

Análisis de Tendencias de Defunciones en Chile (1990-2021)

La presente propuesta de proyecto, tiene como objetivo relacionar varios datasets: ingresos hospitalarios y las defunciones en Chile, entre los años 1990-2021. Esta motivación surge de problemáticas interesantes, como los desafíos que tiene el sistema de salud, relacionado a enfermedades crónicas más comunes, accidentes, envejecimiento de la población, disparidades en los accesos a la salud por región, entre otros.

También el análisis busca responder preguntas valiosas para la prevención y conciencia en la población, además del mismo sistema de salud, para poder adoptar medidas claves en la gestión sanitaria y en la atención hospitalaria. Con la fusión de los datasets es posible analizar las saturaciones de los hospitales o las causas de mortalidad más comunes en Chile. Además es importante revisar la calidad de los datos proporcionados por el Ministerio de Salud (MINSAL). Hurgar en la precisión y veracidad de estos datos tiene un aspecto crucial, ya que, estos errores podrían afectar directamente a la población. Por lo anterior, esta información puede ser de ayuda para gestores de salud, la población común, y autoridades sanitarias.

Descripción de los Datos:

Ambos datasets han sido recopilados desde la fuente pública de datos del departamento de estadísticas e información de salud del gobierno, obtenido de: <https://deis.minsal.cl/>.

1) Defunciones por causa 1990-2021:

- Descripción de cada variable y tipo:
- <https://github.com/vaniaulloa/ATDC--imt2200/tree/main/defunciones%201994-2021>
- formato: archivo csv
- volumen: 04 Mb

2) Egresos hospitalarios:

- Descripción de cada variable y tipo:
- <https://github.com/vaniaulloa/ATDC--imt2200/tree/main/Datos%20egresos%20hospitalarios>
- formato: archivo csv
- volumen: 5,75 GB (datos desde el 2001 hasta el 2021)

- **Preguntas de investigación:** proponga al menos 5 preguntas específicas que buscarán responder con base en los datos presentados. Estas preguntas pueden ser relativas a relaciones entre variables, capacidad de predicción de una variable, clasificación, identificación de clusters o categorías, etc.
 - 1) ¿Cuál es la causa externa más común dentro de los datos analizados? se considera causa externa, como aquellas no relacionadas a enfermedades.
 - 2) ¿Cómo influyen los periodos de saturación hospitalaria en la tasa de mortalidad en ciertas patologías?
 - 3) ¿Existe una tendencia con las causas de muerte en cada uno de los grupos etarios? considerando también las disparidades de género en las tasas de mortalidad.
 - 4) ¿Qué tan efectivas son las hospitalizaciones en la reducción de la mortalidad por enfermedades? Aquí consideramos una investigación, con las causas más comunes de muerte por enfermedades, considerando una diferencia entre enfermedades tratables y no tratables.
 - 5) ¿Existen factores que explican las variaciones en las tasas de mortalidad y hospitalizaciones? Considerando un análisis por región (demográfico, socioeconómico, otros).

Para este análisis, utilizaremos varios métodos. Dentro de los **métodos computacionales**, tenemos planeado emplear los siguientes:

- **Análisis Exploratorio de Datos (EDA):** Visualización de datos a través de gráficos como histogramas, diagramas de dispersión y mapas de cloropletas, entre otros que se consideren adecuados según cada ítem y/o pregunta, lo que nos permitirá analizar tendencias y patrones relacionados con nuestras preguntas de investigación.
- **Preprocesamiento de Datos:** Limpieza y transformación de los datos, que incluye la eliminación de duplicados, el manejo de valores nulos, la normalización y la fusión de nuestras base de datos. Esto garantizará la calidad de los datos para su análisis.
- **Fusión de Datos:** Integraremos diferentes fuentes de datos para analizarlas en conjunto y responder varias de nuestras preguntas, lo que nos permitirá extraer conclusiones importantes y descubrir relaciones entre los conjuntos de datos.

En cuanto a los **métodos estadísticos**, planeamos usar los siguientes:

- **Medidas de tendencia central:** como la media, mediana y moda, para describir la distribución de los datos, en diferentes ciudades o en general.
- **Tasa de Mortalidad General (TMG):** Indicará el número de defunciones por cada 1,000 habitantes en un año determinado.
- **Tasa de Mortalidad Específica por Causa (TMEC):** Refleja el número de defunciones por una causa específica por cada 100,000 habitantes, permitiendo un análisis más detallado de los factores que influyen en la mortalidad.

