Comparación de variables en poblaciones pediátricas y no pediátricas[[1]](#footnote-1)

Esteban Paduro[[2]](#footnote-2)

Junio 2024

**Resumen:** En el presente documento se estudian las diferencias cuantitativas en los perfiles de demanda observados en los datos atenciones de urgencias respiratorias contenidas en la base de datos del Departamento de Estadísticas e Información de Salud (DEIS). Se concluye que existen diferencias significativas en el comportamiento de las curvas de demanda correspondientes a poblaciones pediátricas y no pediátricas. Adicionalmente para poblaciones no pediátricas se observan marcadas diferencias dadas por la ubicación geográfica, entre las zonas más extremas del país versus las zonas más centrales.

# Introducción

El problema de la estimación de la demanda hospitalaria, particularmente la asociada a problemas respiratorios corresponde a un importante desafía que deben enfrentar los diferentes centros de salud pues afecta la asignación de manera eficiente de recursos. Es en este contexto que la pregunta del nivel de agrupación apropiado para el análisis de datos toma relevancia. Por un lado, el tener más datos permite realizar mejores análisis estadísticos pues amortigua los efectos asociados a la incerteza, sin embargo, el agrupar datos que posean comportamientos significativamente diferentes puede dificultar la extracción de buenas conclusiones a partir de las observaciones.

La confección de este informe se enmarca en el proceso de generar una plataforma de predicciones de la fecha y la magnitud de la demanda máxima (peak) de las curvas de demanda de hospitalizaciones respiratorias. En la primera etapa dicho proyecto se enfoca en estudiar poblaciones pediátricas, ya que análisis preliminares sugieren que su comportamiento es cualitativamente diferente al de poblaciones no pediátricas. Dichas observaciones se respaldan por estudios que indican que algunos parámetros como la tasa de reinfección de ciertas enfermedades respiratorias varían de manera significativa entre grupos etario[[3]](#footnote-3). En el presente informe exploraremos esta pregunta poniendo énfasis a las particularidades de la realidad chilena en base a la información de atenciones de urgencia en establecimientos de salud pública disponible de la base de datos DEIS6, y considerando no solo diferencias por asociadas a la distribución etaria, sino además posibles diferencias asociadas a distribución geográfica[[4]](#footnote-4) y tipo de centro de salud.

## Objetivos del proyecto

El objetivo del presente documento es realizar una comparación entre las características de las curvas de demanda de atenciones de urgencia asociada a Infecciones Respiratorias Agudas (IRA) correspondientes a poblaciones pediátricas y no pediátricas. Los objetivos específicos son los siguientes:

* Decidir si es apropiado analizar de manera agrupado grupos poblacionales correspondientes a diferentes grupos etarios.
* Identificar variables que expliquen las diferencias observadas en el comportamiento del peak de atenciones de urgencias respiratorias.

## Hipótesis para el análisis

Para guiar nuestro análisis estableceremos las siguientes hipótesis de investigación.

Hipótesis 1: Existen diferencias significativas entre las poblaciones pediátricas y no pediátricas en términos de demanda de atenciones hospitalarias de urgencia.

Algunas observaciones preliminares que motivas este análisis son las siguientes:

* El peak de demanda hospitalaria en las poblaciones menores de un año y mayores a 65 años parecen ser menos pronunciados que en los otros grupos.
* La composición del tipo de atenciones de urgencia asociada varías al considerar distintos grupos etarios. Es necesario un análisis detallado para poder discriminar entre estos grupos determinar niveles de agrupación apropiada para la predicción de demanda de atención hospitalaria.

Hipótesis 2: Es posible estudiar el perfil de demanda hospitalaria porcentual de manera independiente de la estimación del total de atenciones anuales de cierto grupo de población. (Aún no sé como enfrentar esta pregunta, quizás lo saque)

## Descripción de los datos

Nuestra fuente de información corresponde las principales causas de atenciones de urgencia, de manera agrupada, realizadas dentro de los establecimientos públicos, pertenecientes al Sistema Nacional de Servicios de Salud (SNSS)[[5]](#footnote-5), las cuales se reportaron al sistema DEIS [[6]](#footnote-6). Las causas de atenciones de urgencia consideradas corresponden a 6 categorías descritas en los datos del ministerio de salud detalladas en la Tabla 1.

|  |  |
| --- | --- |
| **IdCausa** | **Glosa Causa** |
| 3 | Bronquitis/bronquiolitis aguda (J20-J21) |
| 4 | Influenza (J09-J11) |
| 5 | Neumonía (J12-J18) |
| 6 | Otra causa respiratoria (J22, J30-J39, J47, J6... |
| 10 | IRA Alta (J00-J06) |
| 11 | Crisis obstructiva bronquial (J40-J46) |

Tabla 1 Lista de causas respiratorias según definición MINSAL 3

Los datos se encuentran desagregados a nivel diario con datos desde el 2017 hasta el 2024, sin embargo, en este estudio nos enfocaremos en los datos hasta 2023. Adicionalmente daremos un trato especial a los datos correspondiente al periodo 2020-2021 por encontrarse fuertemente afectado por la pandemia COVID-19.

En términos de grupos etarios, los datos están desagregados en 5 categorías

* Menores de 1 año
* Entre 1 y 4 años
* Entre 5 y 14 años
* Entre 15 y 64 años
* Mayores de 65 años

En términos geográficos, se disponen datos de centros de salud de las 16 regiones del país, además de contar información desagregada a nivel de servicio de salud (Ver apéndice 5.1) los cuales nos permiten considerar desagregaciones a nivel de zonas geográficas. Para determinados análisis nos restringiremos a centro de salud tipo hospitales o ciertas agrupaciones geográficas como todos los servicios de salud de la región metropolitana.

