de las mismas respecto a las categorías que tienen mayor presencia y se definen en los dos primeros componentes, en el plano bidimensional y los factores obtenido con este proceso, nos permite visualizar las similitudes entre las mismas de los diferentes años de estudio. Algunos son irregulares en magnitud y, en algunos de los casos, muestran considerable dispersión, las que poseen mayores discrepancias entre las caracterizaciones en ambos, se presentan los diagramas de dispersión se observa que la

suma de cuadrados (SC), en función de las diferencias entre las ordenaciones individuales y la de consenso dentro de cada caso, es bastante similar ya que la ordenación de consenso se da para todos los casos en el punto medio de las distancias entre las configuraciones individuales y estas distancias son similares para todos los casos.

El análisis de la varianza de la Tabla 8 permite identificar la contribución relativa de los individuos y las categorías sobre SCR, la variabilidad total es el equivalente a la suma de cuadrados total, donde la suma de cuadrados se presenta por grupos a cada categoría "plano bidimensional" esta puede descomponerse por tablas o por individuos, a partir de la proyección de los individuos sobre la primera dimensión que recoge la mayor cantidad de inercia en el plano factorial, los puntos que representan las categorías en las dos primeras dimensiones de cada individuos están más distanciado de las contribuciones, es notorio el comportamiento que tiene la categoría regiones y la de género, diferente del resto que evidencia la máxima suma de cuadrados totales, se presentan los diagramas de dispersión (sobre los que se superpusieron las configuraciones consenso) se observa que la suma de cuadrados (SC), en función de las diferencias entre las ordenaciones individuales y la de consenso dentro de cada caso, es bastante similar ya que la ordenación de consenso se da para todos los casos en el punto medio de las distancias entre las configuraciones individuales y estas distancias son similares para todos los casos, no para los de género.

La Tabla 9 pone de manifiesto los principales factores de variabilidad de los individuos, la configuración de consenso, es obtenida como el promedio de todas las configuraciones individuales, tenemos un primer factor o dimensión

claramente preponderante, que recoge un 48,88%, el segundo factor o dimensión, que recoge un 34,1%, de la variabilidad total de los datos, proyectándose en el plano bidimensional principal (plano factorial óptimo compuesto por las dos primeras dimensiones) tal como se ve indica más de la mitad de la inercia, es determinado a un 82,8% de la misma ($\lambda_1 = 0,396$; $\lambda_2 = 0,277$).

Tabla 6. Cuadro de Análisis de la Varianza (APG)

Categorías	Consenso	residuo	Total	Prop Cons
H	0,192	0,006	0,198	0,971
M	0,169	0,003	0,172	0,983
G1	1,137	0,053	1,19	0,955
G2	0,265	0,119	0,384	0,69
G3	0,246	0,018	0,264	0,933
G4	0,103	0,005	0,108	0,952
G5	0,046	0,009	0,055	0,837
G6	0,122	0,007	0,129	0,944
AMAZONICA	0,017	0,046	0,062	0,265
COSTA	0,061	0,022	0,083	0,736
SIERRA	0,043	0,009	0,052	0,833
DIGES	0,055	0,002	0,057	0,97
EHODGKIN	0,248	0,154	0,402	0,616
FEM	0,683	0,016	0,699	0,976
GENDOC	0,295	0,073	0,367	0,802
LEUCEMIAS	0,4	0,036	0,436	0,918
LNH	0,021	0,056	0,077	0,268
MAMA	0,782	0,038	0,82	0,953
MASC	0,922	0,027	0,948	0,972
METAST	0,029	0,1	0,129	0,228
OJO	1,28	0,441	1,721	0,744
ORL	0,093	0,067	0,16	0,58
OSEO	0,112	0,178	0,29	0,385

OTROS	0,072	0,106	0,178	0,404
PIEL	0,326	0,1	0,426	0,765
RESP	0,175	0,051	0,226	0,775
SNC	0,139	0,053	0,192	0,726
TBLANDO	0,017	0,062	0,079	0,214
URO	0,069	0,026	0,095	0,725
Total	8,118	1,882	10	0,812

