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. Con respecto a la agrupación geográfica se concluye que la agrupación por “Regiones de la Corfo” parece set apropiada para separar grupos comportamientos cualitativamente diferentes con respecto a atenciones de urgencias respiratorias. Adicionalmente para poblaciones no pediátricas (15 años) se observan marcadas diferencias dadas por la ubicación geográfica, entre las zonas más extremas del país versus las zonas más centrales. Con respecto a la dependencia por tipo de centro de salud, se pudo identificar diferencias que los centros de atención rurales (SUR y CEAR) tienen un perfil de demanda muy diferente a otros centros, pero para establecimientos del tipo Hospital, SAPU o SAR, sus curvas de demanda son bastante similares.

Los análisis específicos correspondientes a la región metropolitana revelan que las separaciones en poblaciones pediátricas y no pediátricas, y la separación por centro de salud hospital vs no hospital revela diferencias importantes entre la demanda de atenciones dependiendo de la zona geográfica (representada por la variable servicio de salud). A este nivel de agrupación se evidencia que el peak de demandas de atenciones de urgencia puede presentar variaciones una o más semanas, dependiendo de la ubicación geográfica y variaciones de más de un mes al considerar datos de diferentes años.

# 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 etarios[[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.

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>

## 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.

## 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 5

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 para la mayoría de nuestros análisis ignoraremos los datos correspondientes 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, de manera más específica consideraremos la agrupación geográfica de la CORFO 1950. 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.

Adicionalmente para el caso particular de la región metropolitana consideraremos los datos de hospitalizaciones correspondientes a causas respiratorias, las cuales5 se encuentran registradas en la columna “CAUSAS SISTEMA RESPIRATORIO”.

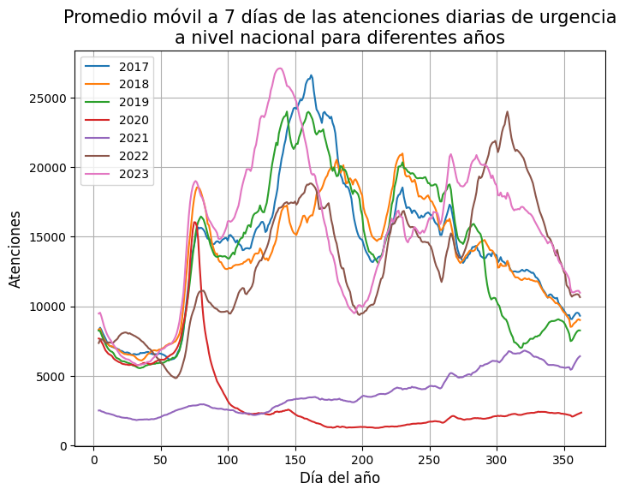
# Análisis de Niveles de Agrupación

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.

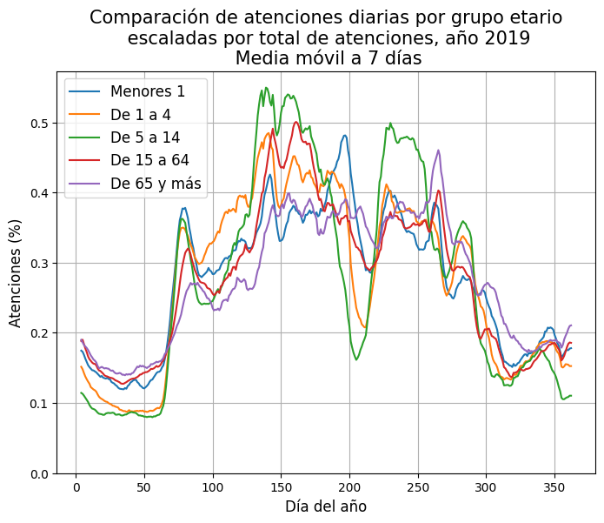


**Ilustración 1** Promedio móvil a 7 días de las atenciones diarias de urgencia   
correspondientes al periodo 2017 -2023.

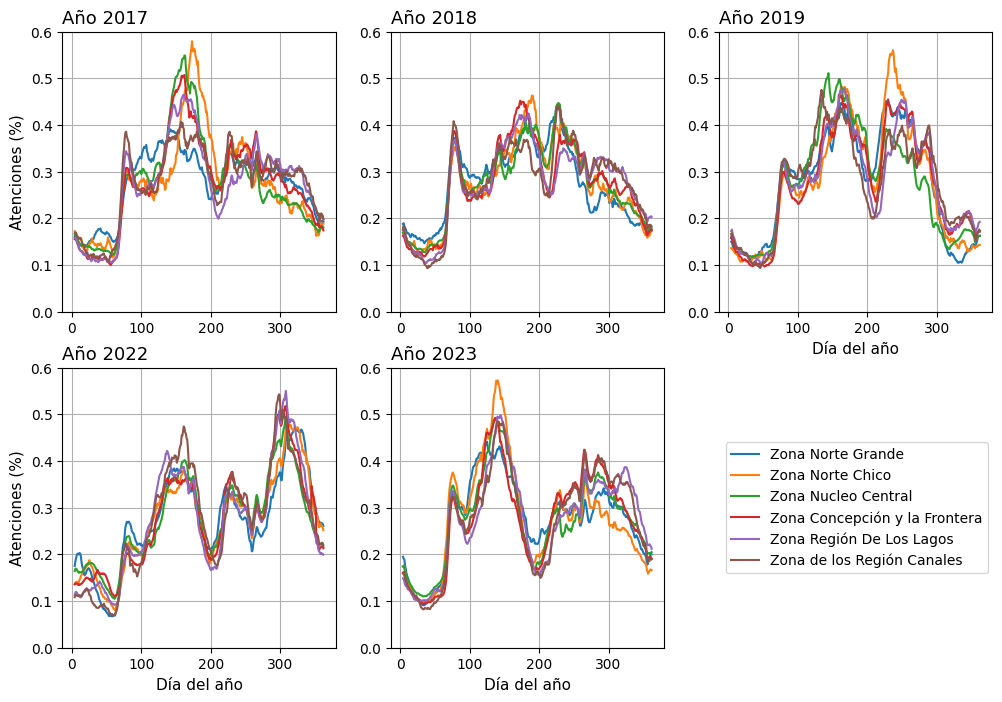
En la Ilustración 1, se muestra el total de atenciones diarias de urgencia en que se han considerado datos de establecimientos públicos de todo el país separado agrupados a nivel de día y donde cada curva representa los datos de un año. Notamos que el comportamiento de los años 2020 y 2021 se ve influenciado fuertemente por los efectos de la pandemia COVID-19 y su comportamiento es muy diferente al resto de los años, razón por la cual éstos serán excluidos de análisis posteriores. 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.