# Análisis de Niveles de Agrupación

El objetivo de esta sección es estudiar el nivel de agrupación apropiado para estudiar los correspondientes a la demanda de atenciones de urgencias respiratorias. La sección se organiza de la siguiente manera. Primero, en la Sección 2.1 se considerarán algunos análisis preliminares que nos permitirán identificar diferencias importantes que guiarán nuestro análisis más detallado. Segundo en la Sección 2.2 se estudiará la agrupación por grupo etario, y se explicará cómo mediante la aplicación de una técnica de reducción de dimensionalidad y métodos de clasificación binaria es posible cuantificar que grupos pediátricos tienen comportamientos significativamente diferentes de los no pediátricos. Luego, en la Sección 2.3 se considerará la dependencia geográfica de los datos, para este propósito consideraremos la división del país en seis zonas y analizaremos si algunas de ellas tienen comportamiento similar. Este análisis se hará de forma separada para grupos pediátricos y no pediátricos. Finalmente, en la Sección 2.4 se estudiará la dependencia del tipo de centro de salud en los perfiles de demanda de atenciones de urgencia respiratorias.

## Análisis preliminar

En objetivo de esta sección es recopilar algunos resultados que permitan guiar nuestro análisis. Nuestro primer análisis será el considerar el total de demanda diaria de atenciones obtenido al agrupar todas las atenciones de urgencias respiratorias a nivel nacional.

En nuestro análisis llamaremos “perfil de demanda” correspondiente a un cierto grupo de población a una lista de diaria de las atenciones a lo largo de un año, divididas por el total de atenciones, de esta manera el número correspondiente a cada día representa el porcentaje de atenciones anuales que se realizaron dicha fecha. La utilidad del perfil de demanda es que permite comparar poblaciones de diferente tamaño.

En la Ilustración 1, se muestra el total de atenciones diarias de urgencia en que se han agrupado datos de establecimientos públicos de todo el país separado a nivel de día y para cada año. Notamos que el comportamiento de los años 2020 y 2021 es muy diferente al resto de los años. Adicionalmente es posible identificar que hay un alza en la demanda de atenciones alrededor del día 150, el cual varía a lo largo de los años, y además es posible identificar un alza de la demanda, con menor magnitud durante la segunda mitad del año. Notamos que incluso a este nivel de agrupación es posible observar diferencias importantes en cuanto a la fecha del peak de atenciones de urgencia.

El perfil demanda de atenciones correspondiente a diferentes años exhibe un comportamiento estacional, más precisamente la demanda correspondiente a atenciones respiratorias de urgencia presentan un máximo durante los meses más fríos de año. Es importantes para la planificación y asignación de recursos el tener buenas estimaciones de la fecha en que ocurre el peak de demanda de atenciones y la magnitud de dicho peak. La fecha precisa correspondiente al máximo puede presentar variaciones significativas de un año a otro.

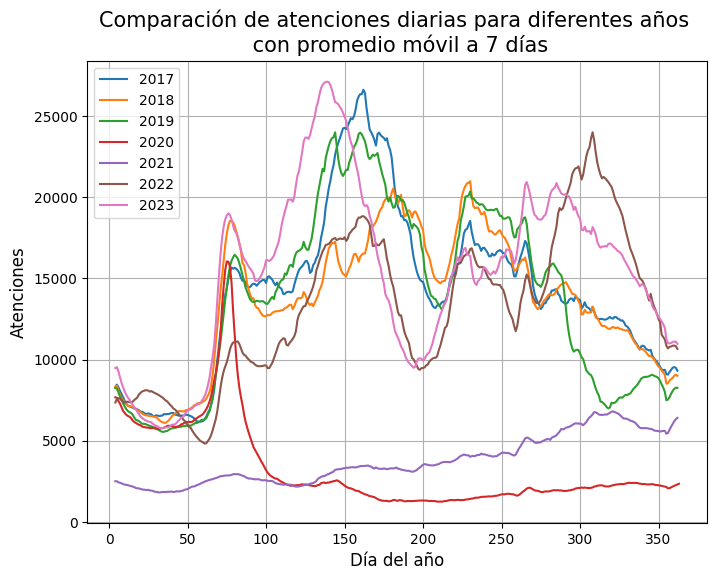


Ilustración 1 Media móvil de las atenciones diarias de urgencia correspondientes al periodo 2017 -2023.

La Ilustración 1 también sugiere que es insuficiente el estimar la fecha del peak solo con información de años anteriores, pues puede presentar variaciones significativas de un año a otro, entre sectores geográficos4 o diferentes grupos etarios. Debido a los motivos antes mencionados resulta relevante estudiar con cuidado la pregunta de la agrupación apropiada de los datos de la demanda de atenciones.

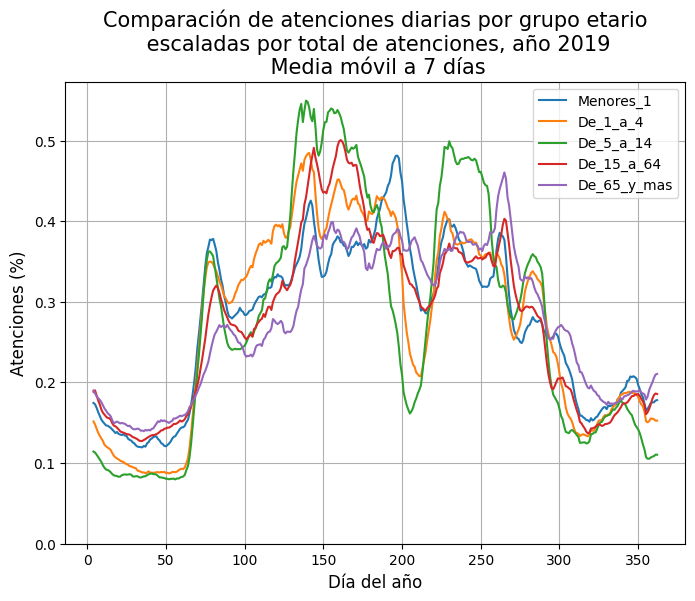


Ilustración 2 Perfil de demanda Año 2019 suavizado a 7 días y desagregado para diferentes grupos etarios.

