



Facultad de Ciencias  
Instituto de Estadística

**Estimación de los días de estadía  
hospitalaria de los adultos mayores de la  
región de Valparaíso para el 2016 a  
través de modelos de conteo: Poisson y  
Binomial Negativo**

Trabajo final presentado por:  
**Johuwsa Francis Cornejo Ramírez**

Trabajo de titulación para optar al título de:  
**Ingeniero Estadístico**

Profesora guía:  
**Claudia Navarro Villarroel, Ph.D.**

Valparaíso, Chile, diciembre de 2016

---

# Índice general

---

AGRADECIMIENTOS	6
ABREVIATURAS	7
RESUMEN	8
ABSTRACT	9
HIPÓTESIS	10
OBJETIVOS	11
NOCIONES PRELIMINARES	12
INTRODUCCIÓN	18
<b>1. MODELOS LINEALES GENERALIZADOS</b>	<b>20</b>
1.1. <b>Componentes del modelo</b> (Alcaide, 2015) . . . . .	21
1.1.1. <b>Componente aleatoria</b> . . . . .	21
1.1.2. <b>Componente sistemática</b> (López, 2016 y Contreras, 2012)	22
1.1.3. <b>Función de enlace</b> . . . . .	22
1.2. <b>Estimación de los parámetros del modelo</b> (López, 2016) . .	24
1.3. <b>Adecuación del modelo</b> (Alcaide, 2015). . . . .	25
1.3.1. <b>Bondad de ajuste (Alcaide 2015)</b> . . . . .	26
1.4. <b>Análisis de los residuos</b> (Osses, 2013) . . . . .	28
1.5. <b>Interpretación</b> (Alcaide, 2015) . . . . .	30

<b>2.</b>	<b>MODELOS LINEALES GENERALIZADOS POISSON Y BINOMIAL NEGATIVO</b>	<b>32</b>
2.1.	<b>Datos de recuento</b> (Alcaide, 2015) . . . . .	32
2.2.	<b>Modelo lineal generalizado Poisson (MLGP)</b> (López, 2016)	33
2.2.1.	<b>Componente aleatoria</b> . . . . .	34
2.2.2.	<b>Componente sistemática</b> . . . . .	34
2.2.3.	<b>Función de enlace</b> . . . . .	34
2.2.4.	<b>Estimación de máxima verosimilitud</b> . . . . .	35
2.3.	<b>Modelo lineal generalizado Binomial Negativo (MLGBN)</b> (López, 2016) . . . . .	39
2.3.1.	<b>Componente aleatoria</b> . . . . .	40
2.3.2.	<b>Componente sistemática</b> . . . . .	40
2.3.3.	<b>Función de enlace</b> . . . . .	40
2.3.4.	<b>Estimación de máxima verosimilitud</b> . . . . .	41
2.3.5.	<b>Método Newton Raphson</b> (López, 2016). . . . .	45
2.3.6.	<b>Adecuación del modelo</b> (Alcaide, 2015) . . . . .	47
2.3.7.	<b>Interpretación del modelo</b> (Alcaide, 2015) . . . . .	49
<b>3.</b>	<b>APLICACIÓN</b>	<b>51</b>
3.1.	<b>Análisis de datos</b> . . . . .	51
3.2.	<b>VARIABLES</b> . . . . .	52
3.3.	<b>Análisis exploratorio</b> . . . . .	53
3.4.	<b>Aplicación del modelo lineal generalizado Poisson (MLGP)</b>	69
3.5.	<b>Diagnóstico del MLGP</b> . . . . .	71
3.6.	<b>Aplicación del modelo lineal generalizado Binomial Negativo (MLGBN)</b> . . . . .	72
3.7.	<b>Diagnóstico del MLGBN</b> . . . . .	74
	<b>CONCLUSIÓN</b>	<b>76</b>
	<b>RECOMENDACIONES</b>	<b>77</b>
	<b>APÉNDICE</b>	<b>78</b>
3.8.	<b>Código en Stata</b> . . . . .	78
3.9.	<b>Salidas computacionales</b> . . . . .	84
	<b>Referencias</b>	<b>92</b>