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.



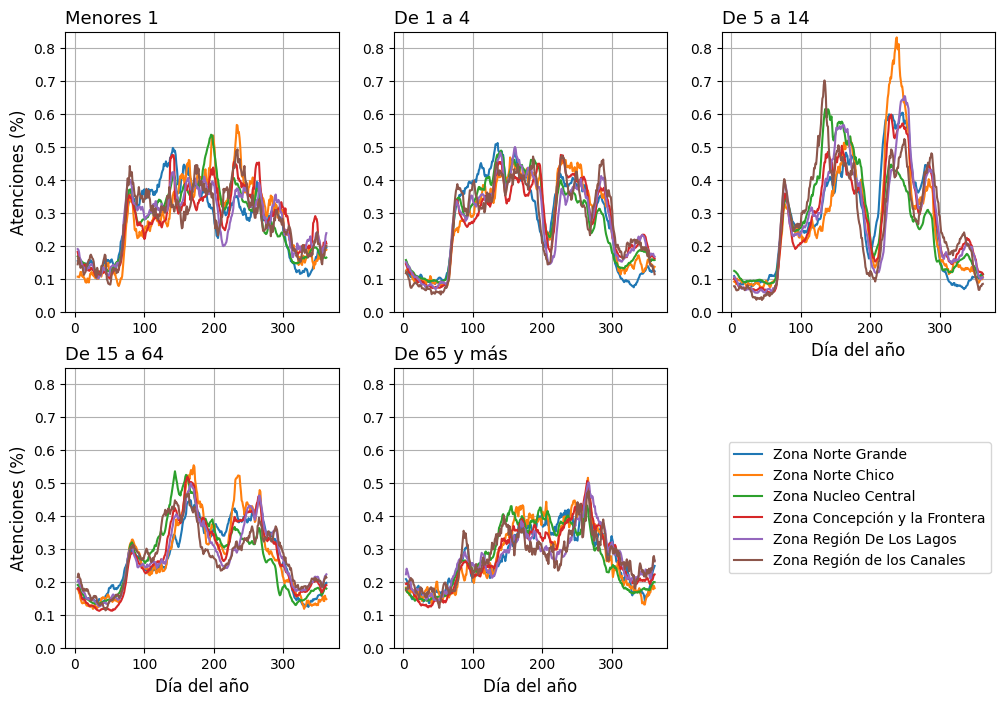
**Ilustración 2** Perfil de demanda Año 2019 suavizado a 7 días y desagregado para diferentes grupos etarios.



**Ilustración 3** Promedio móvil a 7 días de las atenciones diarias de urgencia como porcentaje de las atenciones anuales agrupadas por zona geográfica para cada día del año. Años 2017,2018,2019,2022,2023.

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.

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.



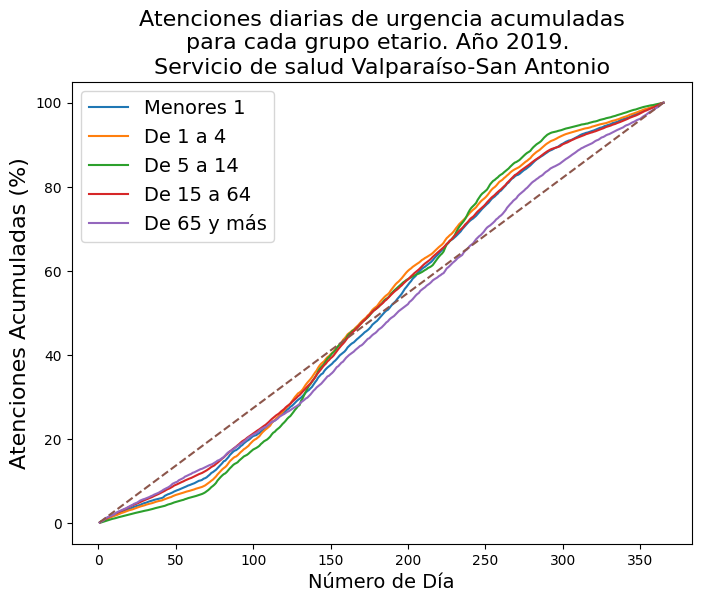
**Ilustración 4** Promedio móvil a 7 días de las atenciones diarias de urgencia como porcentaje de las atenciones anuales, correspondientes al año 2019, separado por grupos etarios y zonas geográficas

Para analizar cómo se relacionan las variables de grupo etario y zona geográfica para un año fijo en la Ilustración 4 se muestra los perfiles de demanda correspondientes al año 2019. A partir de la figura es posible notar importantes diferencias entre zonas geográficas en los perfiles de demanda correspondiente al grupo etario De\_5\_a\_14 y un desfase temporal entre 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 que grupos pediátricos y no pediátricos deben ser estudiados de manera diferenciada, propondremos medidas que permitan identificar diferencias significativas entre sus perfiles de demanda. Para este propósito consideraremos series de datos correspondientes a las atenciones de urgencia diarias asociados a un cierto año específico y otras características asociadas a los grupos que deseamos comparar (distribuciones geográficas, etarias, tipo de establecimiento, etc.).

Para el primer análisis agruparemos las atenciones de urgencias respiratorias agrupadas por año, servicio de salud y grupo etario. De esta forma, por ejemplo, asociadas al triple año 2022, Servicio de Salud Valparaíso-San Antonio 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). Basados en la Ilustración 1 es razonable asumir que las significativas entre años describen factores externos no considerados, por lo cual parece apropiado estudiar datos correspondientes a diferentes años de manera separada.