Figura 20. Matriz de Dimensiones

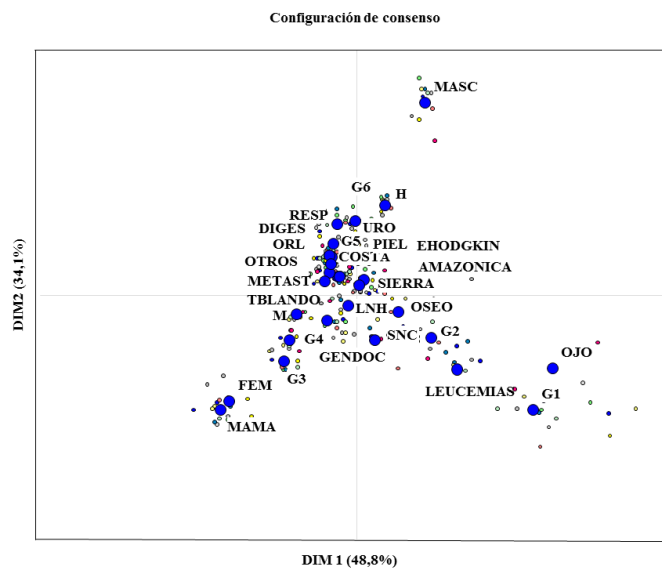


Tabla 7. *Rho*

Coficiente	Valor
Grup01	1,064
Grup02	1,052
Grup03	1,043
Grup04	1,056
Grup05	0,989
Grup06	1,016
Grup07	0,9
Grup08	1,011