---

# Índice de figuras

---

1.	Población total estimada de Chile por grupo etario. 2007, 2010 y 2025. . . . .	14
2.	Perspectiva de la población mundial de adultos mayores para el 2025. . . . .	15
3.	Porcentaje estimado de la población de adultos mayores desde 1992 al 2015. . . . .	16
4.	Cantidad estimada de adultos mayores por cada 100 niños de 15 años en cada región de Chile al 2025. . . . .	17
3.1.	Frecuencia de días de estadía hospitalaria del adulto mayor en la región de Valparaíso. . . . .	61
3.2.	Gráfico de días de estadía hospitalaria por sector del servicio de salud de recurrencia. . . . .	62
3.3.	Gráfico de días de estadía hospitalaria por sexo del paciente. . . . .	63
3.4.	Gráfico días de estadía hospitalaria por previsión del paciente. . . . .	63
3.5.	Gráfico de días de estadía hospitalaria por modalidad de atención de los beneficiarios de Fonasa. . . . .	64
3.6.	Gráfico de días de estadía hospitalaria por comuna de residencia del paciente. . . . .	65
3.7.	Gráfico de días de estadía hospitalaria por condición del egresado. . . . .	65
3.8.	Gráfico de días de estadía hospitalaria por servicio clínico realizado al paciente. . . . .	66
3.9.	Gráfico de días de estadía hospitalaria por sí presenta intervención quirúrgica el egresado. . . . .	67
3.10.	Gráfico de días de estadía hospitalaria por edad del egresado. . . . .	67
3.11.	Residuos Anscombe v/s predicción de la media para días de estadía hospitalaria del adulto mayor . . . . .	71

3.12. Residuos Anscombe v/s predicción de la media para días de estadía hospitalaria del adulto mayor. . . . .	74
---	----

---

## AGRADECIMIENTOS

---

En primer lugar, quiero agradecerle a mi abuela María Malhuen Lizama, el pilar fundamental en mi vida, con quien siempre estaré en deuda, dado que, siempre entregó lo mejor de ella, para que llegaré a ser la persona y el profesional que soy hoy en día. También quiero agradecerles a las personas que más me importan en la vida, a mi familia, quienes siempre me han entregado su apoyo incondicional en todos los desafíos que me he planteado, y me hacen saber día a día lo orgullosos que se sienten de mí.

Agradezco, además, a los profesores del departamento de estadística de la universidad, quienes siempre en cada clase entregaron lo mejor de ellos para transferirme de la mejor forma el conocimiento que necesitaba. Pero, agradezco de forma particular a mi profesora guía Claudia Navarro y el profesor Carlos Felipe Henríquez por su gran paciencia y por ser mis primeros profesores de estadística, los cuales lograron contagiarme el amor a esta ciencia.

Por último, pero no menos importante, quiero agradecerles a mis amigos, tanto de la universidad como de la vida, por siempre ser un apoyo y entregarme consuelo cuando más los necesite. A todos ellos, muchas gracias.

---

# ABREVIATURAS

---

A continuación se presentan algunas abreviaturas que se utilizan a lo largo de este trabajo.

- INE: Instituto Nacional de Estadística.
- DEIS: Departamento de Estadísticas e Información de Salud.
- MINSAL: Ministerio de Salud.
- MLG: Modelo Lineal Generalizado.
- MLGP: Modelo Lineal Generalizado Poisson.
- MLGBN: Modelo Lineal Generalizado Binomial Negativo.
- AIC: Criterio de Información de Akaike.
- BIC: Criterio de Información Bayesiano.
- MV: Máxima Verosimilitud.
- CV: Coeficiente de Variación.

---

## RESUMEN

---

En Chile, se sabe, a través del Instituto Nacional de Estadística (INE, 2007), que en 2015 los adultos mayores alcanzaron cerca del 15 % de la población y además, al observar la proyección demográfica de esta organización para el 2020, se espera que el 17,3 % de la población pertenezca a esta condición; de igual manera, se estima un ascenso en la esperanza de vida de las personas, alcanzando los hombres una esperanza de vida de 77,4 años y las mujeres 82,2 años. Es por esto, que es importante comenzar a implementar nuevas políticas públicas que se preparen para satisfacer el volumen de demandas de este grupo etario. Resulta útil poner en funcionamiento una mejora en cualquier área que implique al adulto mayor, por ello, resulta relevante activar una solución para optimizar el tiempo y designación de los recursos en la hospitalización de personas de la tercera edad en los hospitales en Chile. Para ayudar a satisfacer lo mencionado anteriormente, se estudiarán los datos de egresos hospitalarios de Chile en 2016 publicados por el Departamento de Estadísticas e Información de Salud (DEIS) del MINSAL. Se propone modelar estos datos a través de los modelos lineales generalizados Binomial Negativo y Poisson, para estimar la variable de respuesta “Días de Estadía” a través de las variables más influyentes contenidas en las bases de datos utilizada.