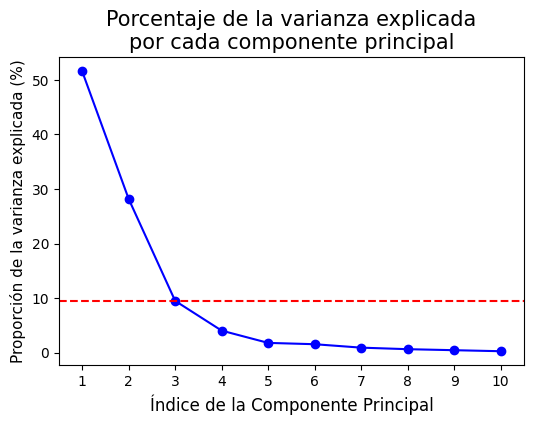


**Ilustración 5** Atenciones acumuladas como porcentaje de las atenciones anuales correspondientes al año 2019 y para el Servicio de Salud Valparaíso-San Antonio, agrupadas por grupo etario.

Una vez que se tienen las atenciones diarias correspondientes a cada uno de estos triples (año, servicio de salud, 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 hasta cierto día del año. Adicionalmente, para poder comparar grupos de población de diferente tamaño, escalaremos la curva de tal manera que se represente el porcentaje de las atenciones anuales que han ocurrido hasta cierta fecha.

En la 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 7 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.

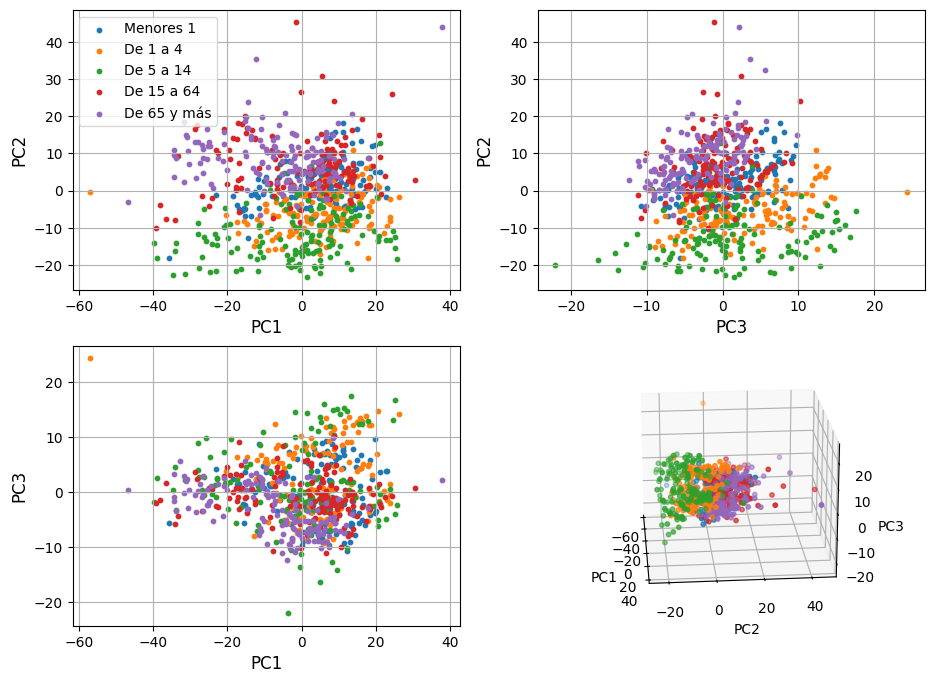


**Ilustración 6** Porcentaje de la varianza explicada por cada componente en el análisis de componentes principales de los datos correspondientes atenciones diarias acumuladas en que cada observación corresponde a un año, servicio de salud y grupo etario específico.

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.

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

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.

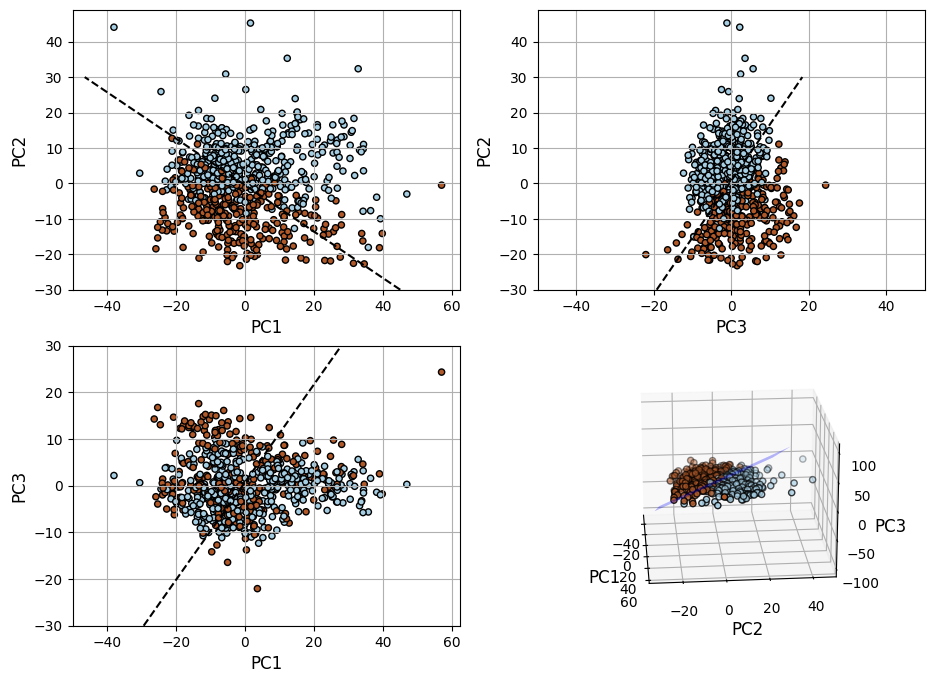


**Ilustración 7** Coordenadas en el espacio generado por las componentes principales. En la figura cada punto representa los datos diarios de un año de atenciones de urgencias respiratorias correspondientes a un determinado año, servicio de salud y grupo etario.

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. Un aspecto muy importante en es análisis es que el tamaño de la población que se está separando debe tenerse en cuenta al evaluar la precisión del clasificador. Por ejemplo si se quieren separar dos poblaciones, donde la población A representa el P% de los datos y la población B el (100-P)% de los datos, entonces un clasificador que diga que todos los datos corresponden a población A se equivocará el (1-P)% de las veces, por lo cual un buen clasificador necesariamente debe predecir correctamente más que este número. Para incorporar este problema en nuestro análisis consideraremos diferentes técnicas para medir la precisión de un clasificador.

https://help.qlik.com/es-ES/cloud-services/Subsystems/Hub/Content/Sense\_Hub/AutoML/scoring-binary-classification.htm#anchor-1

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.