Grup09	0,946
Grup010	0,907

Figura 21. Configuración de Casos

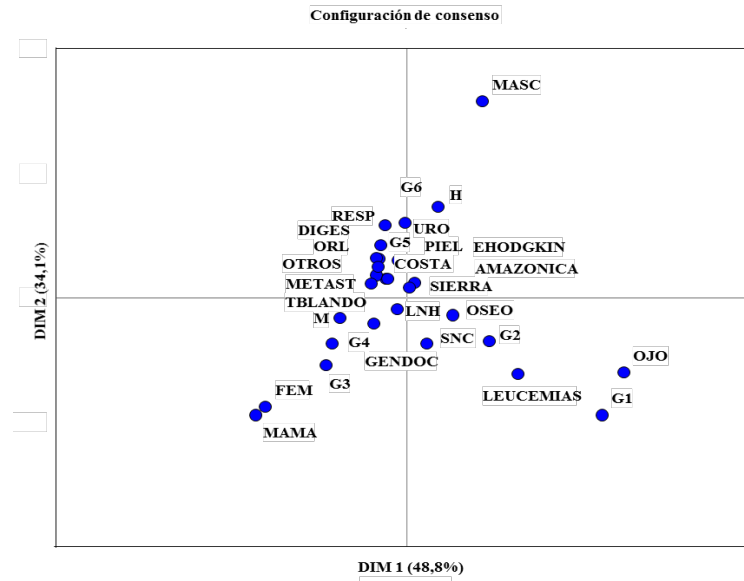


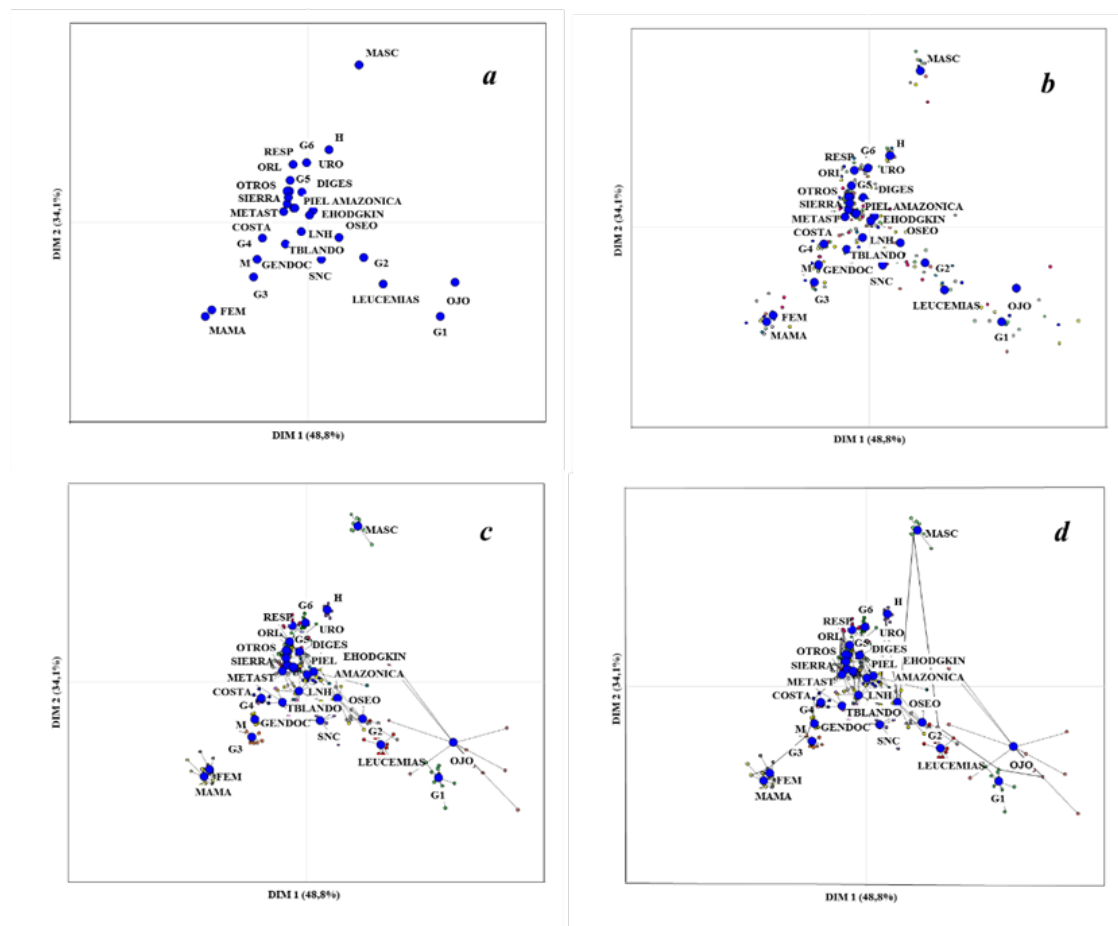
Tabla 8. Análisis de la Varianza por Grupos

	Coefficiente	Consenso	Residuo	Total	Prop	Cons
Grup01	1,031	0,102	1,132	0,91		
Grup02	0,987	0,119	1,106	0,892		
Grup03	0,955	0,132	1,087	0,879		
Grup04	1,002	0,113	1,115	0,898		
Grup05	0,775	0,203	0,978	0,793		
Grup06	0,865	0,167	1,033	0,838		
Grup07	0,497	0,312	0,809	0,614		
Grup08	0,85	0,173	1,023	0,831		
Grup09	0,639	0,257	0,895	0,714		
Grup010	0,518	0,304	0,822	0,63		
Total	8,118	1,882	10	0,812		

Tabla 9. Auto Valores

Lambda	Valor	Proporción	Prop Acum
1	0,396	0,488	0,488
2	0,277	0,341	0,828
3	0,139	0,172	1

Figura 22. Configuración Consenso por Categoría, Getario, Neoplasias, Regiones, Género



DISCUSIONES

Como objetivo de este trabajo es demostrar las características de las dos técnicas ACM y APG a partir de una base de datos de 89926 muertes causadas por cáncer que representa a diez años desde el 2005 hasta el 2014, información que es revisada y validada para su publicación por el INEC en sus boletines anuales en Ecuador, las correlaciones entre las configuraciones obtenidas con ACM y APG , con respecto que en este caso particular, como las variables son cuantitativas, la metodología del ACM se basa fundamentalmente en el ACP simple.

Carlos Alfaro (CIMA, 2016), explica, que en la actualidad se estima que sólo entre el 10 y el 12% de los casos de neoplasias son hereditarios. El 90% aproximadamente está vinculado al envejecimiento, al ambiente y a los hábitos de vida", las neoplasias de mayor incidencia, que son el colon rectal, próstata pulmón y mamas

Como uno de los objetivos es describir los métodos de Análisis de Correspondencia Múltiple (ACM) y proponer su uso en el contexto de estudios epidemiológicos de mortalidad causada por neoplasias malignas, comparar los individuos(en términos de mortalidad), de las variables (en términos de las neoplasias), el ACM y el APG recogen la estructura presente en los datos de mortalidad en los 10 años de estudio, el APG obtiene la conjuración consenso, por cada grupo \promedio, el ACM muestra el coeficiente RV propuesto por Robert y Escoufier (1976) que cuantifica la similitud entre configuraciones, es una herramienta teórica muy útil, sensible a la interacción.