Palabras claves: adulto mayor, Valparaíso, modelo lineal generalizado Binomial Negativo, modelo lineal generalizado Poisson.

---

# ABSTRACT

---

In Chile, it is known, through the Instituto Nacional de Estadística (INE, 2007), that in Chile in the year of 2015 the elderly people reached about the 15% of the population and also, observing the demographic projection of this organization for the year 2020, it is expected that the 17.3% of the population will belong to that condition, likewise, it is expected an increase in life expectation of people, reaching it in 77.4 years in men and 82.2 years for women. That is the reason why is important to start to implement new public policies that are prepared to satisfy the volume of the demands of the age group. It is useful implement an improvement in any area that involves the older adults, therefore, it is important to activate a solution to optimize the time and the designation of resources in the hospitalization of the elderly people in the hospitals of Chile. To help the aforementioned, it will be studied the data of the hospital discharge from Chile used in 2016 published by the Departamento de Estadísticas e Informacion de Salud (DEIS) del MINSAL. It is proposed to model this data through the Negative Binomial and Poisson generalized linear models, to estimate the response variable “Days of Stay” through the most influential variables contained in the databases used.

Keywords: elderly, Valparaíso, Negative Binomial generalized linear model, Poisson generalized linear model.

---

## HIPÓTESIS

---

A través de un modelo lineal generalizado Binomial Negativo, se puede estimar los días de estadía hospitalarios del adulto mayor en la región de Valparaíso de 2016, obteniendo un resultado coherente y óptimo.

---

# OBJETIVOS

---

Los objetivos de este trabajo de titulación se presentan a continuación.

## Objetivo general

Estudiar los modelos lineales generalizados Poisson y Binomial Negativo que pertenecen a los modelos lineales generalizados y aplicarlos en los datos reales de días de estadía hospitalaria del adulto mayor de la región de Valparaíso.

## Objetivo específico

- Analizar las propiedades de los modelos lineales generalizados Poisson y Binomial Negativo.
- Aplicar las propiedades de los modelos lineales generalizados Poisson y Binomial Negativo.
- Comparar los resultados de los modelos lineales generalizados Poisson y Binomial Negativo.

---

## NOCIONES PRELIMINARES

---

En el mundo y principalmente en Latinoamérica, según Ceballos *et al.*(2014) el desbalance entre la demanda y la oferta que existe en los distintos servicios de salud resulta un motivo de preocupación, ya que, estos establecimientos deben estar preparados para acoger el máximo de demanda posible de la población, por esta causa es necesario utilizar los servicios y recursos hospitalarios de manera eficiente, sin dejar de lado las necesidades de los pacientes. Para ello existe un indicador mundial conocido como el “promedio de días de estancia hospitalaria”, que informa y mide el aprovechamiento de las camas y la agilidad con la que se entregan los servicios prestados por los hospitales. Al prolongarse este indicador, más allá del tiempo considerado, se genera un problema de preocupación mundial, perjudicando el correcto flujo de pacientes en los establecimientos, aumentando los costos de atención y el riesgo de eventos adversos provocando, en el peor de los casos, la mortalidad del paciente. Otros efectos negativos que perjudican al sistema de salud se explican a continuación:

En primer lugar, limita la capacidad de los hospitales, dificultando el acceso de pacientes a una cama hospitalaria; ocurriendo esto cuando se presenta una mayor estancia de lo programado, provocando menos disponibilidad de camas y mayor saturación en los servicios de urgencias, que es la principal puerta de entrada a la hospitalización (entre el 70 y 75 % de los ingresos). Se suma a esta situación, el escaso número de camas, un problema común en la capacidad de los hospitales en países de Latinoamérica según la encuesta del Departamento de Desarrollo Humano del Banco Mundial (2007).

En segundo lugar, se incrementa el costos de operación por el uso no apropiado de los recursos. Ocurre cuando el paciente cae en la denominación de estancia prolongada implicando un desperdicio o mala utilización de los recursos económicos y humanos, dado que esto lleva a costos adicionales en términos de personal, alimentación, insumos, uso de equipos y utilización de recursos extras del hospital (Ceballos *et al*, 2014).