De manera similar en la Ilustración 2 consideramos la comparación de las curvas de demanda de atenciones correspondientes a diferentes grupos etarios. Esta vez consideramos las atenciones correspondientes a todos los establecimientos de salud pública del país, pero escalamos las curvas usando su total anual, de manera que las atenciones diarias se presentan como un porcentaje de las atenciones anuales. Como ya identificamos que existen variaciones importantes de año a año, decidimos fijar el año (2019 en este caso) y comparamos las curvas de demanda. A partir de esta Ilustración es posible identificar que los datos para menores de un año y para mayores de 65 presentan un comportamiento con peaks menos marcados, pero para realizar observaciones más precisar necesitaremos herramientas más elaboradas que desarrollaremos en la siguiente sección.

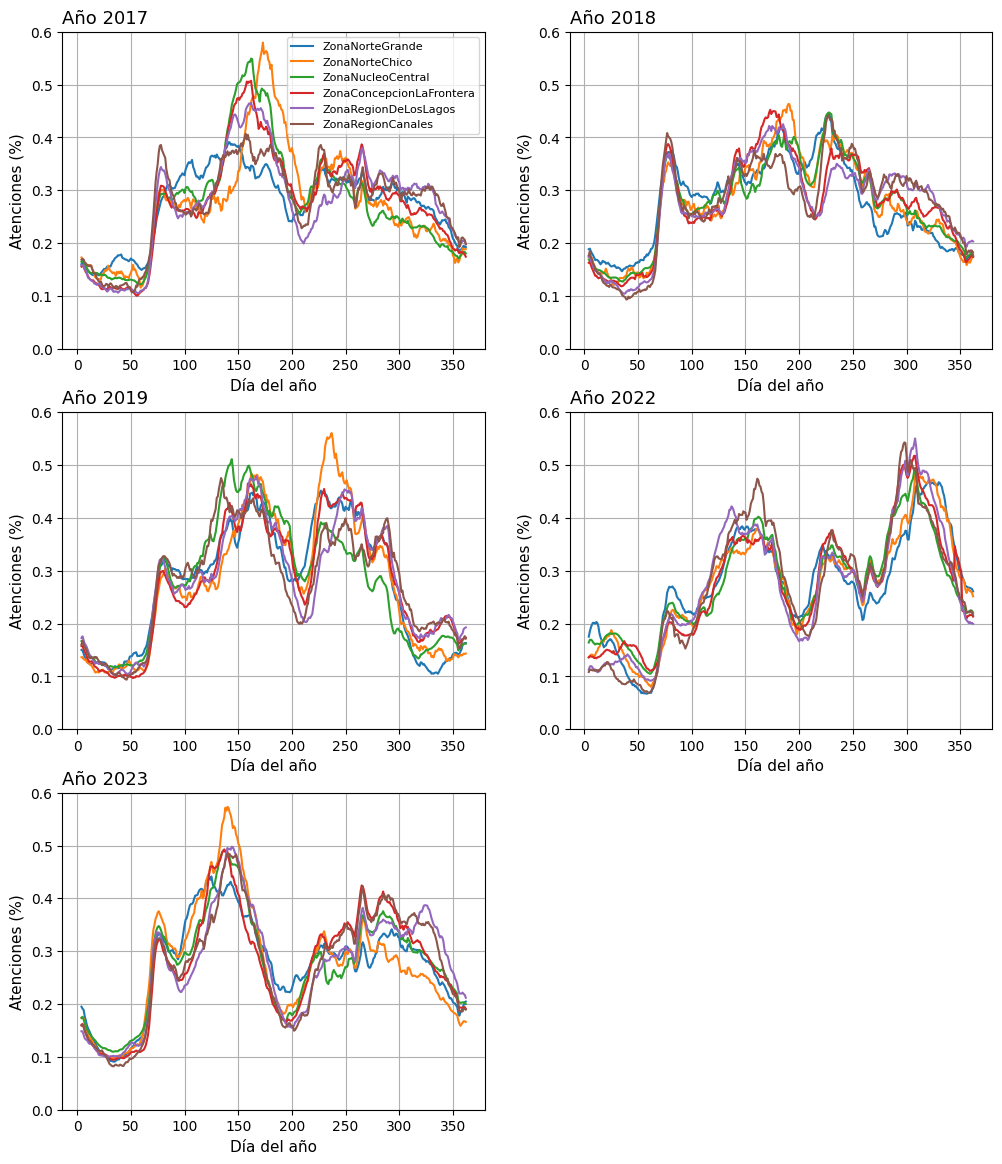


Ilustración 3 Atenciones (%) diarias agrupadas por zona geográfica para cada día del año, Años 2017,2018,2019,2022,2023

El siguiente factor que queremos analizar es la dependencia geográfica de las demandas, para esto separaremos nuestros datos en 6 zonas geográficas identificadas por los correspondientes servicios de salud asociados (Ver apéndice 5.1). La ilustración 3 muestra el comportamiento del perfil de demanda de las atenciones de urgencia, escaladas para mostrar las atenciones diarias como un porcentaje de las atenciones anuales. Esta visualización nos permite identificar diferencias cualitativas entre grupos poblacionales de diferente tamaño.