Medina, y colaboradores (Medina, et al, 2009) se refieren a las neoplasias generalmente se presenta en pacientes de edad avanzada, se asocia con hipertensión arterial y diabetes, por ser estas enfermedades concurrentes en el grupo de edad

involucrado en su estudio, más que ser un factor patogénico, utilizaron métodos descriptivos, se basaron en registros clínicos pero también señalan, valores faltantes derivados de la pobre calidad de los registros, por los tipos de datos obtenidos, en nuestro caso la información que maneja el INEC, es validada para su reporte y puesto a disposición en su página web.

El uso de las diferentes técnicas de análisis para obtener una reducción de la dimensión depende del tipo de datos con el que se trabaja y el objetivo del análisis, los artículos publicados que se enfocaban en el cáncer y sobre su mortalidad, consultados en la Web of Science. en el campo de la salud, como métodos multivariantes, tenemos las Componentes Principales (PCA), Análisis Factorial (AF), Análisis de Correspondencia Múltiples (ACM) entre otros, expuesto en su investigación por los autores (Tenenhaus, M. y Young, F. , 1985) el ACM ha tenido varios enfoques en muchas ocasiones, por diversos autores, y bajo nombres o enfoques diferentes. No obstante, la tradición más conocida fue establecida por la denominada Escuela Francesa (Benzécri J. , 1973), en la cual se adopta propiamente la expresión 'ACM' (Ledezma, 2008).

El análisis de correspondencia múltiple (ACM), es un método de la escuela francesa y el análisis de Procrustes generalizado (APG), es un método de la escuela anglosajona, su funcionalidad parte, de la búsqueda de una estructura común (configuración consenso) para las diferentes condiciones entre individuos y variables, con la diferencia de su matriz inicial y como se la elabora, el ACM, parte de la descomposición de valores singulares (DVS) realizado en la matriz de individuos. La matriz de individuos Z es para las categorías por variables de matriz columnas, obteniendo la matriz de burt. Para el APG es una técnica enteramente de

análisis geométrico, se requiere obtener una configuración consenso óptima, en el sentido de aproximar lo máximo posible las distintas configuraciones asociadas a cada matriz individuos-variables, que es construida a partir del análisis del ACM, en nuestro caso sobre las tres primeras dimensiones por años y por cada categoría, una matriz 29 x 33, más allá de las estadísticas descriptivas para ambos modelos los resultados que proporcionan desde el punto de vista de interpretación, son sus representaciones gráficas de la nube de individuos caracterizados por las variables de cada tabla (individuos categorizados) y por el conjunto de variables, la cual muestran proximidades entre los puntos individuos de las diferentes tablas.

Como propuesta al uso de estos dos métodos de análisis multivariantes en el contexto de estudios epidemiológicos de mortalidad causada por neoplasias malignas esta información que se analiza tiene un universo de 89926 individuos es de carácter multidimensional, donde todos los casos en los cuales una variable representa a los ítems o individuos, por la cantidad de información es favorable usar tablas de contingencia multidimensionales, lo que denominamos Análisis de Correspondencias Múltiples, usando la información del ACM, la misma matriz de individuos a través de una serie de pasos geométricos, cálculos matriciales, y distribuciones multivariantes que incluyen traslación, rotación y escalamientos de los datos bajo dos criterios: que se mantengan las distancias entre individuos de las configuraciones originales para aprovechar que se minimice la suma de los cuadrados entre los puntos análogos, es decir correspondientes al mismo elemento, y su centroides se denomina análisis de Procrustes generalizados (APG), estos dos tipos de análisis multivariantes juegan un papel fundamental en este trabajo de investigación sobre las neoplasias.

Tabla 10. *Comparación de Métodos Utilizados*

Métodos	Características	Variable Indicadora	Reglas de Interpretación	Conclusión
ACM	variables son cualitativas	Las nubes de modalidades tienen las mismas propiedades inerciales en ambos análisis	Buscamos los puntos (categorías) mejor explicados por un factor o dimensión (contribuciones). Cuanto mayor sea la contribución mejor representada está la categoría en la dimensión.	Se puede afirmar que los dos análisis se complementan y proporcionan la mayor información en las dos primeras dimensiones
AGP	variables son cuantitativas	Las distancias entre individuos y entre modalidades definidas en APG y ACM mantienen una estrecha relación.		

