



CONACYT



CIMAT



2022 *Ricardo Flores*  
*Año de Magón*

PRECURSOR DE LA REVOLUCIÓN MEXICANA

Proyecto 10667

Análisis espacio-temporal y sociodemográfico de enfermedades crónicas no-transmisibles (ECNT) en la Cuenca del Alto Atoyac y evaluación de las capacidades de atención de acuerdo a los padecimientos detectados: elementos esenciales para la formulación de estrategias de atención-prevención en beneficio de la salud de los habitantes.

## **Metodología del índice de mortalidad municipal para la Cuenca del Alto Atoyac**

Centro de Investigación en Matemáticas, A.C.  
Gerencia de Ciencia de Datos e Inteligencia Artificial

*28 de octubre de 2022*





# Tabla de Contenidos

<b>1</b>	<b>Objetivo del documento</b> .....	<b>4</b>
<b>2</b>	<b>Terminología</b> .....	<b>5</b>
<b>3</b>	<b>Índice de Mortalidad</b> .....	<b>6</b>
<b>3.1</b>	<b>Tasas de Mortalidad</b>	<b>6</b>
<b>3.2</b>	<b>Definición del Índice de Mortalidad</b>	<b>8</b>
3.2.1	Variables seleccionadas para la construcción del índice de mortalidad .....	9
3.2.2	Construcción del índice de mortalidad .....	10
<b>3.3</b>	<b>Insuficiencia Renal Crónica</b>	<b>11</b>
<b>3.4</b>	<b>Leucemia Linfoide</b>	<b>14</b>
<b>3.5</b>	<b>Malformaciones</b>	<b>16</b>
<b>3.6</b>	<b>Efecto de las Poblaciones Pequeñas en el Índice de Mortalidad</b>	<b>18</b>



## 1. Objetivo del documento

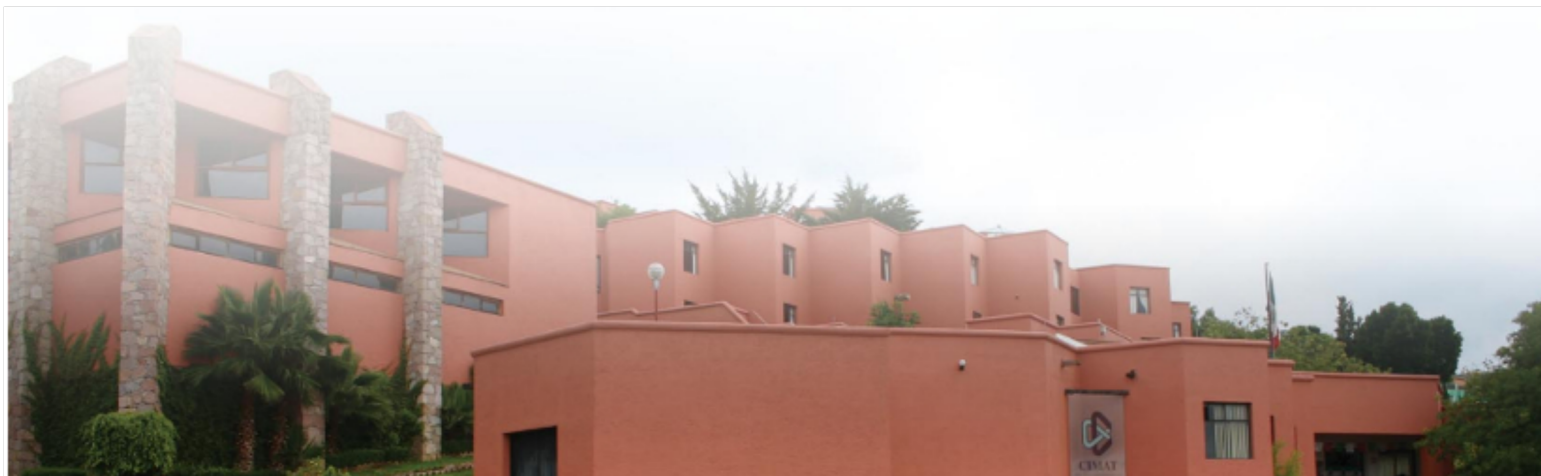
El objetivo del presente documento es describir la metodología del índice de mortalidad municipal desarrollado para la enfermedad renal crónica, la leucemia linfocítica y las malformaciones congénitas.



## 2. Terminología

A lo largo del documento, se utilizarán los siguientes términos de manera equivalente:

<b>"Malformaciones"</b>	Para referirse al grupo de causas "Malformaciones congénitas, deformidades y anomalías cromosómicas", correspondiente a los códigos CIE-10 Q00-99 [1].
<b>"Insuficiencia renal" o "Enfermedad Renal"</b>	Para referirse a la enfermedad "Insuficiencia Renal Crónica" correspondiente al código CIE-10 N18 [1].
<b>"Leucemia"</b>	Para referirse a la enfermedad "Leucemia Linfóide" correspondiente al código CIE-10 C91 [1].
<b>"Grupo de Insuficiencia Renal crónica"</b>	Para el grupo de enfermedades conformada por los códigos CIE-10 N00-23 [1].
<b>"Grupo de Leucemia Linfóide"</b>	Para el grupo de enfermedades conformada por los códigos CIE-10 C90-95 [1].



## 3. Índice de Mortalidad

En esta sección se describe técnicamente la definición del índice de mortalidad propuesto para las enfermedades crónicas no transmisibles consideradas en el proyecto: 1) insuficiencia renal, 2) leucemia y 3) malformaciones. Además se presentan algunas visualizaciones exploratorias que permiten comprender e interpretar el índice.

### 3.1 Tasas de Mortalidad

Dos fuentes de datos principales se consideraron en el análisis de la mortalidad:

1. Defunciones anuales por municipio, grupo de causa, edad y sexo desde 2009 hasta 2019.
2. Población municipal por grupos de edad desde 2000 hasta 2019, separada por sexo a partir del 2003.

Con base en dichos datos se calculan las tasas brutas de mortalidad para cada municipio, año, sexo, edad y grupo de causa, mediante la siguiente fórmula:

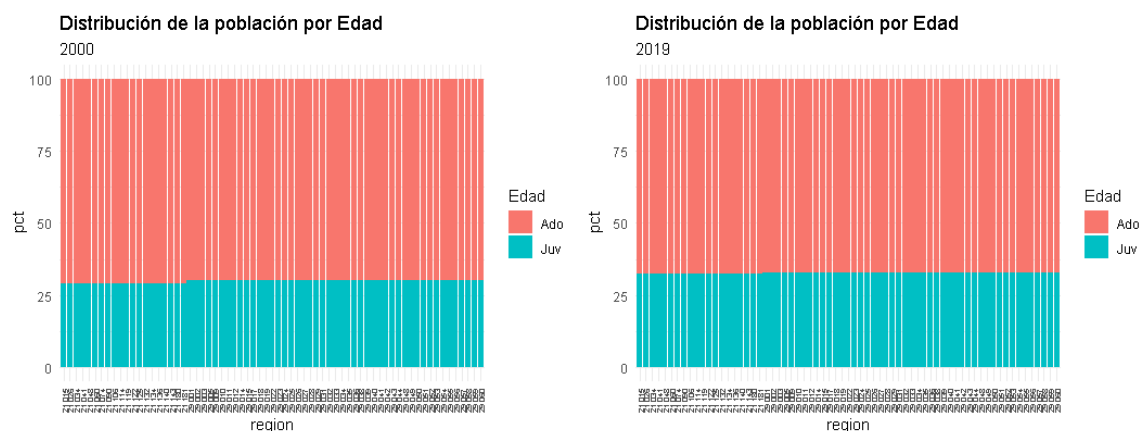
$$TBM = \frac{\#defunciones}{Población} \times 1000. \quad (3.1)$$

Al trabajar con tasas de mortalidad para distintas poblaciones y años, es usual utilizar tasas de mortalidad estandarizadas con el fin de que sean comparables. Lo anterior se recomienda debido a que es usual que la distribución de edad varíe mucho de población a población o de año a año.