Y, por último, afecta la calidad de atención en términos de la seguridad del paciente. Debido a que si incrementa por cualquier circunstancia la duración de la estancia hospitalaria, ésta provoca el aumento de manera significativa a la probabilidad de que el paciente tratado sufra un evento adverso o de contraer infecciones intrahospitalarias que llevan al paciente a elevar el número de días de estadía hospitalaria y en el peor de los casos a la muerte. (Ceballos *et al*, 2014).

Por lo explicado anteriormente, nace en las distintas instituciones de salud la necesidad de buscar reducir, dentro de lo posible, la estancia hospitalaria para mejorar el flujo de pacientes, buscando provocar un impacto en la eficiencia de la operación interna del centro de salud, sin verse afectada la calidad de la atención de los pacientes. Pero la estancia hospitalaria no solo está afectada por problemas de eficiencia, también es afectada por la demora de procedimientos quirúrgicos y entrega de diagnósticos a tiempo; y esto no es solo provocado por la administración hospitalaria, sino que también por las características propias de cada paciente; se conoce este factor como el “ajuste por riesgo”, que hace referencia a cuando la estancia hospitalaria se prolonga por factores externos donde no puede intervenir el centro de salud (socio-familiares y edad del paciente) (Ceballos *et al*, 2014).

Al hablar de Latinoamérica nace la pregunta puntual ¿y qué sucede en Chile en esta área? A través de los datos del MINSAL, se aprecia que el rendimiento de los hospitales ha mejorado, ya que de los ocho días promedio que permanecía un paciente internado en los centros de salud públicos en el 2010, este promedio bajó a seis días en el 2012, mejorando en un 15 % la capacidad hospitalaria, pero lamentablemente, con los datos publicados hasta el 2014 este promedio se ha mantenido fijo (Sandoval, 2013).

Como se ha visto en los párrafos anteriores, la reducción de los días de estadía hospitalaria tiene un impacto directo en la red de salud pública de Chile, al contar con 27 mil camas a nivel nacional, las que deben tener una rotación constante para evitar listas prolongadas de espera. Al optimizar este recurso por parte de los hospitales, se puede llegar a ahorrar a nivel nacional por paciente cerca de

\$1.403.765. Por esto al pensar en reducir sólo 1 día la cantidad de días de hospitalización de los 56.349 adultos mayores atendidos en la región de Valparaíso, se hubieran ahorrado cerca de \$79.100.753.985, lo que provoca ahorros millonarios para el Gobierno de Chile y además, en el 2012 se aprecia que las re hospitalizaciones bajaron, ya que al tomarse 30 días de hospitalización como límite, solo un 0,7 % de los pacientes tratados fueron reingresados (Sandoval, 2013).

También, resulta necesario revisar otra área de interés mundial, lo que es el envejecimiento de la población general productiva; esto se está convirtiendo en una característica principal de los países desarrollados y en vías de desarrollo, debido a que el porcentaje de personas de 60 años y más, está en considerable aumento, produciendo impactos múltiples en la sociedad (salud, educación, economía, entre otros).

Figura 1: Población total estimada de Chile por grupo etario. 2007, 2010 y 2025.

**Chile: Población total estimada al 30 de junio, según grandes grupos de edad e Índice de Dependencia Demográfica. 2007, 2010 y 2015 (Cifras absolutas y relativas)**

GRUPO DE EDAD	2007		2010		2015	
	Número	%	Número	%	Número	%
Total	16.598.074	100,0	17.094.270	100,0	17.865.185	100,0
0-14	3.958.354	23,8	3.814.756	22,3	3.745.210	21,0
15-59	10.634.036	64,1	11.066.078	64,7	11.481.644	64,3
60 o +	2.005.684	12,1	2.213.436	13,0	2.638.331	14,7
IDD		56,1		54,5		55,6

IDD = Índice de Dependencia Demográfica (Número de personas menores de 15 y de 60 o más, por cada cien de 15-59 años).