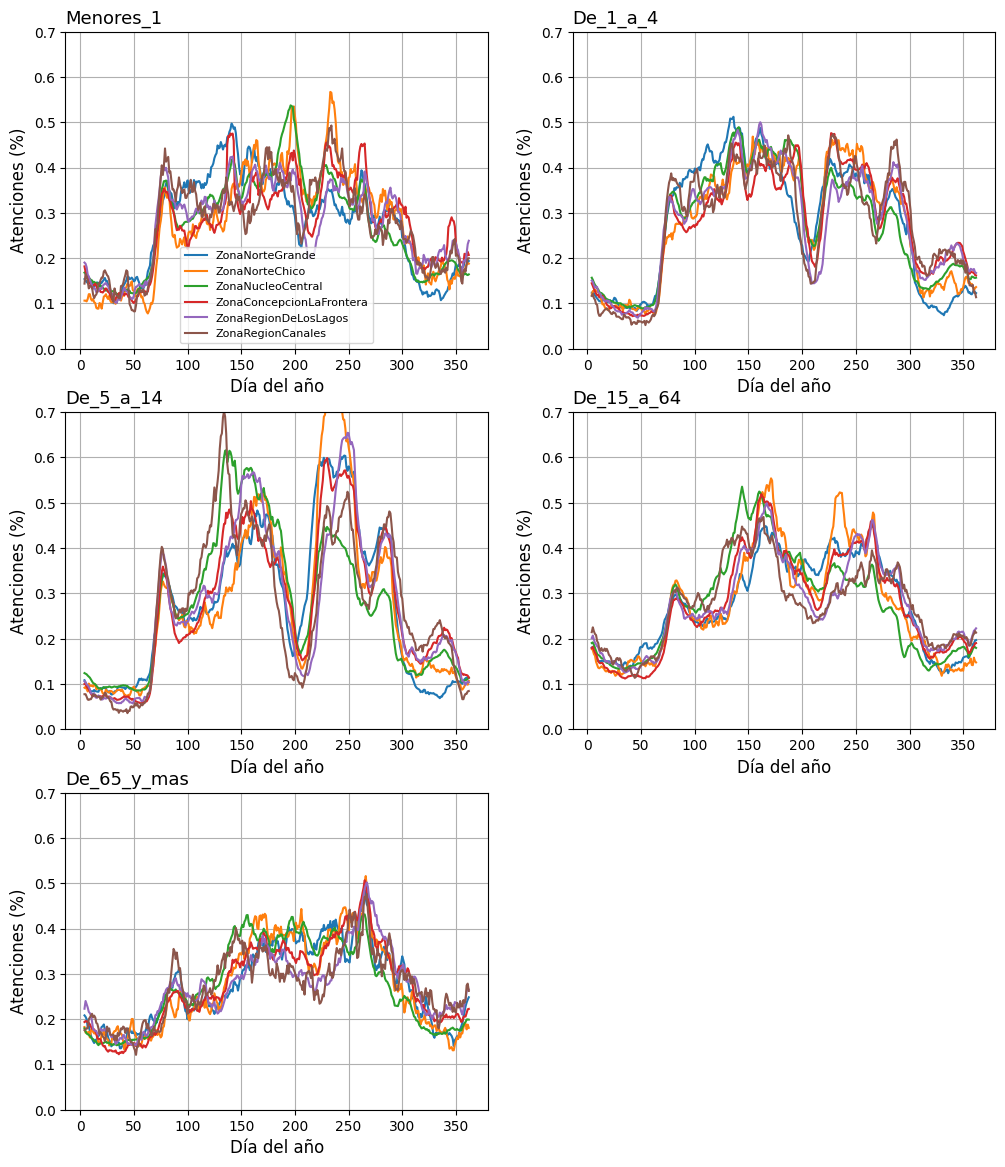


Ilustración 4 atenciones diarias como porcentaje de las atenciones anuales, correspondientes al año 2019, separado por grupos etarios y zonas geográficas

Adicionalmente nos podemos preguntar cómo se relacionan las variables de grupo etario y zona geográfica para un año fijo, la Ilustración 4 muestra los perfiles de demanda correspondientes al año 2019. A partir de la figura es posible notar importantes diferencias en la forma del perfil de demanda para el grupo etario De\_5\_a\_14 y un desfase de las curvas correspondientes al grupo etario De\_15\_a\_64.

## Diferencias por Grupo Etario

Una de las preguntas claves del análisis es el poder decidir cuál es el nivel de agrupación apropiado de los datos. Diferentes niveles de agrupación permiten estudiar diferentes aspectos de los datos. Con el objetivo de evaluar nuestra hipótesis de análisis que grupos pediátricos y no pediátricos deben ser estudiados de manera diferenciada, propondremos medidas que permitan identificar que permitan identificar diferencias significativas entre sus perfiles de demanda. Para este propósito consideremos el siguiente procedimiento para preparar la información: consideraremos series de datos correspondientes a datos diarios asociados a un cierto año y características asociadas a los grupos que deseamos comparar.

Para el primer análisis agruparemos las atenciones de urgencias respiratorias agrupadas por año, región y grupo etario. De esta forma, por ejemplo, asociadas al triple año 2022, Región 5 y grupo etario De\_5\_a\_14 le corresponderá una lista de atenciones diarias que se indizará por el número de día del año (365 variables). Considerando la influencia de factores externos que explican diferencias significativas entre años, es conveniente estudiarlos de manera desagregada.