En la Figura 3.1, se presenta la distribución por edad para cada uno de los municipios de la Cuenca del Alto Atoyac para los años más distintos entre sí: 2000 y 2019.

Puede observarse que la diferencia entre las distribuciones, tanto por municipio como por año, es mínima; el cambio más drástico es de una diferencia del 3%. Lo anterior, nos indica que los municipios y años son comparables entre sí y las tasas brutas podrían utilizarse. Sin embargo, por completez, se decidió estandarizar las tasas para Enfermedad Renal y





**Fig. 3.1:** Comparación de las distribuciones de edad por municipio en la Cuenca del Alto Atoyac por edad para los años 2000 (Izq.) y 2019 (Der.).

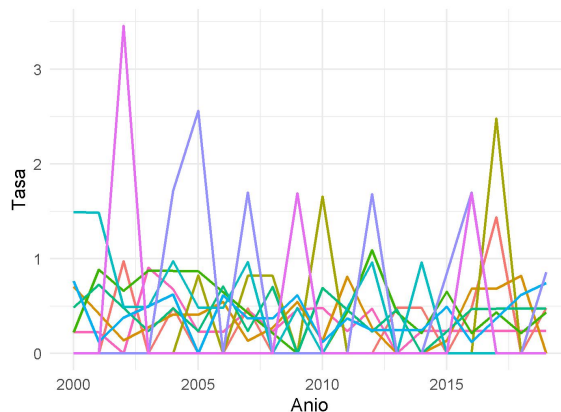
Leucemia Linfode cuando se consideran los grupos de edad "Adolescencia" y "Juventud", juntos. En el caso en que se separan los datos por grupo de edad para Malformaciones, que solamente tiene un grupo de edad (Infancia), se utiliza la tasa bruta.

Para estandarizar las tasas, se utilizó el método directo y se consideró la población de México por grupos de edad (Adolescencia y Juventud) de 2010 como población estándar, por ser la mitad del período considerado en el estudio. La tasa estandarizada por municipio, año, sexo, edad y grupo de causa, está dada por:

$$TEM = \frac{\sum_{i \in Edad} TBM_i \times PobEstándar_i}{\sum_{i \in Edad} PobEstándar_i}, \quad (3.2)$$

donde  $Edad = \{Adolescencia, Juventud\}$ .

A partir de las tasas de mortalidad, se construyen series de tiempo de mortalidad municipales. En la Figura 3.2, se presentan las series de tasas de mortalidad de infantes por Malformaciones para diez de los sesenta y siete municipios de la Cuenca del Alto Atoyac.

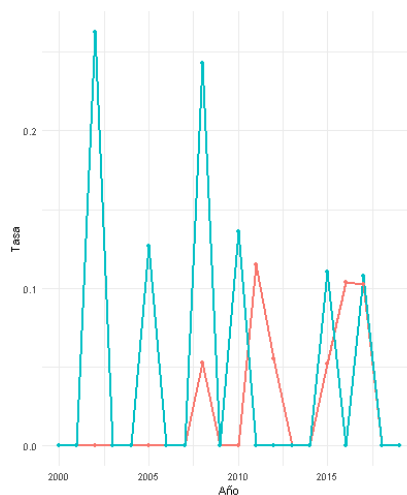


**Fig. 3.2:** Series de tasas de mortalidad de infantes por Malformaciones para diez de los sesenta y siete municipios de la Cuenca del Alto Atoyac.

A través de las series de tiempo, se puede observar el comportamiento general de la mortalidad en cada municipio por grupo causa, sexo y edad. La serie de tiempo proporciona diferente información que la observación aislada de uno de los años en particular. Por ejemplo, en la serie es posible observar si la mortalidad se ha incrementado, disminuido o ha sido estable a través del tiempo. Sin embargo, como puede imaginarse a partir de la Figura 3.2, es difícil visualizar el total de series para compararlas entre sí. Con el objetivo de comparar dichas series y resumir la información de estas en un número, se propone un índice de mortalidad.

### 3.2 Definición del Índice de Mortalidad

El objetivo del índice de mortalidad es resumir la información de las series de tiempo de mortalidad en un número, de forma tal que permita ordenar los municipios según su gravedad en cuanto a mortalidad. Aunque existen muchas formas directas y sencillas de elegir un número para representar cada serie, elegir uno que resuma todas las propiedades de la serie no es una tarea trivial. Consideré por ejemplo, las series presentadas en la Figura 3.3 y evaluemos algunas de las opciones directas para resumir estas series.



**Fig. 3.3:** Ejemplo: series cuya información se desea resumir en un único número.

Una opción de representante de la serie es la última observación. Bajo esta elección, se ignora el pasado y por tanto, dado que en 2019 en ambos municipios no hubo muertos, se tendría que ambos municipios tienen el mismo nivel de mortalidad. Sin embargo, es claro que las dos series son diferentes entre sí. Otra opción es elegir el promedio de la serie. En ese caso, la serie roja no sería catalogada como un caso "grave" pues su promedio es chico. No obstante, se puede observar que la serie roja muestra un comportamiento creciente en los últimos años, a diferencia de la azul que muestra uno decreciente, y esa información no se refleja en el promedio. Por otro lado, si se elige únicamente una estimación de la tendencia de la serie, se pierde la información que da el promedio, que en este caso es que el municipio correspondiente a la serie azul tiene tasas de mortalidad más altas en comparación a la serie roja, por lo que debería prestársele atención. Otra opción directa sería utilizar la mediana. Sin embargo, al igual que en estos dos casos particulares, las series de mortalidad municipales tienen muchos "picos", de forma tal que la mediana de la serie para la mayoría de municipios es cero.



Dado que ninguna de las opciones directas para representar las series con un número es una buena opción, se decidió considerar un conjunto de variables de medidas de tendencia central y tendencia de la serie, de las cuales se seleccionó un subconjunto para construir el índice. En la siguiente subsección 3.2.1, se describen las variables consideradas en la construcción del índice.

### 3.2.1 Variables seleccionadas para la construcción del índice de mortalidad

Una serie de mortalidad requiere atención si su comportamiento es creciente o si presenta niveles de mortalidad más altos que el resto. Bajo este supuesto, además del análisis exploratorio, se consideraron las siguientes pruebas, para resumir el comportamiento de las series de mortalidad:

- Prueba de Mann-Kendall con hipótesis alternativa de tendencia positiva para la serie completa.
- Prueba de Mann-Kendall con hipótesis alternativa de tendencia positiva para la década 2010-2019.
- Pruebas de Fliger-Killeen y Mann-Whitney-Wilcoxon de la década 2000-2009 vs. 2010-2019.
- Prueba de cambios en la media de la serie completa (2000-2019).

El procedimiento de elección de las variables finales basadas en las pruebas para las series por Grupo causa, Edad y sin separar por sexo, se describe a continuación.

Con las pruebas de Mann-Kendall a un nivel de significancia de 0.1, se detectaron cuatro series con tendencia creciente para la Insuficiencia Renal Crónica y una serie para Leucemia Linfóide al considerar únicamente la última década (2010-2019). Cuando se consideró la serie completa, se obtuvo el mismo resultado excepto por una serie de las detectadas para Insuficiencia Renal. La serie extra detectada como creciente al considerar únicamente la última década corresponde a la serie azul de la Figura 3.3. Como puede observarse, efectivamente si se considera únicamente la última década, la serie parece tener una tendencia creciente. Sin embargo, los valores observados en esta década son considerablemente más bajos que en la década anterior y la serie más bien tiene un comportamiento decreciente. Por dicha razón, se eligió la prueba de Mann-Kendall para la serie completa. Además, se realizó la prueba para detectar tendencias decrecientes. Sin embargo, esta variable no se eligió, pues no era tan discriminativa y no permitía una interpretación directa del índice de mortalidad.