Bibliografía

- Aguilar, I., Feja, C., Compés, M. L., Rabanaque, M. J., Esteban, M., Alcalá, T., & Martos, M. C. (2011). Desigualdades y mortalidad por cirrosis en varones (Zaragoza, 1996-2003). *Gaceta Sanitaria Volume 25, Issue 2*, Pages 139–145.
- Aguirre, K. F., Calvo, M. I. L., & Herrán, J. I. M. (2013). Nuevo procedimiento metodológico para el análisis exploratorio de una tabla estructurada en diversos conjuntos de individuos. . *Estadística española*, 55(182), 305-322.
- Alaba, O., & Chola, L. (2014). Socioeconomic inequalities in adult obesity prevalence in South Africa: a decomposition analysis. *International journal of environmental research and public health*, 11(3), 3387-3406.
- Benzécri, J. (1973). *Analyse des Données (2 vols.)*. . París: Dunod. .
- Benzécri, J. P. (1978). Problème sur la classification. . *Les cahiers de l'analyse des données*, 3(1), 95-101.
- Bray, F., Jemal, A., Grey, N., Ferlay, J., & Forman, D. (2012). Global cancer transitions according to the Human Development Index (2008–2030). *The lancet oncology*, 13(8), 790-801.
- BuenVivir. (30 de 03 de 2016). *buenvivir.gob.ec*. Recuperado el 1 de 04 de 2016, de buenvivir.gob.ec: <http://www.buenvivir.gob.ec/web/guest>
- Carter, R. R., DiFeo, A., Bogie, K., Zhang, G. Q., & Sun, J. . (2014). (2 Crowdsourcing awareness: exploration of the

ovarian cancer knowledge gap through Amazon Mechanical Turk. *PloS one.*, 9(1), e85508.

Cecchini, G., Paganini, G., D'Amico, M., Cannone, M., Bertuletti, C., & Barberis, M. C. . (2009). Cecchini, G., Paganini, G., D'Amico, M., Cannone, M., Bertuletti, C., & Barberis, M. C. (2009). Cervical cancer screening programs in low-income communities. Experiences from Ecuador. Low cost detection of HPV infection in a developing country. *Pathologic*, 101(2), 76-79.

Constitución de la República del Ecuador. (2008.). *Constitución de la República del Ecuador*,. Quito.

De Luna Ortega, C A; Martínez Romo, J C; Luna Rosas, F J; Medina Veloz, G; López Rivas, V; . (2015). Detección Automatizada de Cáncer de Mama Basada en Espectroscopía Raman y el Clasificador de Mahalanobis. *Conciencia Tecnológica*, 9-26.

Demey, J. R., Vicente-Villardón, J. L., Galindo-Villardón, M. P., & Zambrano, A. Y. . (2008). Identifying molecular markers associated with classification of genotypes by External Logistic Biplots. *Bioinformatics*, 2832-2838.

Di Rienzo J.A., Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., Robledo C.W. . (24 de 03 de 2016). *InfoStat versión 2016*. . Obtenido de Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina: <http://www.infostat.com.ar>

Domínguez, L. P., Martínez, M. T. G., Alvarado, N. C., Novella, Á. T., Grosso, A. P. H., & Núñez, J. E. C. (2014). Morbilidad y mortalidad de la ileostomía derivativa temporal en la cirugía por cáncer de recto. *Cirugía Española*, 92(9), 604-608.

Donoso, E., & Cuello, M. . (2006). Mortalidad por cáncer en la mujer chilena: análisis comparativo entre los años

- 1997 y 2003. . *Revista chilena de obstetricia y ginecología*, , 71(1), 10-16.
- Escofier, B. E. (1990). Análisis factoriales simples y múltiples: objetivos, métodos e interpretación. . *Servicio Editorial Universidad del País Vasco*,.
- Escoufier, Y. (1980). *L'analyse conjointe de plusieurs matrices de données*. In: Jolivet M, ed. *Biométrie et Temps*. . Paris:: Société Francaise de Biométrie, 59–76.
- Gil, M., Gamboa, O., & Orjuela, M. E. (2015). Occupational history documented in the medical records of patients diagnosed with lung cancer. *Revista Colombiana de Cancerología*, 19(3), 156-165.
- Globocan. (30 de 10 de 2012). *summary_table_site_sel.aspx*. Recuperado el 1 de 04 de 2016, de http://globocan.iarc.fr/Pages/summary_table_site_sel.aspx
- González, F. Á. (1999). Algunas aportaciones al Análisis de Datos, utilizando técnicas de representación Multivariante . (*Doctoral dissertation, Tese de Doutoramento. Faculdade de Ciências, Departamento de Matemática, Universidade de Cádiz*).
- Hurtig, A. K., & San Sebastián, M. . (2002). Geographical differences in cancer incidence in the Amazon basin of Ecuador in relation to residence near oil fields. . *International Journal of Epidemiology*, 31(5), 1021-1027.
- IARC. (2014). *Informe mundial sobre el cáncer 2014*, . IARC.
- IARC-OMS. (31 de Marzo de 2016). *IARC*. Obtenido de <http://globocan.iarc.fr/Pages/Map.aspx>