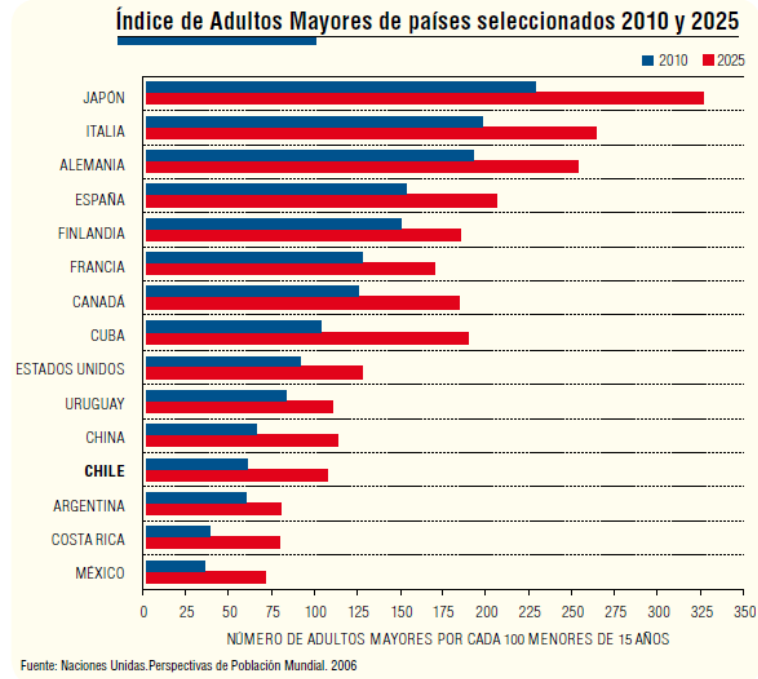
Fuente: INE, Chile: Proyecciones y Estimaciones de Población. Total País: 1990-2050. Agosto 2005.

Fuente: INE, Enfoque Estadístico Adulto Mayor en Chile (2007).

Chile, al igual que los países desarrollados, está pasando por una etapa avanzada de transición al envejecimiento demográfico de sus habitantes, produciéndose este envejecimiento a partir de 1960, con cifras leves, donde existieron alrededor de 20 adultos mayores por cada cien menores de 15 años. Posteriormente, a partir de 1963 el descenso sostenido de los embarazos, junto con el aceleramiento del aumento porcentual de los adultos mayores, dio inicio al fenómeno de la transición al envejecimiento demográfico. Esto disminuyó el aporte porcentual de los menores de 15 años y aumentó el aporte de forma gradual del adulto mayor. En 2006, una de cada diez personas corresponde a un adulto mayor y se espe-

ra que para el 2025, esta cifra aumente a uno por cada cinco personas (INE, 2007).

Figura 2: Perspectiva de la población mundial de adultos mayores para el 2025.

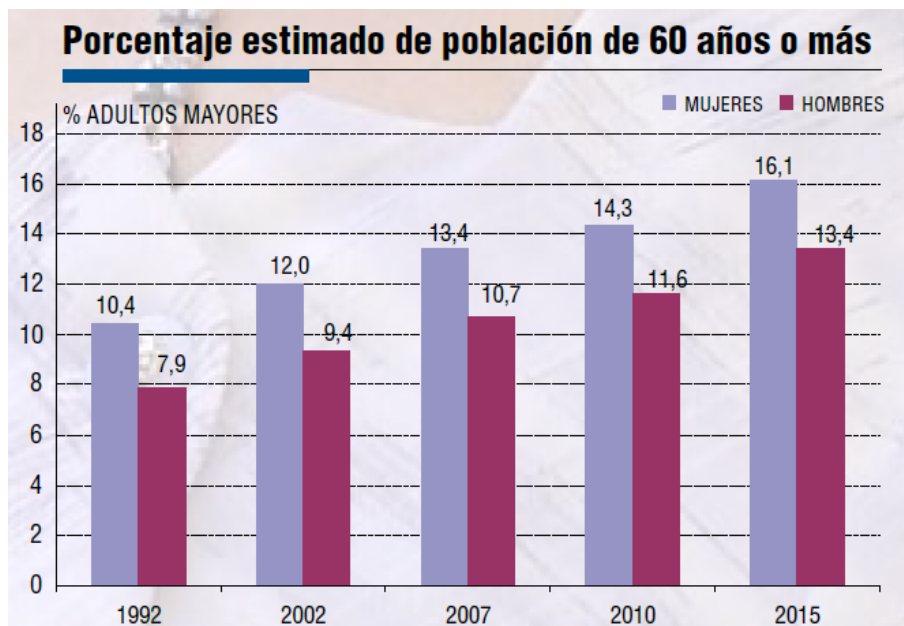


Fuente: INE, Enfoque Estadístico Adulto Mayor en Chile (2007).