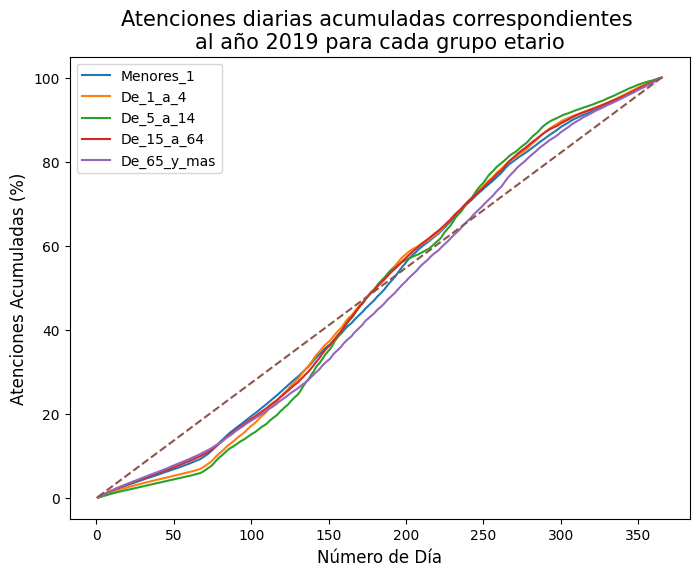


Ilustración 5: Atenciones acumuladas año 2019, por grupo etario, V región. En rojo se indica la fecha correspondiente al peal de atenciones de urgencias respiratorias para niños de 1 a 4 años.

Una vez que se tienen las atenciones diarias correspondientes a cada uno de estos triples (año, región, grupo etario), notamos que los datos tienen un comportamiento bastante ruidoso. Con el objetivo de disminuir los efectos de las variaciones diarias, consideraremos la cantidad de atenciones acumuladas a cierto día del año. Adicionalmente para poder comparar la forma del perfil de demanda de atenciones, escalaremos la curva de tal manera que se represente el porcentaje de las atenciones anuales que han ocurrido hasta cierta fecha. Puede ser útil notar que esta forma de analizar los datos contiene la misma información que una media móvil con una ventana de tres días.

A partir de esta curva Ilustración 5 es posible observar diferencias cualitativas entre los perfiles de demanda de atenciones correspondientes a diferentes grupos etarios. Un perfil más cercano a la diagonal, como lo es del grupo etario De\_65\_y\_mas representa que las atenciones se distribuyen de manera más uniformes a lo largo del año. Esta forma de analizar los datos preserva la habilidad de estudiar la época del año en que ocurre la mayor demanda de atenciones, lo cual se puede hacer fijando una ventana temporal (digamos una semana) y buscar el rango de 5 días consecutivos en que se produjo la mayor demanda de atenciones.

Para evaluar nuestra hipótesis de similitud o diferencias de comportamientos de diferentes grupos poblacionales utilizaremos una técnica de reducción de dimensionalidad conocida como análisis de componentes principales (PCA)[[7]](#footnote-7).

Para realizar el PCA, consideraremos que cada observación corresponde a un vector con largo la cantidad de días del año (365). Las atenciones correspondientes a un mismo día son estandarizadas a lo largo de todas las observaciones para tener media cero y varianza 1. Las componentes principales corresponden a vectores propios asociados a los valores propios más grandes de la matriz de covarianza. Para decidir cuantas componentes usar usaremos el concepto de la proporción de la varianza explicada por cada una de las componentes principales. Esta se obtiene como el cociente entre el valor propio correspondiente y la varianza total del conjunto de datos.

Para nuestro primer análisis consideraremos la parte de los datos que tenemos a disposición. Se considerarán los datos de los años 2017, 2018, 2019, 2022, 2023, para todas las regiones. La agrupación se realizará a nivel de servicio de salud (29 en total). Se considerarán por separado las atenciones correspondientes a cada uno de los grupos etarios definidos en la introducción, adicionalmente nos restringiremos a las atenciones diarias por urgencias respiratorias correspondientes a las 6 causas consideradas en la Sección Descripción de los datos. Usando la proporción de varianza explicada que se muestra en la Ilustración 6 decidimos que para nuestro análisis consideraremos solo las 3 primeras componentes principales, pues ellas explican el 89% de la varianza.

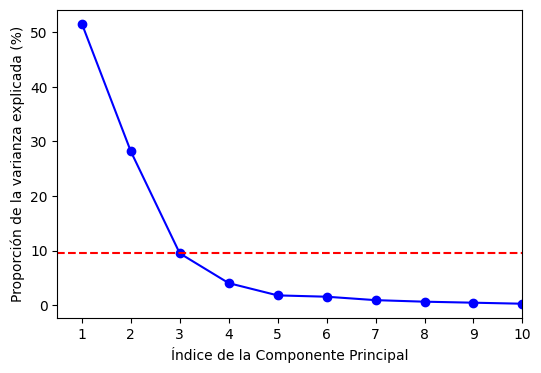


Ilustración 6 Porcentaje de la varianza explicada por cada componente en el análisis de componentes principales.

Durante el resto de este trabajo siempre usaremos las 3 primeras componentes principales para describir nuestros datos.