Se estimaron las tendencias de las series con el estimador de Sen para la pendiente, pero se observó que en varios casos, a pesar de que la prueba de Mann-Kendall fuese estadísticamente significativa, el estimador de Sen era cero. Lo anterior es debido a que el estimador de Sen es la mediana de las diferencias en el tiempo y como se mencionó anteriormente, estas series están caracterizadas por tener muchos ceros o muchos picos. Por lo anterior, se eligió no utilizar el estimador de la pendiente sino construir una variable binaria que indica si la serie tiene una tendencia creciente o no.

La prueba de cambios en la media se realizó con base en la verosimilitud de los datos y se eligió el nivel  $\alpha = 0.1$ . Para encontrar los puntos de cambio en la media, se calcula la logverosimilitud negativa de los datos bajo los diferentes escenarios (sin puntos de cambio, con un punto de cambio, con dos puntos de cambio, etc.) y se elige el modelo bajo el cual se minimiza la logverosimilitud negativa. La idea surge de una prueba de hipótesis

no paramétrica que utiliza el cociente de verosimilitud generalizado como estadístico de prueba de no tener ningún cambio versus tener un cambio. Se identificaron tres series para Insuficiencia renal con cambio creciente en la media, una para Leucemia Linfóide y ocho para Malformaciones. Al hacer la inspección de las series seleccionadas, se observó que efectivamente las series presentaban un cambio creciente, por lo que se decidió crear las tres variables que corresponden a los promedios de la serie en cada temporada, si hay cambio y el promedio de la serie completa, si no. Además, se creó una variable indicadora de si alguno de estos cambios crecientes sucedieron en la última década del período considerado; indicando que el crecimiento es más reciente.

Se realizó la prueba de Fligner-Killeen de homocedasticidad de varianzas y posteriormente la prueba de Mann-Whitney-Wilcoxon para diferencias en la mediana entre las décadas 2000-2009 y 2010-2019, para detectar posibles cambios crecientes en las series. No se encontraron diferencias estadísticamente significativas en los grupos al 95% de confianza, por lo que no se eligió ninguna variable para la tendencia asociada a estas pruebas.

Como ya se mencionó, las series cuentan con la particularidad de tener muchos ceros, por lo que no fue posible detectar datos extremos a través de diagramas de caja de los datos anuales: cualquier dato distinto de cero, era considerado extremo. Por lo anterior, se construyeron variables de proporciones de ceros, mediana y promedios de la serie y se compararon entre sí para diversos grupos de años. El objetivo de la variable de proporciones de ceros es identificar series que requieren atención por tener baja proporción de ceros, lo cual significa que tienen pocos años en los que no hay muertos por el grupo causa. Se crearon variables de proporciones por décadas del período de análisis, ya que en el análisis exploratorio se observó que los datos extremos se capturaban mejor bajo esta agrupación. En el caso de la mediana, los datos extremos se identificaban tanto por década como con la serie completa, por lo que finalmente se eligió la serie completa. Para los promedios, se obtuvieron resultados individuales similares a la prueba de cambios en la media. No obstante, al separar por décadas se identificó un cambio grupal en la distribución de los promedios para la Insuficiencia Renal crónica en Juventud, por lo que se decidió elegir las variables de promedio por década.

Finalmente, las variables seleccionadas para la construcción del índice son:

1. Una variable binaria correspondiente a tendencia creciente o no detectada con la prueba de Mann-Kendall de un sólo lado, al 90% de confianza, sobre la serie completa.
2. Tres variables continuas: Prom1, Prom2 y Prom3, provenientes de la prueba de cambios en la media al 90% de confianza.
3. Una variable binaria, indicadora de si el cambio creciente en la media de la tasa de mortalidad, detectado en la prueba de cambio en la media, se presentó en la última década o no.
4. Una variable continua correspondiente a la mediana de la serie completa.
5. Dos variables continuas correspondientes a las proporciones de ceros por década (2000-2009 y 2010-2019).
6. Dos variables continuas correspondientes a los promedios por década.

### 3.2.2 Construcción del índice de mortalidad

El análisis de componentes principales (PCA) [2] es una técnica que permite describir un conjunto de datos en función de nuevas variables no correlacionadas llamadas componentes principales. Las componentes principales son combinaciones lineales de las variables

originales de los datos. Dicha técnica es muy utilizada para la construcción de índices debido a que la primera componente principal es la variable univariada que posee la mayor información de los datos en términos de varianza.

El objetivo del índice de mortalidad es resumir en una sola variable el comportamiento general de la mortalidad municipal, por lo que el análisis de componentes principales es una técnica adecuada para su construcción. Además, queremos que las variables que discriminan o distinguen más entre municipios sean las que tengan más peso. Si una variable es casi constante en los municipios, no sirve para discriminar ya que diríamos que la situación de mortalidad es la misma para todos. Por lo anterior, elegir las variables con mayor varianza como se hace en PCA, es adecuado.

Se realizó un análisis de componentes principales sobre las variables seleccionadas, presentadas en la sección anterior, para cada uno de los grupos de enfermedad. Con el fin de capturar el comportamiento general de los grupos de enfermedades, el análisis se realizó sobre las variables provenientes de las series sin separar por sexo ni edad, sino únicamente por Grupo de causa. Después de obtener los pesos de cada variable, dados por la primera componente principal, para cada grupo causa, se obtienen las proyecciones para las series separadas por edad y por sexo. Finalmente, los valores obtenidos se reescalan a  $[0, 1]$  calculando el máximo y el mínimo dentro de cada Grupo de causa, de forma tal que los índices son comparables entre sexos y edades dentro de cada Grupo causa.

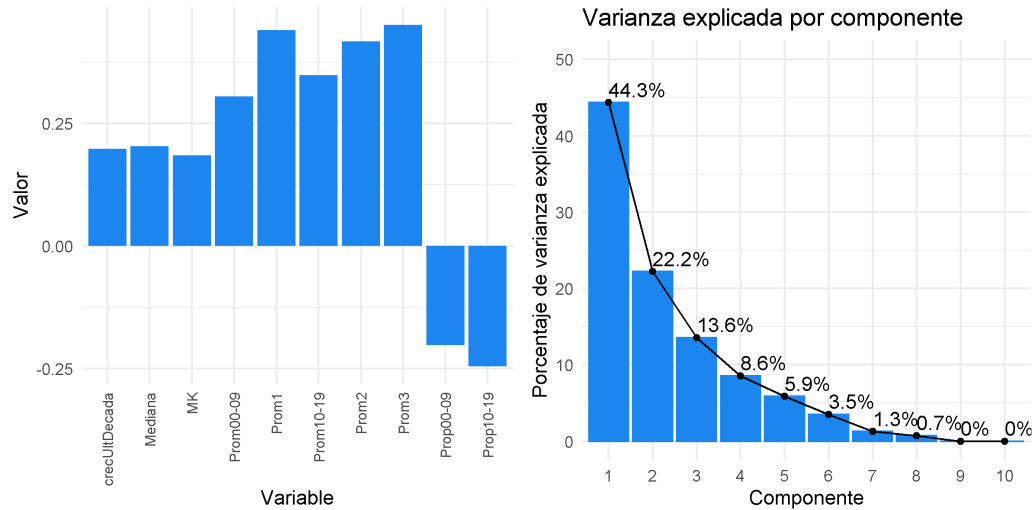
En las siguientes secciones, se presentan los resultados obtenidos mediante este análisis para cada grupo de causa.

### 3.3 Insuficiencia Renal Crónica

En la Figura 3.4, se presenta el resultado obtenido del análisis de componentes principales sobre las variables seleccionadas a partir de las series de tiempo de Adolescencia y Juventud juntos y sexo juntos, para los 67 municipios de la Cuenca.