**Ilustración 8** Representación gráfica de la clasificación binaria obtenida mediante SVM y el correspondiente plano que separa las clases usando las componentes principales usadas en la Ilustración 7. En este caso se indican en rojo el grupo formado al agrupar las clases De\_1\_a\_4 y De\_5\_a\_14, y en azul el correspondiente a la agrupación Menores\_1, De\_15\_a\_64 y De\_65\_y\_mas.

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 clasificaciones correspondientes a (De\_1\_a\_4, De\_5\_a\_14) vs (Menores\_1, De\_15\_a\_64, De\_65\_y\_mas); (Menores\_1,De\_1\_a\_4, De\_5\_a\_14, De\_15\_a\_64) vs (De\_65\_y\_mas) y por último (Menores\_1, De\_1\_a\_4, De\_5\_a\_14) vs (De\_15\_a\_64, De\_65\_y\_mas) son las que presentan mayores índices de precisión en la clasificación, lo cual se asocia a comportamientos cuantitativamente diferentes entre los grupos. Es destacable a partir de este resultado se valida la idea de que los grupos de pacientes pediátricos presentan un comportamiento diferentes a los pacientes no pediátricos en términos de demanda de atenciones de urgencia y por lo tanto es apropiado estudiarlos de forma separa con respecto al 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 en el perfil de demanda asociado a las atenciones de urgencia para diferentes tipos de centros de salud particularmente entre hospitales vs no hospitales.

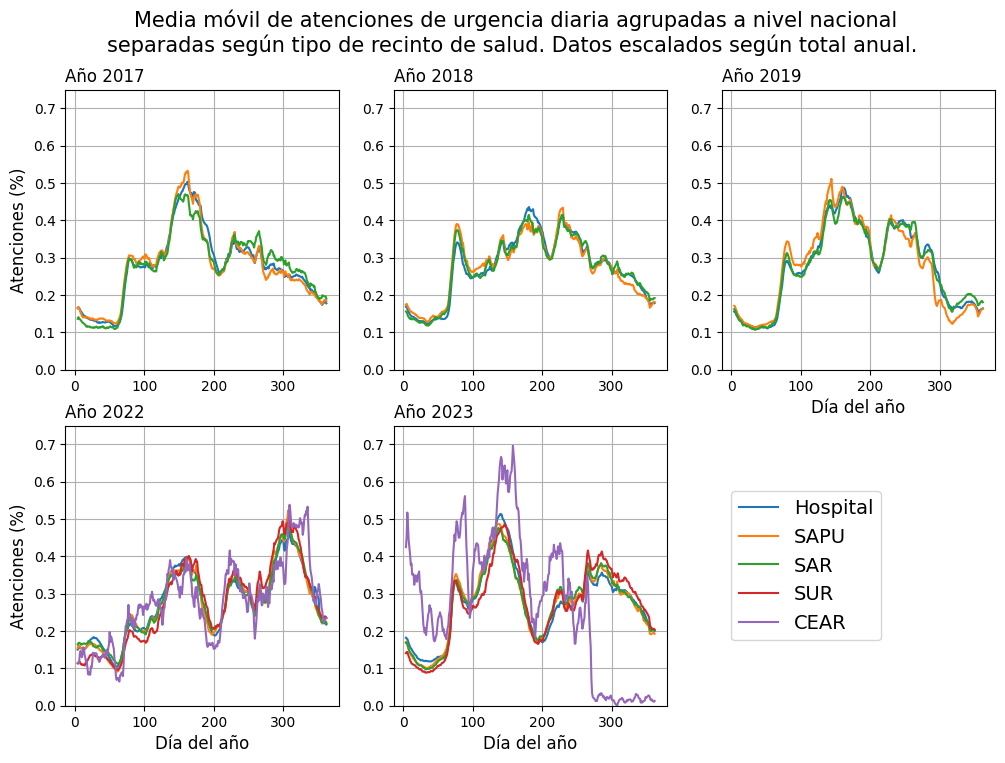


Ilustración 9 Atenciones diarias de urgencia para diferentes tipos de centros de salud separados por años. Media móvil a 7 días, datos escalados según total anual de atenciones.

Dentro de los datos considerados existen cinco tipos de centro de salud

1. Hospital
2. SAR: Servicio de Atención Primaria de Urgencias de Alta Resolución
3. SAPU: Servicio de Atención Primaria de Urgencias
4. CEAR: Centros Exclusivos de Atención Respiratoria. Solo tiene datos desde 2020
5. SUR: Servicio de Urgencia Rural. Solo tiene datos desde 2020

Para hacernos una idea de la magnitud de las atenciones consideradas, durante el año 2023 las atenciones de urgencias respiratorias se distribuyeron como se muestra en la Tabla 5.

Tabla 5 Cantidad e atenciones de urgencia por tipo de centro para el año 2023

|  |  |  |
| --- | --- | --- |
| Tipo de centro | Atenciones de Urgencia | % del Total |
| Hospital | 26.687.660 | 38,7% |
| SAPU | 22.775.071 | 33,0% |
| SAR | 15.132.586 | 21,9% |
| SUR | 4.336.901 | 6,3% |
| CEAR | 57.930 | 0,1% |
| Total | 68.990.148 | 100% |

A partir de la Ilustración 9 notamos que no parece haber diferencias significativas entre los perfiles de demanda asociados a los centros de tipo Hospital, SAR y SAPU. Para el tipo de centro CEAR tienen un comportamiento más ruidoso, lo cual probablemente se explica a que corresponden a una cantidad mucho menor de atenciones totales.

Para formalizar nuestras observaciones intentaremos hacer el mismo tipo de análisis que se usó para segregar grupos etarios usando PCA y SVM. Realizaremos este análisis en dos partes. Consideraremos datos separados a nivel de año, servicio de salud, tipo de establecimiento y grupo etario. En el primer caso consideraremos solo los datos correspondientes a los años 2022 y 2023 e intentaremos el análisis entre los 5 tipos de centro de salud.