La muestra de los datos, se puede obtener mediante el repositorio del proyecto, que se encuentra en el siguiente enlace:

<https://github.com/vaniaulloa/ATDC--imt2200/tree/main/Muestras>

Estas se encuentran como archivos del formato csv, cada una contiene 20 datos, seleccionados aleatoriamente, a modo ilustrativo.

Visualización de los dataframes:

- Defunciones:

AÑO	FECHA_DEF	SEXO_NOMBRE	EDAD_TIPO	EDAD_CANT	COD_COMUNA	COMUNA	NOMBRE_REGION	DIAG1	CAPITULO_DIAG1	...	DIAG2	CAPITULO_DIAG2	GLOSA_CAPITULO_DIAG2	CODIGO_GRUPO
1308	2020	2020-03-07	Mujer	1.0	87	3101	Copiapó	De Atacama	C449	C00-D48	...	NaN	NaN	NaN
1370	2020	2020-06-27	Hombre	1.0	95	2201	Calama	De Antofagasta	B572	A00-B99	...	NaN	NaN	NaN
1731	2020	2020-03-25	Mujer	1.0	95	9101	Temuco	De La Araucanía	A419	A00-B99	...	NaN	NaN	NaN
1732	2020	2020-03-13	Mujer	1.0	89	10107	Llanquihue	De Los Lagos	C159	C00-D48	...	NaN	NaN	NaN
1983	2020	2020-06-26	Mujer	1.0	91	8110	Talcahuano	Del Biobío	A047	A00-B99	...	NaN	NaN	NaN
...
2785632	2020	2020-10-30	Mujer	1.0	77	7101	Talca	Del Maule	U072	U00-U99	...	NaN	NaN	NaN
2785633	2020	2020-11-08	Hombre	1.0	52	13119	Maipú	Metropolitana de Santiago	U072	U00-U99	...	NaN	NaN	NaN
2785634	2020	2020-11-18	Hombre	1.0	51	13126	Quinta Normal	Metropolitana de Santiago	U072	U00-U99	...	NaN	NaN	NaN
2785635	2020	2020-12-13	Mujer	1.0	70	14101	Valdivia	De Los Ríos	U072	U00-U99	...	NaN	NaN	NaN
2785636	2020	2020-12-21	Hombre	1.0	61	12101	Punta Arenas	De Magallanes y de La Antártica Chilena	U072	U00-U99	...	NaN	NaN	NaN

- Egresos hospitalarios:

AÑO	FECHA_DEF	SEXO_NOMBRE	EDAD_TIPO	EDAD_CANT	COD_COMUNA	COMUNA	NOMBRE_REGION	DIAG1	CAPITULO_DIAG1	...	DIAG2	CAPITULO_DIAG2	GLOSA_CAPITULO_DIAG2	CODIGO_GRUPO
0	2019	2019-06-11	Hombre	1.0	57	13401	San Bernardo	Metropolitana de Santiago	G122	G00-G99	...	NaN	NaN	NaN
1	2012	2012-09-02	Mujer	1.0	65	8110	Talcahuano	Del Bío-Bío	J440	J00-J99	...	NaN	NaN	NaN
2	2014	2014-06-22	Mujer	1.0	78	9116	Saavedra	De La Araucanía	J90X	J00-J99	...	NaN	NaN	NaN
3	1996	1996-05-12	Mujer	2.0	2	7101	Talca	Del Maule	9349	NaN	...	911X	NaN	NaN
4	2001	2001-04-25	Hombre	2.0	2	13201	Puente Alto	Metropolitana de Santiago	K729	K00-K93	...	NaN	NaN	NaN
5	2020	2020-08-18	Mujer	1.0	89	13501	Melipilla	Metropolitana de Santiago	E86X	E00-E99	...	NaN	NaN	NaN
6	1991	1991-09-30	Mujer	1.0	88	13109	La Cisterna	Metropolitana de Santiago	1809	NaN	...	NaN	NaN	NaN
7	2018	2018-10-23	Mujer	1.0	90	8110	Talcahuano	Del Bío-Bío	J841	J00-J99	...	NaN	NaN	NaN
8	1990	1990-09-20	Hombre	1.0	65	13130	San Miguel	Metropolitana de Santiago	485X	NaN	...	NaN	NaN	NaN
9	2006	2006-10-16	Hombre	1.0	80	7301	Curicó	Del Maule	G700	G00-G99	...	NaN	NaN	NaN
10	1999	1999-02-05	Mujer	1.0	89	13123	Providencia	Metropolitana de Santiago	I64X	I00-I99	...	NaN	NaN	NaN

Bibliografía

- Minsal. <https://deis.minsal.cl/> (sección datos abiertos)