- INEC. (2010). *Instituto Nacional de Estadística y Censos, Anuario de estadísticas vitales: nacimientos y defunciones.* . Quito: INEC.
- INEC. (2014). *Base de datos de defunciones 2014.* Quito: INEC.
- INEC. (31 de Marzo de 2016). *Censo 2010.* Obtenido de Ecuador en Cifras: <http://www.ecuadorencifras.gob.ec/censo-de-poblacion-y-vivienda/>
- Jones, I. R., Papacosta, O., Whincup, P. H., Goya Wannamethee, S., & Morris, R. W. . (2011). Class and lifestyle 'lock-in' among middle-aged and older men: a Multiple Correspondence Analysis of the British Regional Heart Study. . *Sociology of health & illness*, 33(3), 399-419.
- Lebart, L.; Morineau, A.; y Warwick, K. M. . (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices.* . Nueva York: John Wiley & Sons, Inc. , 198 p.
- Ledesma, R. (2008). SOFTWARE DE ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES: UNA REVISIÓN COMPARATIVA. *Metodología de Encuestas Volumen 10*, 59-75 .
- López-Roldán, P. &. (2013). *Análisis de correspondencias.* . Diapositivas. Bellaterra: Universitat Autònoma de Barcelona.
- Mardi, K.V.; Kent,J.T.; Bibby,J.M. (1982). *Multivariate analysis.* Londres. Ed. Academic Press., 521 pp.

- Martínez, J. C. (2016). Factores asociados a la mortalidad por enfermedades no transmisibles en Colombia, 2008-2012. *Biomédica*, 36(4).
- Martínez, P., & Guzmán, L. (2013). El valor de la estadística para la Salud Pública. *Salus*, 4-11.
- Moolgavkar, Suresh H.; Chang, Ellen T.; Watson, Heather; et ál.. (2014). Cancer mortality and quantitative oil production in the Amazon region of Ecuador, 1990–2010. *Cancer Causes & Control*, 25(1), 59-72.
- Natal, C. (2016). Variaciones en el proceso de confirmación diagnóstica entre unidades de cribado poblacional de cáncer de mama. *Gaceta Sanitaria*.
- Nenadic, O., Greenacre, M. . (2016). Correspondence Analysis in R, with two- and three-dimensional graphics: The ca package. . *Journal of Statistical Software* , 20(3):1-13.
- OMS. (2014). *Determinantes de Inequidades en Salud*, O.P.S., O.M.S. Salud en las Américas.
- OMS. (2016). *Organizacion Mundial de la Salud*. Obtenido de http://www.who.int/cancer/iccd_2016/en/#
- Pacheco-Ojeda, L., De La Torre, M., Guerrón, M., & Letort, M. (1997). Cancer of the larynx in Ecuador. *Acta otorrinolaringologica espanola*, 48(3), 215-219.
- PAHO. (2012). *Salud en las Américas, Edición de 2012*. Edición de 2012.
- PAHO. (2016). *Health system profile. Ecuador. Monitoring and analysis of the change and reform processes*. Washington, DC: Pan American Health Organization.
- Palencia, A. (2015). Punto de corte de la cotinina urinaria y cotinina/creatinina en niños. *Avances en Biomedicina*.