Tabla 6 Resultados de la agrupación usando PCA y SVM. Se consideran solo datos de 2022 y 2023 pues son los únicos en el periodo considerado que contienen información de todos los tipos de centros. En la tabla se indica que tipos de centros se usaron en la separación, lo cual debe leerse como dichos centros vs el resto. La precisión indica la precisión promedio del clasificador sobre 1000 repeticiones.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Agrupación | Precisión |  | Agrupación | Precisión |
| **(CEAR,)** | **0.975784** |  | (SAR, SUR) | 0.581580 |
| **(SUR,)** | **0.784693** |  | (Hospital, SAPU) | 0.575807 |
| **(SUR, CEAR)** | **0.756943** |  | (Hospital, SUR) | 0.562813 |
| (Hospital, CEAR) | 0.658932 |  | (SAPU, CEAR) | 0.560875 |
| (SAR, CEAR) | 0.657761 |  | (SAPU, SAR) | 0.553557 |
| (SAR,) | 0.652722 |  | (SAPU, SUR) | 0.550659 |
| (Hospital,) | 0.650028 |  | (Hospital, SAR) | 0.505545 |
| (SAPU,) | 0.599187 |  |  |  |

A partir de los resultados que se muestran en la Tabla 5 obtenemos que separar los centros tipo CEAR y SUR del resto parece mostrar un comportamiento marcadamente diferente. En el caso de CEAR esta separación es poco significativa pues dichos centros atienden un grupo más bien pequeño de población. La separación con precisión del 78% para SUR vs el resto es un poco más interesante pues éstos atienden una cantidad relevante de atenciones de urgencia (6.3% en 2023, ver Tabla 5) pero las características asociadas a la población rural que atiende justifica el que el comportamiento de su perfil de demanda se pueda considerar

La precisión del clasificador en el resto de las clasificaciones es más bien baja por lo cual no lo consideraremos.

En un segundo análisis removeremos los datos correspondientes a CEAR y SUR de manera para enfocarnos en los otros tres tipos de centro y para ello consideraremos datos correspondientes a los años 2017, 2018, 2019, 2022 y 2023. Los resultados de este segundo análisis se presentan en la Tabla 7. Los resultados confirman nuestras observaciones basadas en la Ilustración 9 de que los perfiles de demanda para atenciones de urgencia para los centros de salud tipo Hospital, SAR y SAPU no presentan diferencias significativas.

Tabla 7 Resultados del análisis PCA-SVM para discriminar el comportamiento de los perfiles de demanda correspondiente a diferentes tipos de centros de salud con los datos correspondientes a los años 2017, 2018. 2019, 2022 y 2023. (a) Clasificación obtenida al considerar todos los grupos etarios. (b) Considerando solo población pediátrica (Menores 1, De 1 a 4, De 5 a 14). (c) Solo población no pediátrica (De 15 a 64 y De 65 y más).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (a) | Agrupación | Precisión | (b) | Agrupación | Precisión | (c) | Agrupación | Precisión |
|  | (SAR,) | 0.613862 |  | (SAR,) | 0.608916 |  | (Hospital,) | 0.635453 |
|  | (Hospital,) | 0.604138 |  | (Hospital,) | 0.604471 |  | (SAR,) | 0.614160 |
|  | (SAPU,) | 0.565497 |  | (SAPU,) | 0.566427 |  | (SAPU,) | 0.580993 |

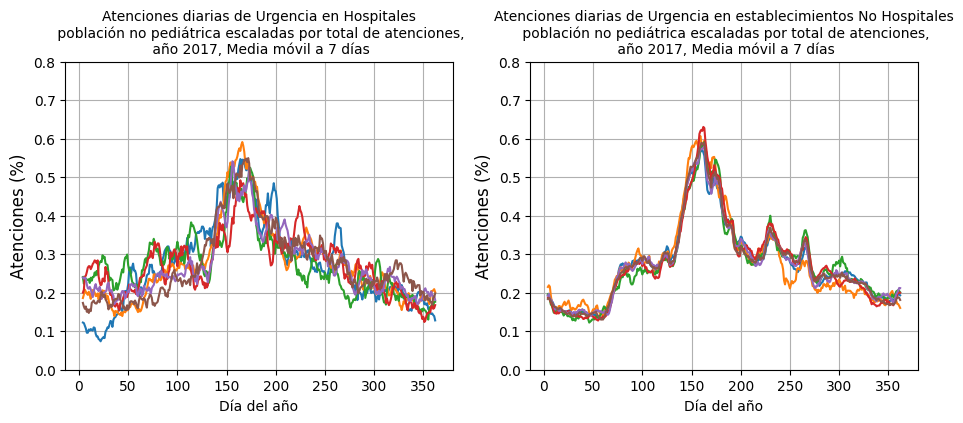
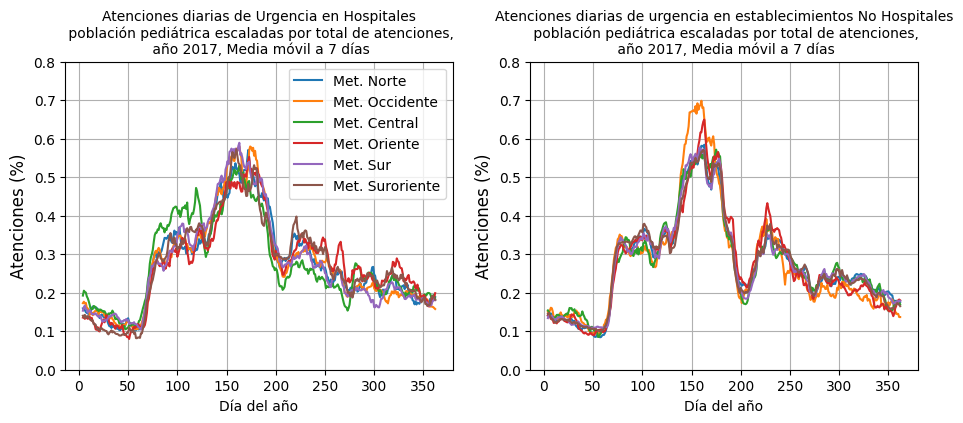
# Análisis específico para la Región Metropolitana

En esta sección nos enfocaremos en características más específicas asociadas a la población de la región metropolitana, la cual se organiza en 6 servicios de salud: Para este análisis nos restringiremos a los datos correspondientes a los servicios de salud: Metropolitano Norte, Metropolitano Occidente, Metropolitano Central, Metropolitano Oriente, Metropolitano Sur, Metropolitano Suroriente. Ellos coordinan ciertos aspectos administrativos asociados a sus correspondientes su zona geográfica. 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.