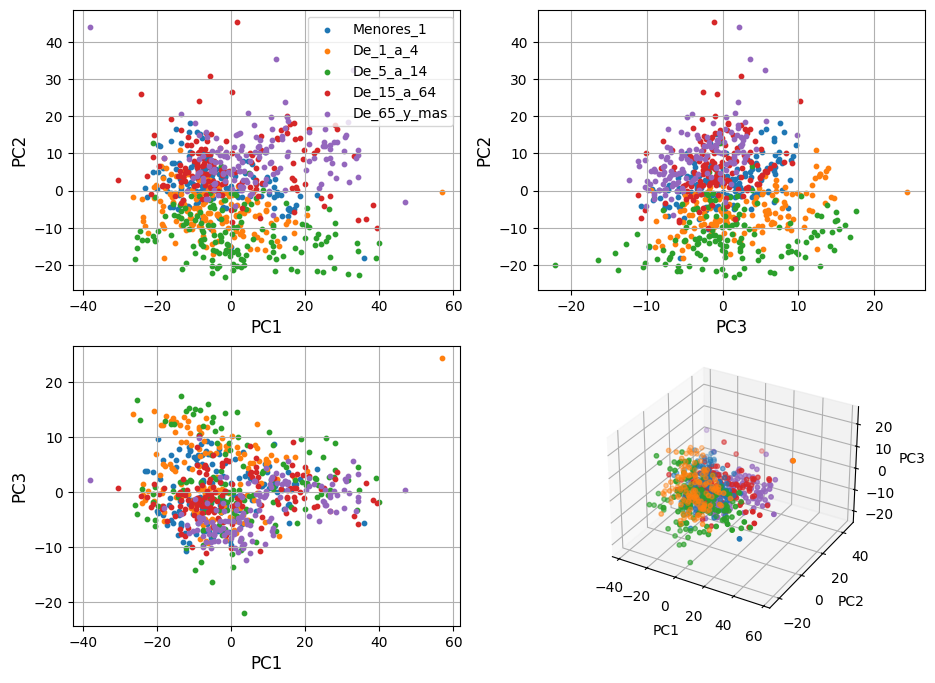


Ilustración 7 Componentes principales asociados a los datos de urgencias respiratorias agrupados por año, servicio de salud, grupo etario.

La Ilustración 7 nos permite identificar que existe un comportamiento marcadamente diferente entre la agrupación De\_1\_a\_4 y De\_5\_a\_14 vs el resto de los individuos, también a partir de las gráficas notamos que el comportamiento del grupo Menores\_1 no parece poder separarse de las demás clases.

Para formalizar estas observaciones aplicaremos una técnica de Machine Learning diseñada para clasificación conocida como Support Vector Machine (SVM). De manera más específica utilizaremos un clasificador lineal regularizado, que se resuelve utilizando la técnica de aprendizaje no supervisado de Stochastic Gradient Descent[[8]](#footnote-8). Con esta técnica buscaremos un plano en el subespacio descrito por las componentes principales (PC1, PC2, PC3) que permita separa los elementos de una clase vs los elementos fuera de ella. La idea es que, si somos capaces de hallar un plano que separe las clases un alto nivel de precisión, entonces es una indicación razonable que los conjuntos de datos tienen comportamientos cuantitativamente diferentes, y por tanto hay razones para estudiarlos de forma.

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Ilustración 8 Diagrama de la clasificación Binaria y el correspondiente plano que separa las clases usando las componentes principales usadas en la Ilustración 1. En este caso se indican en rojo el grupo formado al agrupar las clases De\_1\_a\_4 y De\_5\_a\_14.

Para el ejemplo anterior, consideraremos como una clase a la agrupación De\_1\_a\_4 y De\_5\_a\_14, y trataremos de separarla del resto. En este caso se obtiene una precisión del clasificador[[9]](#footnote-9) del 87% (promedio sobre 1000 repeticiones), lo cual sugiere que sugiere una separación bastante efectiva de este grupo versus el resto de los grupos etarios. El plano que separa las clases se muestra en la Ilustración 8.

Aun cuando la agrupación obtenida es bastante efectiva, no es claro que tal separación sea la mejor de entre todas las posibles. Por este motivo consideraremos un procedimiento más exhaustivo. Como tenemos cinco grupos etarios consideraremos todas las posibles clasificaciones en dos clases entre ellos. Debido a la simetría del problema, existen solo dos casos posibles, separar uno de los otros cuatro y separar dos de los otros tres. De esta forma la Tabla 2 indica en la fila i, columna j el porcentaje de clasificación correcta cuando se intenta separar el conjunto (i,j) del resto, además los elementos en la diagonal representan que solo separamos una clase de las otras cuatro.

Los resultados de la clasificación son presentados en la Tabla 2. En ella se puede observar que dentro de nuestros existen subclases que tienen comportamientos significativamente diferentes. Los grupos identificados nos dicen que las observaciones correspondientes a grupos pediátricos (De\_1\_a\_4, De\_5\_a\_14) o (Menores\_1, De\_1\_a\_4, De\_5\_a\_14) tienen comportamientos cuantitativamente diferentes a los correspondientes de a no pediátricos, y por lo tanto es apropiado estudiarlos de forma separa con respecto a el análisis de las curvas de demanda.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Menores\_1 | De\_1\_a\_4 | De\_5\_a\_14 | De\_15\_a\_64 | De\_65\_y\_mas |
| Menores\_1 | 0,69439 |  |  |  |  |
| De\_1\_a\_4 | 0,59142 | 0,73553 |  |  |  |
| De\_5\_a\_14 | 0,60337 | **0,86971** | 0,81717 |  |  |
| De\_15\_a\_64 | 0,60044 | 0,54125 | 0,53636 | 0,72374 |  |
| De\_65\_y\_mas | 0,65523 | 0,53515 | 0,58135 | **0,82211** | **0,84009** |

Tabla 2 Precisión de la clasificación como promedio de 1000 repeticiones. Observamos que los tres mejores resultados corresponden a separar De\_1\_a\_4 y De\_5\_a\_14 del resto, separar De\_65\_y\_mas del resto y finalmente separar De\_15\_a\_64 con De\_65\_y\_mas del resto

## Influencia de la Agrupación Geográfica

Siguiendo en análisis de la sección anterior esta vez estudiaremos la pregunta de cuál es la mejor segregación según zona geográfica. Motivados por los resultados de la sección anterior consideraremos separar nuestro análisis entre Población pediátrica (Menores\_1, De\_1\_a\_4, De\_5\_a\_14) vs Población no pediátrica (De\_15\_a\_64, De\_65\_y\_mas).

Para el análisis consideraremos 3 años de manera simultánea (2017,2018,2019) pues la Ilustración 3 sugiere comportamiento parecido. En primera instancia miraremos los datos correspondientes a todos los grupos etarios y luego separaremos entre pediátricos y no pediátricos. Para estudiar cuan propicio es agrupar los datos utilizaremos la misma estrategia de PCA y SVM considerada en la sección anterior. Para estudiar diferentes zonas geográficas consideraremos la agrupación descrita en el Apéndice 5.1. Usaremos las siguientes abreviaciones para referirnos a las zonas geográficas.