- Plan del Buen Vivir. (30 de 03 de 2016). *planificacion.gob.ec*. Recuperado el 1 de 04 de 2016, de [planificacion.gob.ec/resena-historica/](http://www.planificacion.gob.ec/resena-historica/):
<http://www.planificacion.gob.ec/resena-historica/>
- Presidencia del Ecuador. (1 de 04 de 2016). *presidencia.gob.ec*. Recuperado el 27 de 03 de 2016, de <http://www.presidencia.gob.ec/>
- Quaglino, M. B., & Pagura, J. A. . (1998). Una propuesta para algunas aplicaciones de análisis de correspondencias múltiples. *Rosario, Argentina: Universidad Nacional de Rosario, Facultad de Ciencias Económicas y Estadística, Instituto de Investigaciones* .
- Registro de Tumores-SOLCA Matriz Guayaquil. (31 de Marzo de 2016). *Registro de Tumores SOLCA Matriz*. Obtenido de <http://www.estadisticas.med.ec>
- Restrepo LF, Rodríguez H. (2014). Análisis comparativo de la esperanza de vida en Sudamérica, 1980 - 2010. *Rev Univ. salud.*, 16(2):177 - 187.
- Rodrigues, S., Paulos, K., Pereira, E., Oliveira, A. F. G. D. F., & Teixeira, A. (2011). Análisis sensorial de carne seca y salada de ovinos y caprinos. . *XIV Jornada sobre Producción Animal*, , 715-717.
- San Sebastian, M., Armstrong, B., Cordoba, J. A., & Stephens, C. A. R. O. L. Y. N. (2001). Exposures and cancer incidence near oil fields in the Amazon basin of Ecuador. *Occupational and environmental medicine*, 58(8), 517-522.
- Sansó Soberats, F. J., Alonso Galbán, P., & Torres Vidal, R. M. (2010). Mortalidad por cáncer en Cuba. *Revista Cubana de Salud Pública*, 36(1), 78-94.

- Secretaría Nacional de Planificación y Desarrollo. (05 de 01 de 2013). *objetivos-nacionales-para-el-buen-vivir*. Recuperado el 12 de 05 de 2016, de <http://www.buenvivir.gob.ec/objetivos-nacionales-para-el-buen-vivir>
- Sociedad Americana del Cáncer. (2014). *EL ATLAS DEL CANCER*. Atlanta, Georgia: Sociedad Americana del Cáncer.
- Souza, M. C. D., Vasconcelos, A. G. G., Rebelo, Rebelo, P. A. D. P., & Cruz. . (2014). Profile of patients with lung cancer assisted at the National Cancer Institute, according to their smoking status, from 2000 to 2007. *Revista Brasileira de Epidemiologia*, 17(1), 175-188.
- Tenenhaus, M. y Young, F. . (1985). An analysis and synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and other methods for quantifying categorical data. *Psychometrika*, 50, 91-119. .
- TheGlobeconomy. (1 de 4 de 2016). *theglobeconomy*. Recuperado el 30 de 3 de 2016, de http://es.theglobeconomy.com/Ecuador/Death_rate/
- Varon, M. J. (2011). Aplicación de técnicas estadísticas multivariadas en perfilación y segmentación. *Universitas Scientiarum*, 16(3), 254-262.
- Vidal C, Hoffmeister L, Biagini L. . (2013). Tendencia de la mortalidad por cáncer de cuello uterino en Chile: aplicación de modelos de regresión joinpoint. *Rev Panam Salud Publica.*, 33(6):407–13.
- Wang, C., Wen, M., Bai, L., & Zhang, T. (2016). Auto-classification for confocal back-scattering micro-spectrum at single-cell scale using principal component

analysis. *Optik-International Journal for Light and Electron Optics*, 127(3), 1007-1010.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Journal of Statistical Software*, 1-25.

Zucca, M., Ugalde, J., Arteaga, F., Biggio, G., Flore, V., Nonne, T. & Ennas, M. G. . (2013). Leukemia in children and youths of the Azuay province, Ecuador: 2000–2010. . *International journal of environmental health research*, 23(1), 58-65.

Ing. Juan Tarquino Calderon Cisneros, PhD.

Investigador – Senescyt-Reg-Inv:18-02756

Coordinador de la Red Académica "Herramientas de Estadística Multivariante para el Análisis de Big Data". Nro. Senescyt-2018-040-Red –18-0011.

Docente Investigador Externo

UNIVERSIDAD ECOTEC

Dra. Vilma Raffo Babici, Mgtr.

COORDINADORA ACADÉMICA DE LA FACULTAD CIENCIAS DE LA SALUD Y DESARROLLO HUMANO

UNIVERSIDAD ECOTEC

PhD. José Luis Vicente-Villardón

Coordinador de Doctorado y Postgrado

Director del Departamento de Estadística

UNIVERSIDAD DE SALAMANCA

ISBN: 978-9942-33-571-5



compAs
Grupo de capacitación e investigación pedagógica



@grupocompas.ec
compasacademico@icloud.com