## Comparación atenciones de urgencias respiratorias

En una primera etapa validaremos que la segregación entre poblaciones pediátricas y no pediátricas siga siendo significativa cuando nos restringimos a esta población.

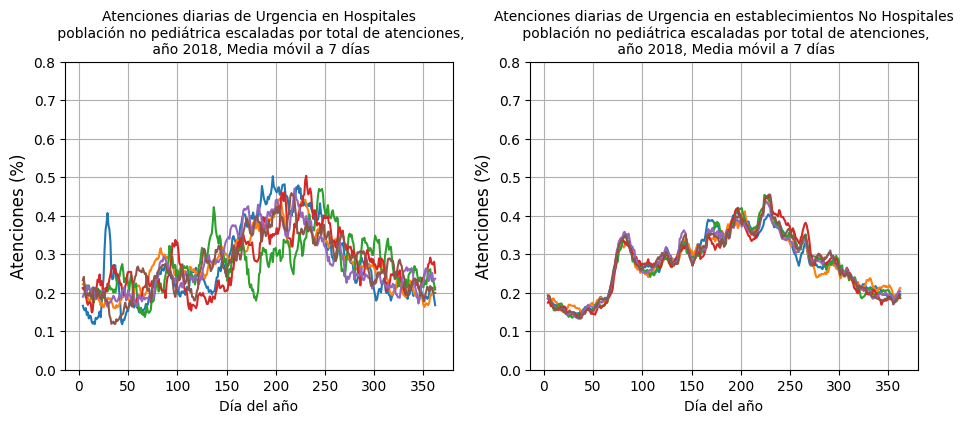
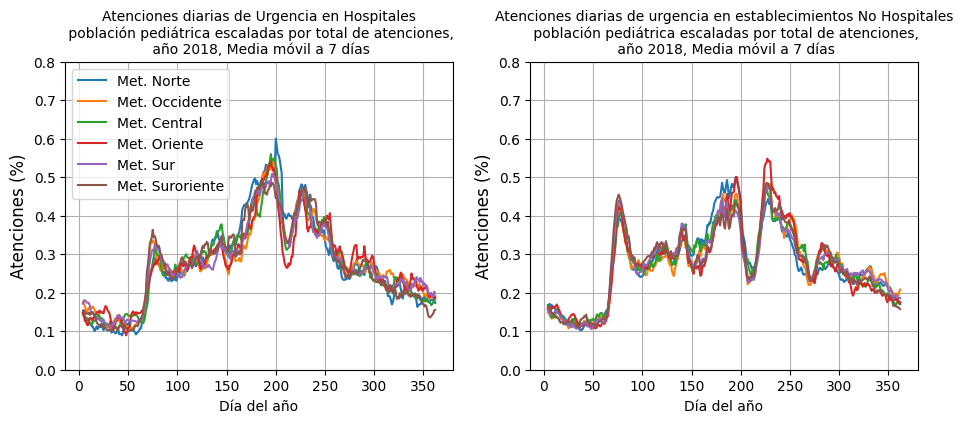
En nuestro segundo análisis queremos estudiar la dependencia geográfica (caracterizada por los servicios de salud), dependencia etaria (mediante la separación de poblaciones pediátricas y no pediátricas) y la dependencia por tipo de recinto de salud (Hospitales vs centros no hospitalarios). Para esto construiremos ilustraciones que permiten identificar para un año fijo, grupo etario fijo y tipo de recinto fijo, las curvas de demanda correspondientes a cada uno de los servicios de salud de la región metropolitana. Luego en base a tales Ilustraciones obtendremos diferencias cualitativas identificadas gracias a estas variables.



**Ilustración** **10** Promedio móvil de las curvas de demanda de atenciones de urgencia separadas según grupo etario y tipo de recinto de Salud. Cada curva representa los datos correspondientes a uno de los servicios de salud de la región metropolitana para el año 2017.

Basados en las Ilustraciones 9, 10, 11 y 12 podemos identificar algunas diferencias importantes en el comportamiento de las curvas de demanda. Notamos que para establecimientos no Hospitales (CEAR, SAR, SUR, SAPU) las curvas de demanda son mucho más parecidas a lo largo de los diferentes servicios de salud comparados con los Hospitales. En el otro extremo notamos que la demanda de atenciones de urgencias Hospitalarias para poblaciones no pediátricas presenta variaciones significativas a entre diferentes servicios de salud.