1. Zona Norte Grande (NG)
2. Zona Norte Chico (NC)
3. Zona Núcleo Central (C)
4. Zona Concepción y la Frontera (CF)
5. Zona Región de los Lagos (RL)
6. Zona Región de los Canales (RC)

Utilizando estas zonas trataremos de determinar si es apropiada agruparlas en el contexto de estudio de la curva de demanda de atenciones de urgencia. Basados en los resultados de la sección anterior, sabemos que es apropiado separa el análisis de grupos pediátricos y no pediátricos.

Al considerar solamente poblaciones pediátricas se obtienen los resultados que se muestran en la Tabla 3. Observamos que en este caso el clasificador sugiere que la agrupación por zonas inicialmente considerada es una medida razonable para discriminar comportamientos de las curvas de demanda (clasificación sobre el 75% de precisión al separa una zona de todas las demás). Además de entre los casos que efectivamente agrupan zonas, notamos que las que parecen ser más apropiadas son las que corresponden a asociar las zonas más extremas del territorio, es decir: (NG, C) vs (NC, CF, RL, RC); (NG, NC, C) vs (CF, RL, RC) y (NG, NC, C, CF) vs (RL, RC).

Tabla 3 Resultados de las clasificaciones agrupando zonas geográficas considerando solo gropos poblacionales pediátricos correspondientes a los años 2017, 2018 y 2019

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Agrupación | Precisión del clasificador |  | Agrupación | Precisión del clasificador |  | Agrupación | Precisión del clasificador |
| (NG) | **0,78930** |  | (NC, CF) | 0,65977 |  | (NG, NC, RC) | 0,54860 |
| (NC) | **0,75791** |  | (NC, RL) | 0,53256 |  | (NG, C, CF) | 0,55465 |
| (C) | **0,79302** |  | (NC, RC) | 0,58698 |  | (NG, C, RL) | 0,66953 |
| (CF) | **0,76535** |  | (C, CF) | 0,59070 |  | (NG, C, RC) | 0,57535 |
| (RL) | **0,78767** |  | (C, RL) | 0,61628 |  | (NG, CF, RL) | 0,55488 |
| (RC) | **0,81651** |  | (C, RC) | 0,61279 |  | (NG, CF, RC) | 0,51860 |
| (NG, NC) | 0,64465 |  | (CF, RL) | 0,64047 |  | (NG, RL, RC) | 0,67581 |
| (NG, C) | **0,74140** |  | (CF, RC) | 0,69674 |  |  |  |
| (NG, CF) | 0,57558 |  | (RL, RC) | **0,73349** |  |  |  |
| (NG, RL) | 0,70163 |  | (NG, NC, C) | **0,74209** |  |  |  |
| (NG, RC) | 0,61744 |  | (NG, NC, CF) | 0,57744 |  |  |  |
| (NC, C) | 0,66047 |  | (NG, NC, RL) | 0,54953 |  |  |  |

Al considerar las poblaciones no pediátricas (ver Tabla 4) notamos que nuevamente la separación por zonas se presenta como una segregación apropiada para el análisis de curvas de demanda, sin embargo, para estos grupos etarios adultos los resultados sugieren que la separación más efectiva se logra al considerar las agrupaciones (NG, RL, RC) vs (NC, C, CF), es decir las zonas más extremas versus las centrales.

Tabla 4 Análisis de dependencia geográfica, considerando solo grupos etarios correspondientes a población no pediátrica, años 2017, 2018 y 2019

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Agrupación | Precisión del clasificador |  | Agrupación | Precisión del clasificador |  | Agrupación | Precisión del clasificador |
| (NG) | **0,83071** |  | (NC, CF) | 0,72607 |  | (NG, NC, RC) | 0,64929 |
| (NC) | **0,86286** |  | (NC, RL) | 0,62321 |  | (NG, C, CF) | 0,58750 |
| (C) | **0,81714** |  | (NC, RC) | 0,58321 |  | (NG, C, RL) | 0,53250 |
| (CF) | **0,80500** |  | (C, CF) | 0,73607 |  | (NG, C, RC) | 0,64357 |
| (RL) | **0,85393** |  | (C, RL) | 0,61357 |  | (NG, CF, RL) | 0,59679 |
| (RC) | **0,82821** |  | (C, RC) | 0,64786 |  | (NG, CF, RC) | 0,55500 |
| (NG, NC) | **0,80393** |  | (CF, RL) | 0,75643 |  | (NG, RL, RC) | **0,90500** |
| (NG, C) | 0,66857 |  | (CF, RC) | 0,64929 |  |  |  |
| (NG, CF) | 0,55929 |  | (RL, RC) | **0,84500** |  |  |  |
| (NG, RL) | 0,71321 |  | (NG, NC, C) | **0,85357** |  |  |  |
| (NG, RC) | 0,78607 |  | (NG, NC, CF) | 0,63750 |  |  |  |
| (NC, C) | **0,82643** |  | (NG, NC, RL) | 0,66536 |  |  |  |

## Tipo de Centro de Salud

En objetivo de esta sección es estudiar si existen diferencias entre las curvas de demanda asociadas a centro distintos tipos de centros de salud, es de particular interés investigar si existen diferencias significativas entre hospitales vs no hospitales.

# Análisis específico para la Región Metropolitana

En esta sección nos enfocaremos en características más específicas asociadas a la población de la región metropolitana. Ésta esta organizada en 5 servicios de salud que coordinan ciertos aspectos administrativos asociados a sus correspondientes centros de salud. Un aspecto que vale la pena considerar a este nivel de agregación son los posibles desfases en el peak de atenciones de urgencia que se producen en las diferentes zonas de la región para poblaciones pediátricas y no pediátricas.