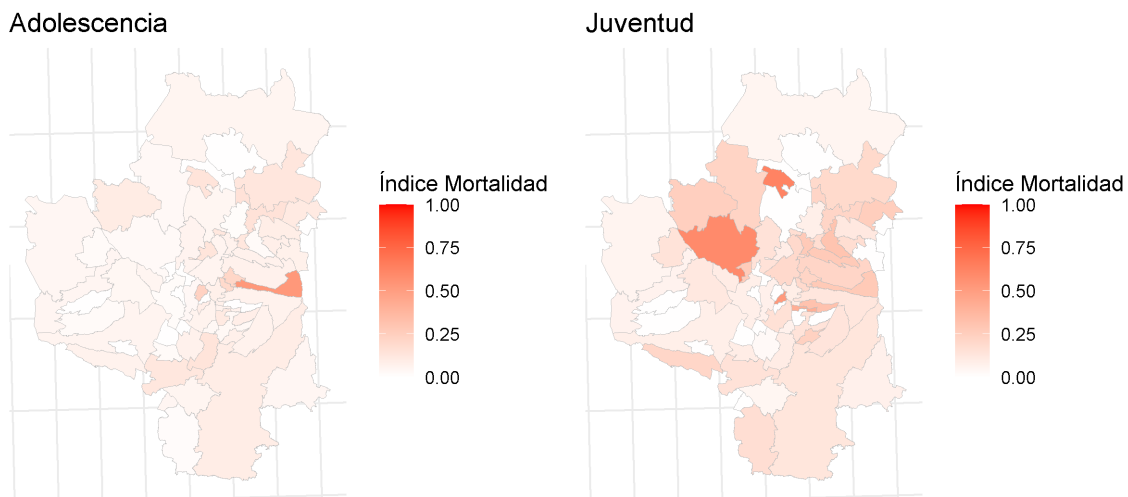
El índice de mortalidad tiene una interpretación directa: entre mayor su valor más grave la situación de mortalidad en el municipio. Lo anterior se debe a que las variables que tienen un peso positivo, representan mayor gravedad de la situación de mortalidad: tendencia creciente, promedios y mediana más grandes y cambios en la media crecientes en la última década. Por el contrario, las variables con peso negativo representan una situación menos grave de mortalidad pues entre más grande la proporción de muertos es mayor la cantidad de años en los que no hubo muertos por ese grupo de causa en ese municipio. Para este grupo de causa, las variables de promedios y de proporciones de cero, son las más importantes y la primera componente explica el 44.3% de la varianza total de los datos.

El municipio "San Lucas Tecopilco" del estado de Tlaxcala (29-055), destaca por su alto índice de mortalidad en el grupo de adolescentes, en comparación al resto de los municipios; para Juventud, destacan los municipios "San Lorenzo Axocomanitla" (29-054) e "Ixtacuixtla de Mariano Matamoros" (29-015), ambos también del estado de Tlaxcala. Los 19 municipios con índice de mortalidad más altos en Adolescencia o Juventud para esta enfermedad, pertenecen al estado de Tlaxcala.

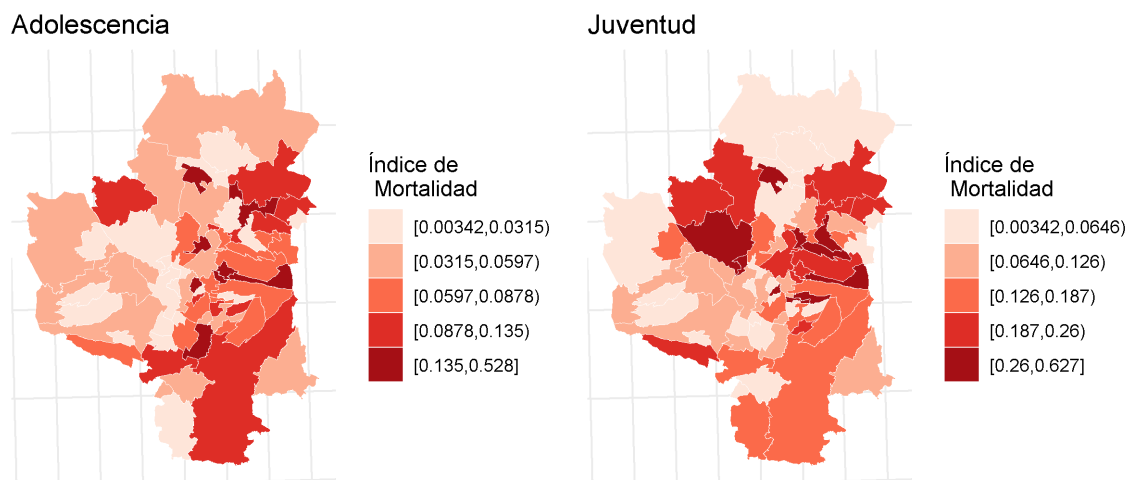


**Fig. 3.4:** Pesos (Loadings) dados por la primera componente principal y porcentaje de la varianza total, explicada por componente para las series de Insuficiencia Renal Crónica N18.

Con el fin de agrupar los municipios y visualizar de mejor manera el índice de mortalidad, se construyó una estratificación del índice utilizando el método de Dalenius y Hodges [3]. El método de Dalenius y Hodges, construye estratos de forma tal que estos sean lo más homogéneos posible: varianza mínima al interior de cada estrato y máxima entre cada uno de ellos. En la Figura 3.6, se pueden visualizar los 5 estratos construidos a partir del índice de Mortalidad por insuficiencia renal crónica en cada grupo de edad.

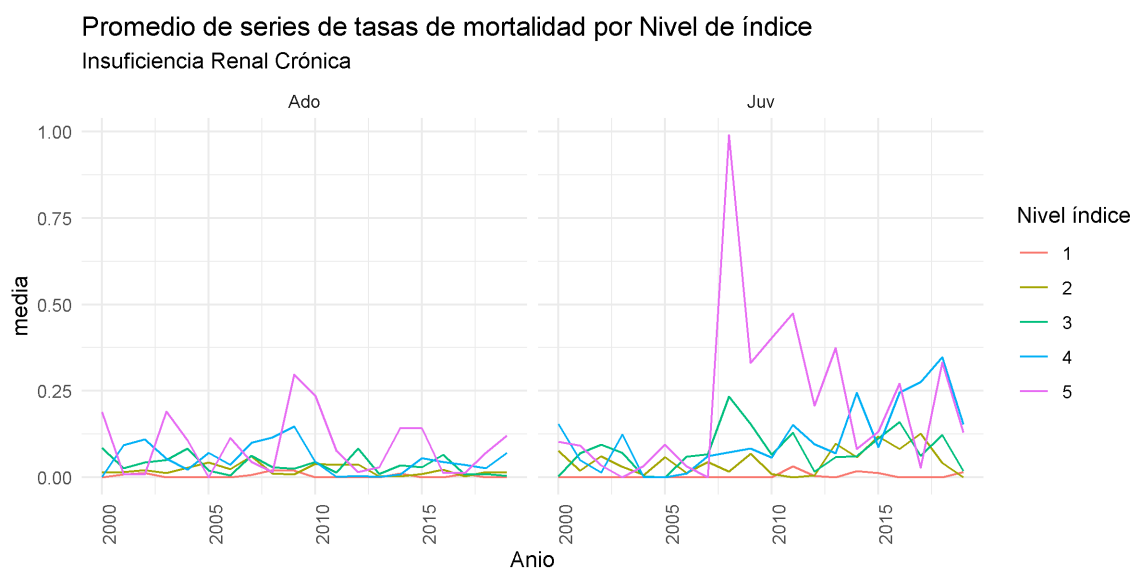


**Fig. 3.5:** Índice de Mortalidad municipal por Insuficiencia Renal Crónica (N18) para adolescentes y jóvenes.



**Fig. 3.6:** Estratos del índice de Mortalidad por Insuficiencia Renal Crónica (N18) para adolescentes y jóvenes construidos con el método de Dalenius y Hodges [3].