Como se visualiza en la Figura 2, Chile ocupa el puesto décimo segundo de la lista de aumento de adultos mayores por país entre el 2010 y el 2025 y se visualiza un aumento para el 2025 de casi el doble del que obtuvo en el 2010. Al analizar el crecimiento de la población y el de menores de 15 años, los adultos mayores corresponden a uno por cada dos niños y niñas, visualizándose en el gráfico de la Figura 2 que para el 2025, esta relación sería de 103 adultos mayores por cada 100 menores de 15 años. Esta alza en el envejecimiento de la población se encuentra en diferentes etapas, dependiendo del país que se analice; la proporción de adultos mayores en los países desarrollados es 21 % mayor respecto de los países en vías de desarrollo, quienes están en un 8 %. Según la afirmación de la Organización de las Naciones Unidas (ONU) en 2010, Japón, cuenta con 225 adultos mayores por cada 100 menores de 15 años. Italia y Alemania con cerca de 190 adultos mayores por cada 100 niños menores de 15 años, quienes serán los países con mayor población en condiciones de vejez proyectado para el 2025. A la vez, se proyecta que Japón en el 2025 tendrá cuatro veces más adultos mayores por cada niño/a menor de 15 años en comparación con Chile para el mismo año.

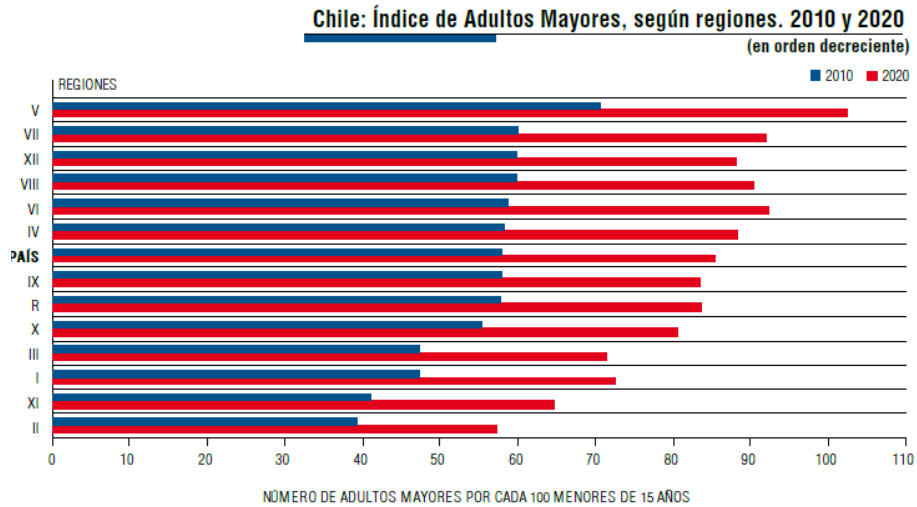
La población de Chile está mayormente conformada por mujeres, existiendo una amplia gama de factores que intervienen en esta afirmación, siendo uno de ellos la menor tasa en la mortalidad, lo que incide en una esperanza de vida superior respecto de los hombres. Las mujeres alcanzan el 16,1% pertenecientes al grupo etario de la tercera edad, mientras que los hombres son el 13,4% pertenecientes a este grupo etario, confirmándose esta información en la Figura 3, en donde también se visualiza cómo desde 1992 se ha mantenido un aumento gradual prolongado de los porcentajes de personas consideradas adultos mayores en el país. Por otro lado, también existe una dependencia demográfica del país que afecta el proceso de transición al envejecimiento, reflejándose en el Índice de Dependencia Demográfica (IDD) que indica el número de personas menores de 15 y de 60 años o más, por cada cien habitantes de entre 15 y 59 años. En 2005 el IDD fue de 57,2 y se espera que para el 2020 alcance el valor de 60, lo que significa, que por cada cien personas activas en 2020 existirán 60 personas en condición de dependientes; además, se visualiza en la Figura 4 el Índice de Adultos Mayores por región, en donde se espera en el peor de los casos, para 2020 de cada 103 niños menores de 15 años existan 86 adultos mayores en el país (INE, 2007).

Figura 3: Porcentaje estimado de la población de adultos mayores desde 1992 al 2015.



Fuente: INE, Enfoque Estadístico Adulto Mayor en Chile (2007).

Figura 4: Cantidad estimada de adultos mayores por cada 100 niños de 15 años en cada región de Chile al 2025.