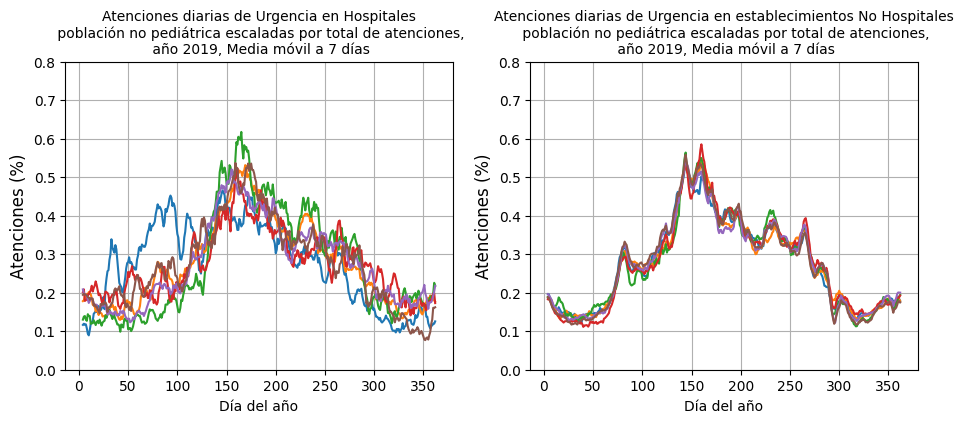
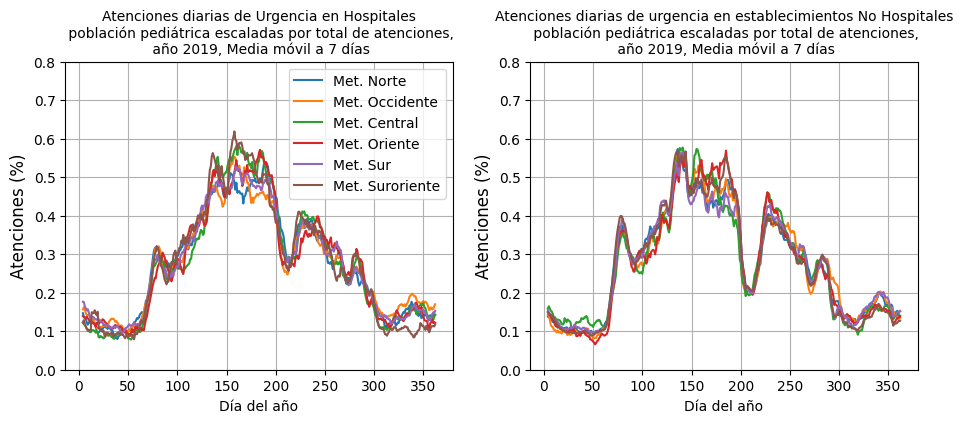
Con respecto a la fecha del peak de demanda a las atenciones de urgencia en recintos que no son hospitales notamos comportamiento muy similar para diferentes servicios de salud y grupos etarios. Por otro lado, en recintos hospitalarios se pueden observar diferencias cercanas a una semana para poblaciones pediátricas, sin embargo, para poblaciones no pediátricas las diferencias entre las fechas pueden ser más significativas como se observa en los años 2018, 2019 y 2023.



**Ilustración 11** Curvas de demanda de atenciones de urgencia separadas según grupo etario y tipo de recinto de Salud. Datos correspondientes a la región metropolitana, año 2018

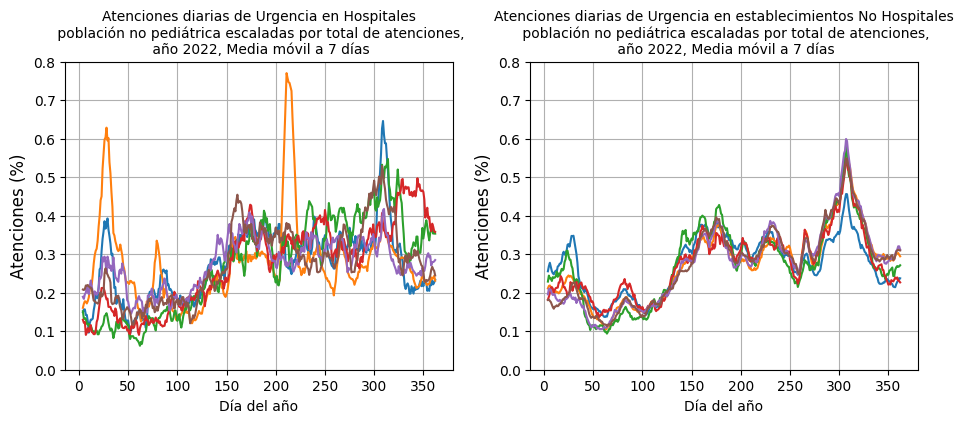
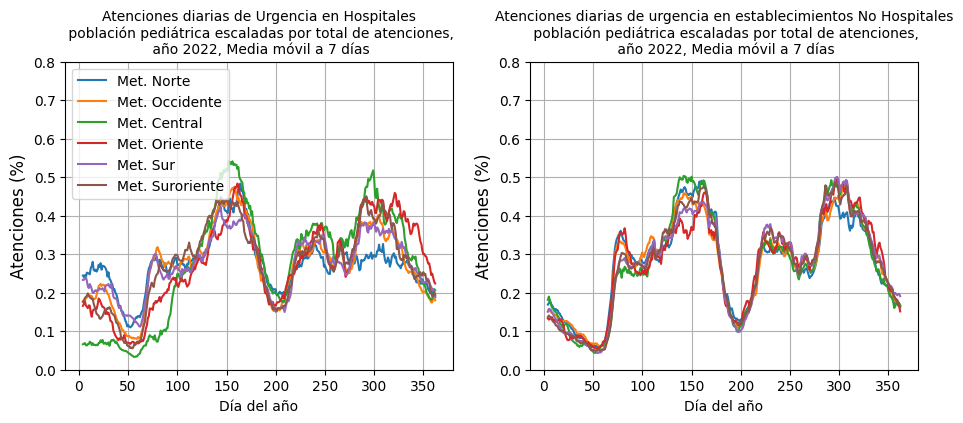
Al enfocarnos en el comportamiento de las poblaciones pediátricas, notamos que ellas presentan dos peak de atenciones de urgencia, uno entre fines de mayo y principios de junio, una caída importante en el periodo correspondiente a las vacaciones de invierno, y luego otro peak de menor magnitud a finales de julio, principios de agosto. Aun analizando esta pequeña muestra de años notamos que variaciones importantes son posibles, así por ejemplo el año 2018 vemos que el mayor peak de atenciones se ubica alrededor del día 200, es decir mediados de julio, lo cual refuerza la idea de que efectuar predicciones usando solo información correspondiente a años anteriores es insuficiente. Este comportamiento presenta una diferencia importante entre los grupos etarios pediátricos y no pediátricos, lo cual justifica analizarlos de manera separada.

Los datos correspondientes al año 2022 (Ilustración 12) presentan particularidades interesantes, vemos que durante el segundo semestre del año hay una cantidad mayor de urgencias respiratorias comparadas con lo que uno espera mirando la información histórica, lo cual se explica en parte por la normalización de la asistencia a clases post pandemia[[10]](#footnote-10).