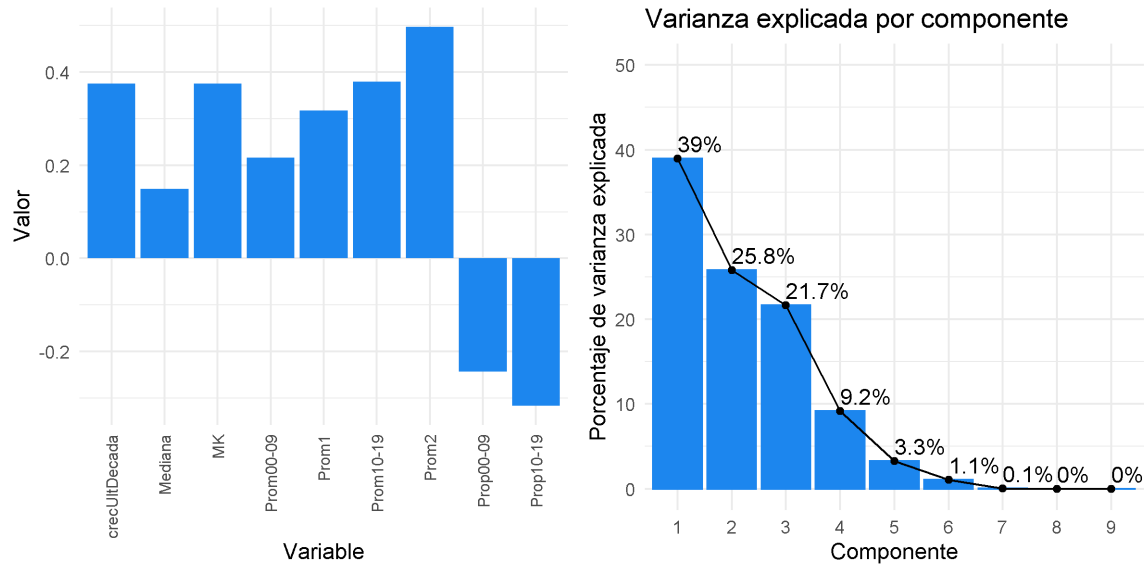
En la Figura 3.7, se presentan las series de los promedios de las tasas de mortalidad anuales para cada uno de los niveles o estratos construidos con el método de Dalenius y Hogdes [3]. Las series de niveles más altos de mortalidad, tienen promedios más altos y las series en niveles bajos tienen promedios más bajos. Lo anterior, es indicativo de que el índice de mortalidad está retratando de manera adecuada la situación general de mortalidad en los municipios.



**Fig. 3.7:** Series de los promedios anuales de tasas de mortalidad por Insuficiencia Renal Crónica (N18) para adolescentes y jóvenes para cada nivel/estrato construidos con el método de Dalenius y Hodges [3].

### 3.4 Leucemia Linfoide

En la Figura 3.8, se presenta el resultado obtenido del análisis de componentes principales sobre las variables seleccionadas a partir de las series de tiempo de Adolescencia y Juventud juntos y sexo juntos, para los 67 municipios.



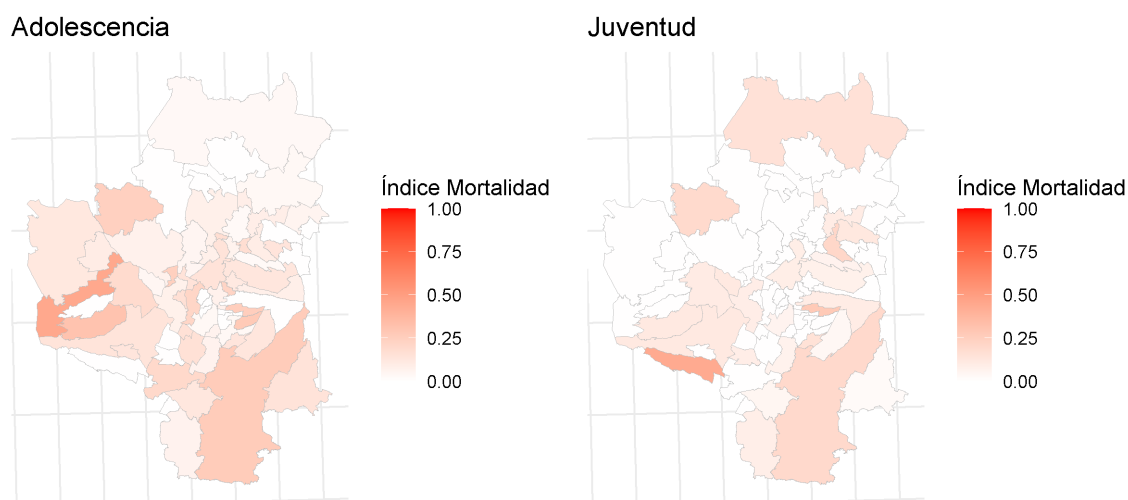
**Fig. 3.8:** Pesos (Loadings) dados por la primera componente principal y porcentaje de la varianza total, explicada por componente para las series de Leucemia Linfoide C91.

Al igual que para los otros grupos causa, el índice de mortalidad tiene una interpretación directa: entre mayor su valor más grave la situación de mortalidad en el municipio. Lo anterior se debe a que las variables que tienen un peso positivo, representan mayor gravedad de la situación de mortalidad: tendencia creciente, promedios y mediana más grandes y cambios en la media crecientes en la última década. Por el contrario, las variables con peso negativo representan una situación menos grave de mortalidad pues entre más grande la proporción de muertos es mayor la cantidad de años en los que no hubo muertos por ese grupo de causa en ese municipio. Para este grupo de causa, las variables de tendencia y de proporciones de cero, son las más importantes y la primera componente explica el 39% de la varianza total de los datos.

El municipio "San Salvador el verde" del estado de Puebla (21-143), destaca por su alto nivel de mortalidad en adolescencia, en comparación con el resto de municipios; "Calpan", en el estado de Puebla (21-026), destaca para juventud.

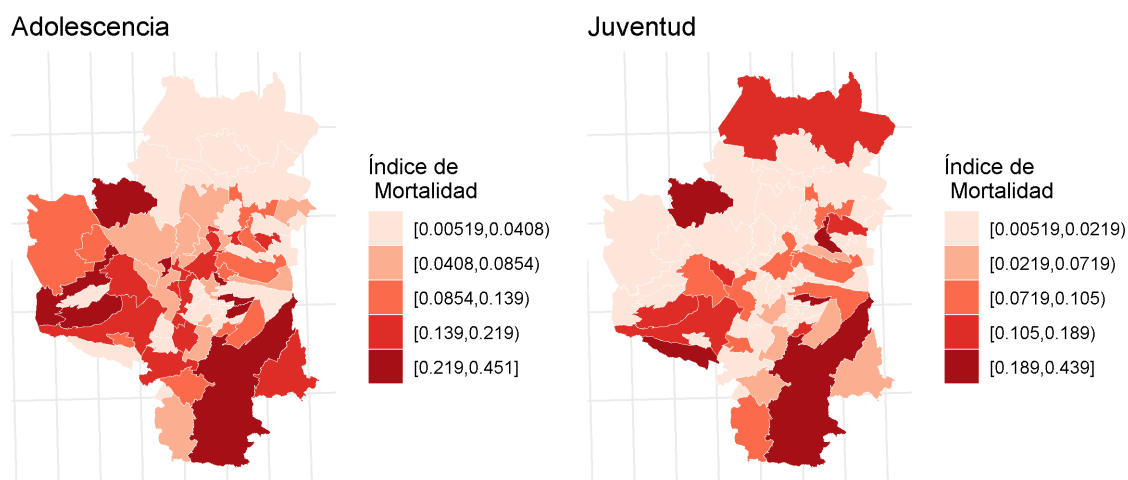
En la Figura 3.6, se pueden visualizar los 5 estratos construidos a partir del índice de Mortalidad por leucemia linfoide en cada grupo de edad.