## Agrupación por Tipo de Centro de Salud

# Disponibilidad de información

Los datos utilizados fueron obtenidos a partir de la base de datos DEIS. Para el análisis los datos correspondientes a los años 2017-2023 fueron consolidados de una tabla con la siguiente estructura: IdEstablecimiento, NEstablecimiento, IdCausa, GlosaCausa, Total, Menores\_1, De\_1\_a\_4, De\_5\_a\_14, De\_15\_a\_64, De\_65\_y\_mas, fecha, semana, GLOSATIPOESTABLECIMIENTO, GLOSATIPOATENCION, GlosaTipoCampana, CodigoRegion, NombreRegion, CodigoDependencia, NombreDependencia, CodigoComuna, NombreComuna.

Detalles de la construcción de las diferentes ilustraciones y análisis están disponibles en un repositorio GitHub[[10]](#footnote-10)

# Apéndice

## Lista de Servicios de Salud

|  |  |  |
| --- | --- | --- |
| CodigoDependencia | NombreDependencia | Zona Geográfica |
| 1 | Arica | Zona norte grande |
| 2 | Iquique | Zona norte grande |
| 3 | Antofagasta | Zona norte grande |
| 4 | Atacama | Zona Norte Chico |
| 5 | Coquimbo | Zona Norte Chico |
| 6 | Valparaíso San Antonio | Zona Núcleo Central |
| 7 | Viña Del Mar Quillota | Zona Núcleo Central |
| 8 | Aconcagua | Zona Núcleo Central |
| 9 | Metropolitano Norte | Zona Núcleo Central |
| 10 | Metropolitano Occidente | Zona Núcleo Central |
| 11 | Metropolitano Central | Zona Núcleo Central |
| 12 | Metropolitano Oriente | Zona Núcleo Central |
| 13 | Metropolitano Sur | Zona Núcleo Central |
| 14 | Metropolitano Suroriente | Zona Núcleo Central |
| 15 | Libertador B. O'Higgins | Zona Núcleo Central |
| 16 | Del Maule | Zona Núcleo Central |
| 17 | Ñuble | Zona Núcleo Central |
| 18 | Concepción | Zona Concepción y La frontera |
| 19 | Talcahuano | Zona Concepción y La frontera |
| 20 | Bíobío | Zona Concepción y La frontera |
| 21 | Araucanía Sur | Zona Concepción y La frontera |
| 22 | Valdivia | Zona Región de los lagos |
| 23 | Osorno | Zona Región de los lagos |
| 24 | Del Reloncaví | Zona Región de los lagos |
| 25 | Aisén | Zona Región de los Canales |
| 26 | Magallanes | Zona Región de los Canales |
| 28 | Arauco | Zona Concepción y La frontera |
| 29 | Araucanía Norte | Zona Concepción y La frontera |
| 33 | Chiloé | Zona Región de los Canales |

1. Zona norte grande: Arica y Parinacota, Tarapacá y Antofagasta. Regiones 15, 1, 2. Servicios de salud con ID: 1, 2, 3
2. Zona norte chico: Atacama y Coquimbo (3,4). Regiones 3,4. Servicios de Salud con ID: 4, 5
3. Zona Núcleo Central: Valparaíso, Metropolitana de Santiago, O'Higgins, Maule y Ñuble. Regiones 5,13,6,7,16. Servicios de Salud con ID: 6,7,8,9,10,11,12,13,14,15, 16, 17.
4. Zona Concepción y La Frontera: Biobío y La Araucanía. Regiones 8,9. Servicios de Salud ID: 18,19,20,21, 28,29,
5. Zona Región de los Lagos: Los Ríos y Los Lagos (continental) . Regiones 14,10. Servicios de Salud ID: 22,23,24.
6. Zona Región de los Canales: Los Lagos (Provincia de Chiloé), Aysén y Magallanes. Regiones 10, 11,12. Servicios de Salud ID: 25, 26, 33.

1. CMM – Línea de Salud Digital, proyecto de estimación de demanda IRA, XXXX. [↑](#footnote-ref-1)
2. Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS IRL2807), Universidad de Chile, Santiago, Chile [↑](#footnote-ref-2)
3. A. S. Monto and S. K. Lim, “THE TECUMSEH STUDY OF RESPIRATORY ILLNESS,” *American Journal of Epidemiology*, vol. 94, no. 3, pp. 290–301, Sep. 1971, doi: [10.1093/oxfordjournals.aje.a121322](https://doi.org/10.1093/oxfordjournals.aje.a121322). [↑](#footnote-ref-3)
4. T. T. Lam et al., “Comparative global epidemiology of influenza, respiratory syncytial and parainfluenza viruses, 2010–2015,” Journal of Infection, vol. 79, no. 4, pp. 373–382, Oct. 2019, doi: 10.1016/j.jinf.2019.07.008. [↑](#footnote-ref-4)
5. Manual de Registro Sistema de Atención Diaria de Urgencia, Departamento de Estadística e Información de Salud (DEIS), Enero 2021 [↑](#footnote-ref-5)
6. Para estudiar esta disyuntiva es relevante identificar cuáles son las características principales que se desea rescatar de las observaciones <https://datos.gob.cl/dataset/atenciones-de-urgencia> [↑](#footnote-ref-6)
7. Jolliffe, I. T. (2002). Principal Component Analysis. Springer Series in Statistics. New York: Springer-Verlag. doi:10.1007/b98835. ISBN 978-0-387-95442-4 [↑](#footnote-ref-7)
8. https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.SGDClassifier.html [↑](#footnote-ref-8)
9. Para evaluar la precisión del clasificador este se calcula usando el 80% de los datos y se evalúa con el restante 20%. [↑](#footnote-ref-9)
10. https://github.com/estebanpaduro/CMM-SD\_comparacion\_poblaciones\_IRA [↑](#footnote-ref-10)