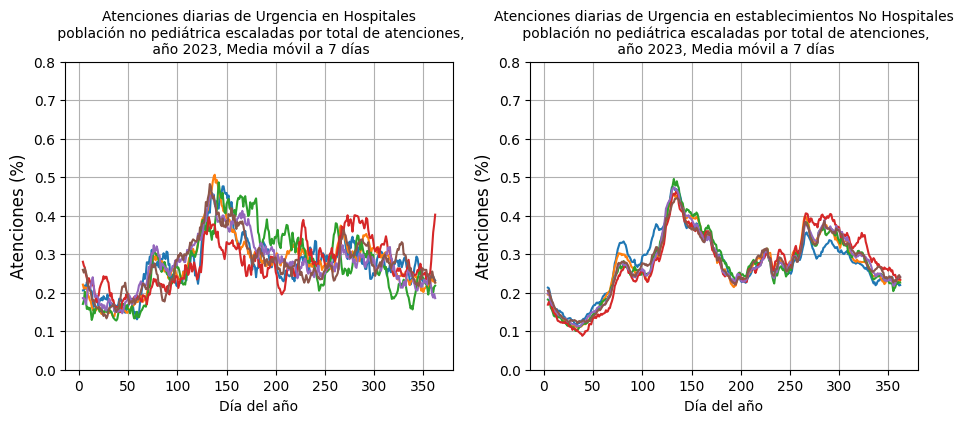
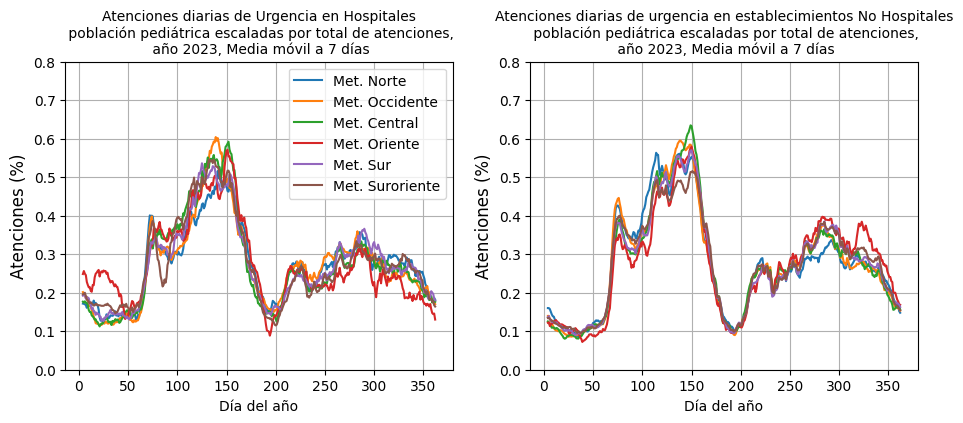


**Ilustración 12** Curvas de demanda de atenciones de urgencia separadas según grupo etario y tipo de recinto de Salud. Datos correspondientes a la región metropolitana, año 2019

Al enfocarnos en el comportamiento de las atenciones de urgencia en hospitales correspondientes a las poblaciones no pediátricas, lo primero que notamos es que ellas presentan una mayor variabilidad cuando se comparan diferentes servicios de salud. Para este grupo de pacientes es esperable observar gran variabilidad pues corresponde a un rango de edades muy grandes, el cual agrupa población con comportamientos muy diferentes. Adicionalmente estas diferencias se ven amplificadas al notar que la distribución geográfica representada por los diferentes servicios de salud codifica información respecto a la distribución socioeconómica de la población.



**Ilustración 13** Curvas de demanda de atenciones de urgencia separadas según grupo etario y tipo de recinto de Salud. Datos correspondientes a la región metropolitana, año 2022



**Ilustración 14** Curvas de demanda de atenciones de Urgencia separadas según grupo etario y tipo de recinto de Salud. Datos correspondientes a la región metropolitana, año 2023

## Análisis de Hospitalizaciones por causa respiratoria

Para considerar las hospitalizaciones por causa respiratoria, primero observamos que a partir de los datos (2017-2023) en la región metropolitana éstas se efectúan en dos tipos de centros: Hospitales (99.1%) y SAR (0.9%). Para analizar las hospitalizaciones no separaremos por tipo de recinto.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Ilustración 15 Hospitalizaciones diarias por causa respiratoria para poblaciones pediátricas para la región metropolitana. Se muestra la media móvil a 7 días y los datos son escalados según el total de hospitalizaciones anuales

Al observar los datos de hospitalizaciones separados entre poblaciones pediátricas (Ilustración 14) y no pediátricas (Ilustración 15), notamos importantes diferencias. Por un lado, consistentemente a lo largo de los años las poblaciones pediátricas presentaron un marcado peak entre los días 150-200 y la magnitud del peak es significativamente más grande que para el caso de las atenciones de urgencia. Para las hospitalizaciones este alcanza un máximo de aproximadamente un 1% de las hospitalizaciones anuales en un solo día, sin embargo, en el caso de las atenciones de urgencia (Ilustración 9 - Ilustración 13) este peak alcanza valores cercanos al 0.6% de las atenciones anuales.

Gráfico, Histograma

Descripción generada automáticamente

Ilustración 16 Hospitalizaciones diarias por causa respiratoria para poblaciones no pediátricas para la región metropolitana. Se muestra la media móvil a 7 días y los datos son escalados según el total de hospitalizaciones anuales

# 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:

1. IdEstablecimiento
2. NEstablecimiento
3. IdCausa
4. GlosaCausa
5. Total
6. Menores\_1
7. De\_1\_a\_4
8. De\_5\_a\_14
9. De\_15\_a\_64
10. De\_65\_y\_mas
11. fecha
12. semana
13. GLOSATIPOESTABLECIMIENTO
14. GLOSATIPOATENCION
15. GlosaTipoCampana
16. CodigoRegion
17. NombreRegion
18. CodigoDependencia
19. NombreDependencia
20. CodigoComuna
21. NombreComuna

Detalles de la construcción de las diferentes ilustraciones y análisis están disponibles en un repositorio GitHub[[11]](#footnote-11).

# Apéndice

## Lista de Servicios de Salud

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

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. [↑](#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. Reporte nacional de asistencia de marzo a octubre de los años 2018, 2022 y 2023, y análisis de la asistencia de octubre. Centro de Estudios MINEDUC. Enero 2024 [↑](#footnote-ref-10)
11. https://github.com/estebanpaduro/CMM-SD\_comparacion\_poblaciones\_IRA [↑](#footnote-ref-11)