Los municipios "Acuamanala de Miguel Hidalgo" (29-022) y "Españita" (29-012) del estado de Tlaxcala, y "Puebla" (21-114), del estado de Puebla, son los municipios que para ambas edades se encuentran en un nivel de alerta máximo en cuanto a mortalidad, en comparación al resto de los municipios de la Cuenca.



**Fig. 3.9:** Índice de Mortalidad municipal por Leucemia Linfoide (C91) para adolescentes jóvenes.

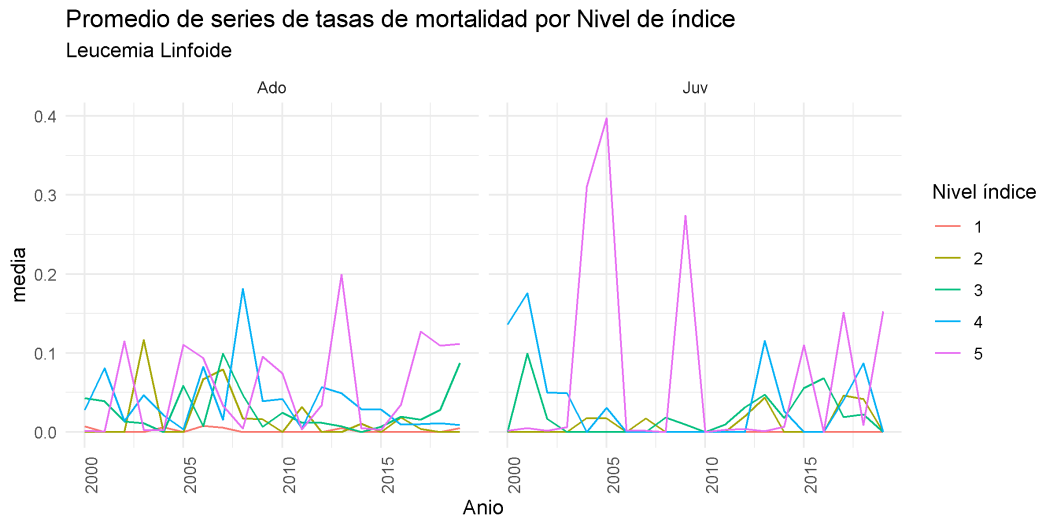
En la Figura 3.11, se presentan las series de los promedios de las tasas de mortalidad anuales para cada uno de los niveles o estratos construidos con el método de Dalenius y Hogdes. Las series de niveles más altos de mortalidad, tienen promedios más altos y las series en niveles bajos tienen promedios más bajos. Lo anterior, es indicativo de



**Fig. 3.10:** Estratos del índice de Mortalidad por Leucemia Linfoide (C91) para adolescentes y jóvenes construidos con el método de Dalenius y Hodges [3].



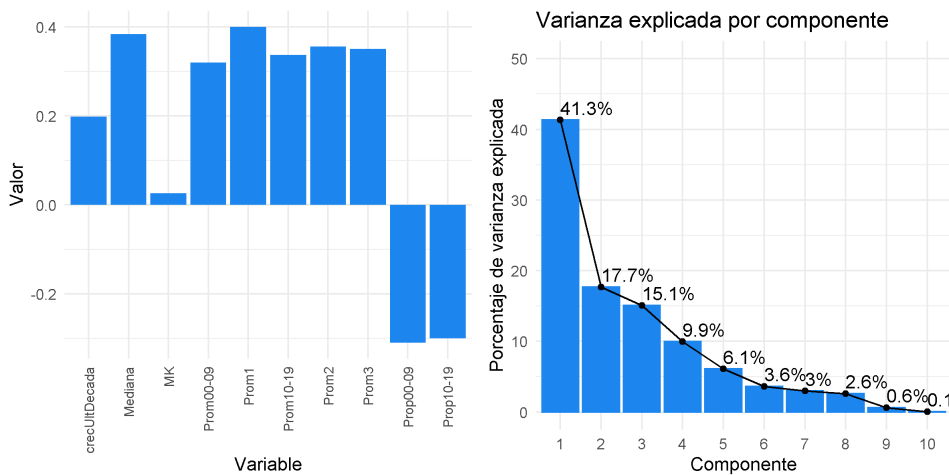
que el índice de mortalidad está retratando de manera adecuada la situación general de mortalidad en los municipios.



**Fig. 3.11:** Series de los promedios anuales de tasas de mortalidad por Leucemia Linfóide (C91) para adolescentes y jóvenes para cada nivel/estrato contruidos con el método de Dalenius y Hodges [3].

### 3.5 Malformaciones

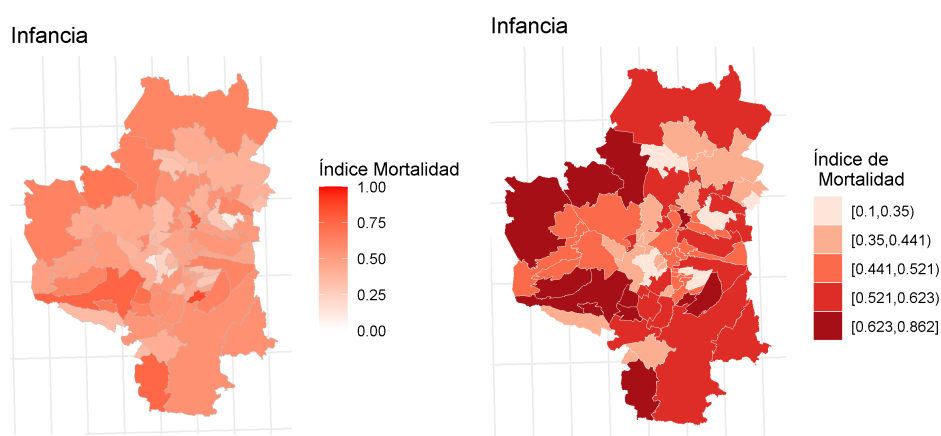
En la Figura 3.12, se presenta el resultado obtenido del análisis de componentes principales sobre las variables seleccionadas a partir de las series de tiempo de Infancia y sexo juntos, para los 67 municipios.



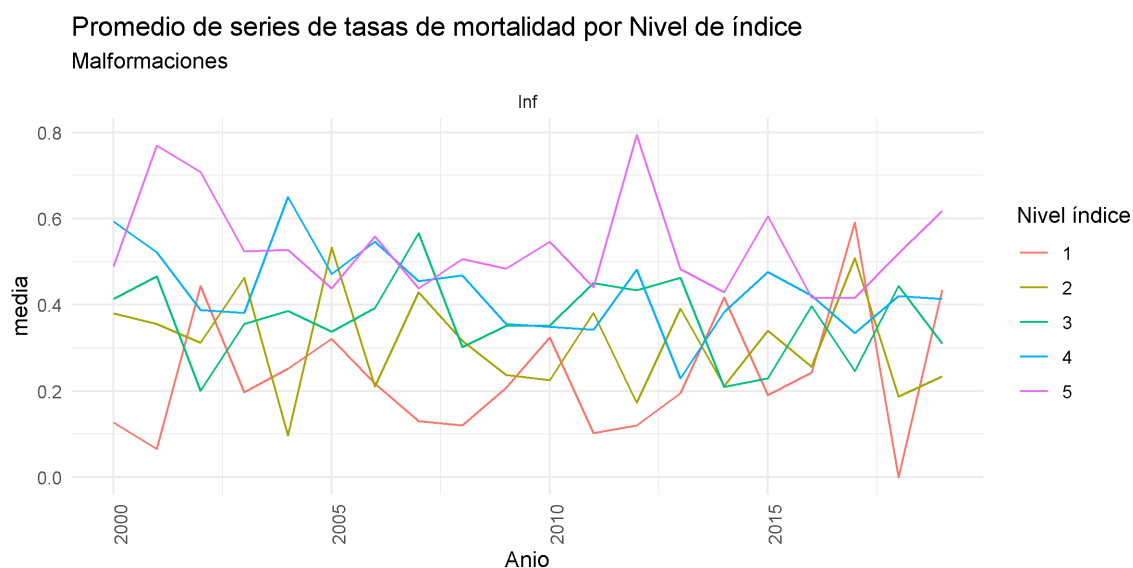
**Fig. 3.12:** Pesos (Loadings) dados por la primera componente principal y porcentaje de la varianza total explicada por componente para las series de Malformaciones.

Al igual que para los otros grupos causa, el índice de mortalidad tiene una interpretación directa: entre mayor su valor más grave la situación de mortalidad en el municipio. Lo anterior se debe a que las variables que tienen un peso positivo, representan mayor

gravedad de la situación de mortalidad: tendencia creciente, promedios y mediana más grandes y cambios en la media crecientes en la última década. Por el contrario, las variables con peso negativo representan una situación menos grave de mortalidad pues entre más grande la proporción de muertos es mayor la cantidad de años en los que no hubo muertos por ese grupo de causa en ese municipio. Para este grupo de causa, la mediana y los promedios, son las variables más importantes, y la primera componente explica el 41.3% de la varianza total de los datos. En la Figura 3.13, se puede visualizar el índice de Mortalidad por Malformaciones en la infancia y los 5 estratos construidos con el método de Dalenius y Hodges[3].



**Fig. 3.13:** Izq.: índice de Mortalidad por el grupo de Malformaciones (Q00-99) en infancia. Der.: estratos del índice de mortalidad por el Grupo de Malformaciones (Q00-99) en la infancia, construido con el método de Dalenius y Hodges [3].



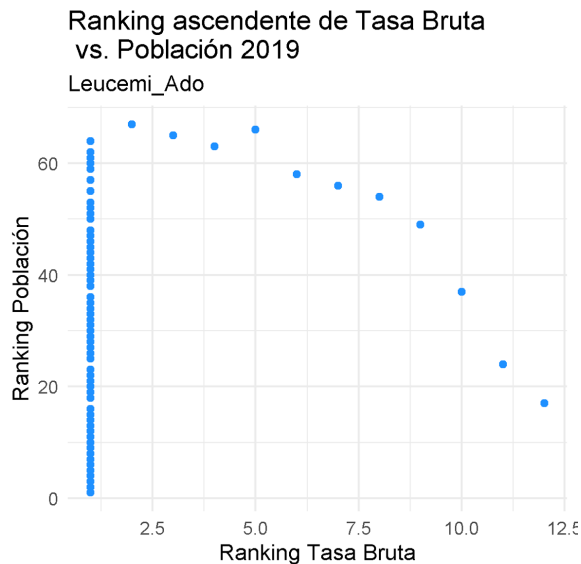
**Fig. 3.14:** Series de los promedios anuales de tasas de mortalidad por el grupo de Malformaciones para adolescentes y jóvenes para cada nivel/estrato construidos con el método de Dalenius y Hodges [3].

Los 5 municipios con nivel de índice de mortalidad por Malformaciones en la infancia más alto en comparación al resto son: Tenancingo (29-027) y Apetatitlán de Antonio Carvajal (29-002) del estado de Tlaxcala, y Huejotzingo (21-074), Ocoyucan (21-106) y Tlaltenango (21-181) de estado de Puebla.

En la Figura 3.14, se presentan las series de los promedios de las tasas de mortalidad anuales para cada uno de los niveles o estratos construidos con el método de Dalenius y Hogdes. Las series de niveles más altos de mortalidad, tienen promedios más altos y las series en niveles bajos tienen promedios más bajos. Lo anterior, es indicativo de que el índice de mortalidad está retratando de manera adecuada la situación general de mortalidad en los municipios.

### 3.6 Efecto de las Poblaciones Pequeñas en el Índice de Mortalidad

Durante el Taller: "Homologando metodologías para la comparación de tasas de mortalidad en poblaciones pequeñas", organizado por la Dirección de Vinculación y Enlace Nacional e Internacional del CONACYT, se discutió sobre el efecto de estimar tasas de mortalidad con poblaciones pequeñas. En epidemiología se acostumbra publicar tasas de mortalidad calculadas sobre 10 defunciones o más. Dado que en los municipios de la Cuenca, para los grupos causa de estudio, se tienen pocas defunciones, analizar la estimación de tasas para poblaciones pequeñas resulta relevante. En la Figura 3.15, se muestra el ranking de los municipios según la tasa bruta de mortalidad y el tamaño de población municipal para el año 2019. Los rankings están dados de menor a mayor, es decir, entre más pequeño el ranking, la población o la tasa bruta de ese municipio es más pequeña y entre más grande el ranking, el tamaño de población o tasa bruta es más grande.

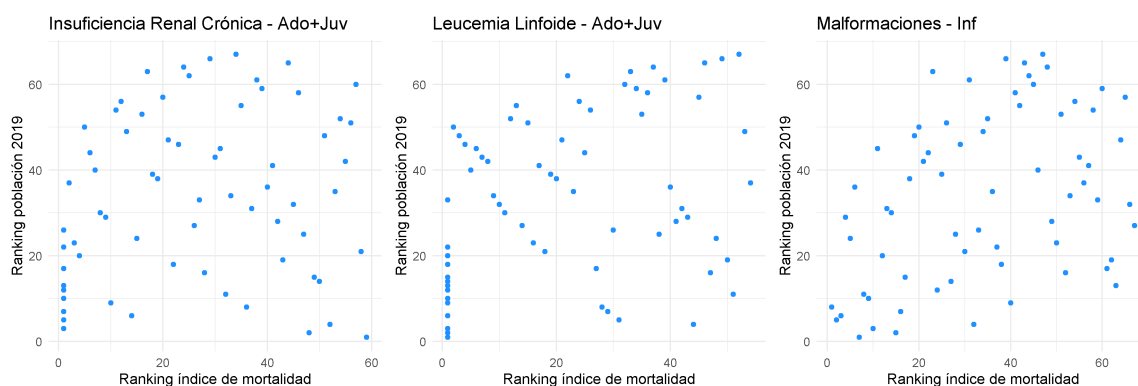


**Fig. 3.15:** Ranking de los municipios de acuerdo a su tamaño de población en 2019 (de menor a mayor población) vs. ranking de los municipios de acuerdo a su tasas brutas de mortalidad de Leucemia Linfocítica en adolescentes (de menor a mayor tasa).

Se observa que los municipios que no tuvieron defunciones en 2019, se encuentran repartidos en los diferentes tamaños de población. Es decir, no se observa un sesgo a que no se presenten defunciones en municipios con tamaño de población grande o con tamaño de población pequeña. Sin embargo, sobre los municipios que presentaron defunciones, se

observa una tendencia a que los municipios con menor población tengan una mayor tasa de mortalidad. En el taller, se discutió la hipótesis de que el efecto anterior, puede deberse al tamaño de la población.

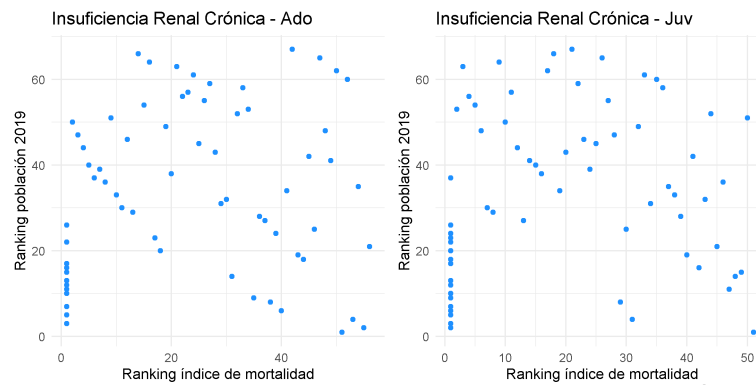
Una forma usual de evitar el efecto de poblaciones pequeñas es calculando las tasas a un mayor nivel de agregación: por región geográfica, por grupos de años, por enfermedades, etc. A pesar de que el índice de mortalidad está construido a partir de las tasas municipales anuales de mortalidad, las variables utilizadas en el análisis de componentes principales consideran varios tipos de agregación de dichas tasas. Recuérdese que dado que las variables se construyen a partir de la serie, el objeto de estudio es en cierta forma un agregado. Los promedios y proporciones se construyeron por década; mientras que la mediana, la tendencia, puntos de cambio y la indicadora de si la serie es creciente en la última década se construyeron con información de los 20 años. Por lo anterior, como se muestra en la Figura 3.16, el índice de mortalidad no presenta un efecto por poblaciones pequeñas como en el caso de las tasas brutas.



**Fig. 3.16:** Ranking de los municipios de acuerdo a su tamaño de población en 2019 (de menor a mayor población) vs. ranking de los municipios de acuerdo a su índice de mortalidad (de menor a mayor). Izq: Insuficiencia Renal Crónica (Adolescencia + Juventud). Centro: Leucemia Linfoide (Adolescencia + Juventud). Der.: Grupo de malformaciones (Infancia).

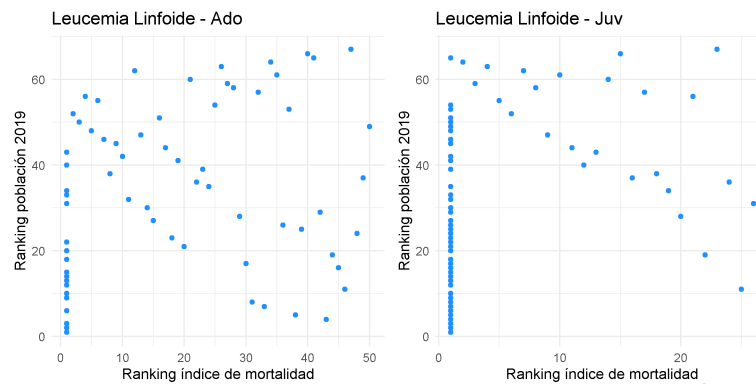
Para las tres enfermedades o grupos de causa, se puede observar que los municipios con índices más altos están repartidos a través de los diferentes tamaños de población, es decir, el índice de mortalidad no presenta un efecto de estimación dado por el tamaño de la población.

Recuérdese que el índice de mortalidad se construye con los grupos agregados como aparecen en la Figura 3.16, pero también es posible obtener los índice desagregados por sexo y edad utilizando los pesos o loadings obtenidos en el análisis de componentes principales. A continuación, se muestra la comparación de los índices desagregados por grupo de edad con la población municipal en 2019.



**Fig. 3.17:** Ranking de los municipios de acuerdo a su tamaño de población en 2019 (de menor a mayor población) vs. ranking de los municipios de acuerdo a su índice de mortalidad por insuficiencia renal crónica (de menor a mayor). Izq.: Adolescencia. Der.: Juventud.

En la Figura 3.17, se puede observar que los índices más altos están razonablemente distribuidos entre los diferentes tamaños de población. Sin embargo, se empieza a observar una leve tendencia a tener índices grandes en poblaciones pequeñas. Se puede considerar aceptable este nivel de desagregación para el índice de mortalidad de insuficiencia renal crónica, puesto que si uno observa los diez índices más altos, por ejemplo, puede darse cuenta que estos municipios tienen muy diferente tamaño de población entre sí.



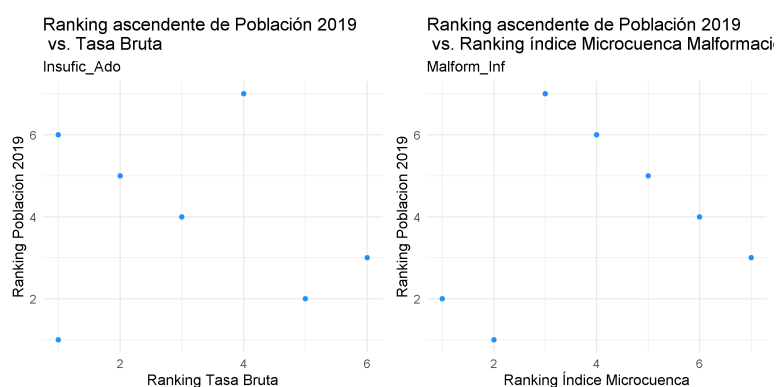
**Fig. 3.18:** Ranking de los municipios de acuerdo a su tamaño de población en 2019 (de menor a mayor población) vs. ranking de los municipios de acuerdo a su índice de mortalidad por leucemia linfóide (de menor a mayor tasa). Izq.: Adolescencia. Der.: Juventud.

En la Figura 3.18, se puede observar que los índices más altos están razonable distribuidos entre los diferentes tamaños de población. No obstante, como en el caso de insuficiencia renal crónica, se observa una leve tendencia a tener índices grandes en poblaciones pequeñas. Se puede considerar aceptable este nivel de desagregación para el índice de mortalidad de leucemia linfóide, puesto que si uno observa los diez índices más altos, por ejemplo, puede darse cuenta que estos municipios tienen muy diferente tamaño de población entre sí.

Si a la desagregación del índice por edad se le añade una desagregación por sexo, entonces sí se puede observar un patrón a tener índices altos en poblaciones pequeñas. No

es obvio que este efecto se deba únicamente al tamaño de la población, como se mostrará más adelante. Sin embargo, se sugiere manejar con precaución este nivel de desagregación para el índice.

El tamaño de la población puede no ser la única razón de la tendencia de las altas tasas o altos índices de mortalidad con dos niveles de desagregación hacia poblaciones pequeñas, puesto que se desconoce si existen variables correlacionadas con el tamaño de la población que puedan estar influyendo en la mortalidad por estas enfermedades. Aún más, se realizó el experimento de agrupar los datos a nivel microcuenca, para verificar si al tener más casos dicho efecto desaparecería. En el caso de Malformaciones, al agrupar de esta manera el 84% de los datos cumplen con el requisito de diez decesos o más sugerido en la Literatura. Para dicho grupo de causa, se realizó el análisis de componentes principales con las mismas variables que se eligieron para el índice municipal y el ranking obtenido de esta forma se presenta en la Figura 3.19 (Der.).



**Fig. 3.19:** Ranking de los municipios de acuerdo a su tamaño de población en 2019 (de menor a mayor población) vs. Izq.: ranking de los municipios de acuerdo a su tasa de mortalidad por Insuficiencia Renal Crónica (de menor a mayor tasa). Der.: ranking de los municipios de acuerdo a su índice de mortalidad por Malformaciones (de menor a mayor)

En la Figura 3.19 puede observarse que aún a este nivel de agregación, existe una tendencia de tasas o índice alto hacia poblaciones pequeñas. Sabemos que el comportamiento anterior no es debido a la construcción del índice, puesto que a nivel municipal con hasta un nivel de desagregación (sexo o edad), no presenta este comportamiento. Por lo anterior, no es claro que esta tendencia hacia poblaciones pequeñas se deba exclusivamente al tamaño de la población; pueden existir otras variables correlacionadas con el tamaño de la población que también influyen de alguna manera en la mortalidad.



## Referencias

- [1] "World health organization. icd-10 : international statistical classification of diseases and related health problems : tenth revision, 2nd edition," *World Health Organization*, 2004.
- [2] R. Maronna, "Alan julian izenman (2008): Modern multivariate statistical techniques: Regression, classification and manifold learning," *Statistical Papers*, vol. 52, pp. 733–734, 08 2011.
- [3] T. Dalenius and J. L. H. Jr, "Minimum variance stratification," *Journal of the American Statistical Association*, vol. 54, no. 285, pp. 88–101, 1959.