Fuente: INE, Enfoque Estadístico Adulto Mayor en Chile (2007).

Finalmente, se observa que son varios los temas que involucran y tienen en el centro al adulto mayor tanto en el país como en el mundo, es por ello, que es necesario analizar cualquier área que pueda aportar de forma positiva en las cifras mencionadas anteriormente esperadas para el futuro. Los ciudadanos tendrán que hacer un cambio cultural y social para sobrellevar de forma positiva los cambios que se avecinan y así tener un resultado óptimo como país. Por eso, en este trabajo se investigará “los días de estancia hospitalaria del adulto mayor” acotándose para este trabajo solo a la región de Valparaíso.

---

# INTRODUCCIÓN

---

En el día a día, en diferentes áreas, nace la necesidad de conocer el comportamiento de variables que aporten información al momento de tomar alguna decisión importante. Es por esto que se suele, la mayor parte del tiempo, considerar alguna técnica que permita obtener esta información, como lo es cualquier técnica de regresión convencional, pero de manera lamentable, se suele dejar de lado los supuestos que implican su utilización, los cuales son, la linealidad entre las variables, la no colinealidad entre variables independientes, la homocedasticidad de varianza y la normalidad. De igual manera, se debe tener en cuenta cómo se distribuye la variable que se espera estimar, dado que su forma limita automáticamente el tipo de modelo que se debe utilizar.

Al presentar limitaciones los modelos convencionales, surgen los modelos lineales generalizados, los que no necesariamente se basan en considerar una variable respuesta con distribución Normal, basta que la variable respuesta tenga una distribución que pertenezca a la familia exponencial.

Este trabajo se limitará solamente a explicar y utilizar modelos lineales generalizados atribuidos para variables de conteos, siendo estos definidos como la cantidad de sucesos que ocurren en un momento específico con una misma unidad en un período de tiempo. La definición contiene las principales características de las variables de conteo, naturaleza discreta y no negatividad en sus resultados, cualidades diferenciadoras de las demás variables (McCullagh y Nelder, 1989).

Una de las distribuciones más utilizada para variables de conteo es la distribución de Poisson, que fue presentada la primera vez por el matemático francés Simeón Denis Poisson en 1838 en la investigación sobre la probabilidad de los juicios en materias criminales y civiles (*Recherches sur la probabilité des jugements en ma-*

*tières criminelles et matière civile*, 1838). El modelo de Poisson es uno de los más utilizados en la literatura para modelar variables de conteo, pero el requisito de igualdad de media y varianza de esta distribución dificulta su aplicación en datos reales, dado que éstos, en la mayoría de los casos, presentan una diferencia mayor en el caso de la varianza con respecto a la media, lo que es provocado por diversas razones, una de ellas la alta frecuencia en una determinada cuantía comparado con las demás. A estos datos se les denomina sobre dispersos y motivan al uso de modelos con mayor flexibilidad para que se adapten de mejor manera a los datos (López, 2016).

Luego, la distribución Binomial Negativa nació como una alternativa a la distribución Poisson para acceder a calcular un número de ocurrencia de un determinado evento cuando los datos presentaban sobredispersión. Quien presentó esto por primera vez fue Pierre Rémond de Montmort, el que publicó un estudio sobre los juegos de azar en 1714, pero fue presentada de forma más reciente y oficial por Blaise Pascal (López, 2016).

Este trabajo se dividirá en tres secciones, la primera hace referencia a los modelos lineales generalizados, para conocer cómo se definen y crear conocimientos previos antes de evaluar de manera exhaustiva el modelo que interesa aplicar al final de la investigación (Binomial Negativo). Luego en la segunda sección se revisarán los modelos lineales propuestos por la literatura para variables de conteo siendo el modelo lineal generalizado Poisson y el modelo lineal generalizado Binomial Negativo los considerados en este trabajo. Y para finalizar, con el fin de analizar la utilidad del modelo lineal generalizado Binomial Negativo, se verá una aplicación en datos de egresos de hospitalización del adulto mayor acotados en la región de Valparaíso de 2016 publicados en el sitio oficial del Departamento de Estadísticas e Información de Salud de Chile (DEIS) del MINSAL. El trabajo del análisis de los datos se realizará en el software estadístico Stata versión 14.

---

# MODELOS LINEALES GENERALIZADOS

---

Los modelos estadísticos se proponen para explicar un fenómeno en particular de cierta población a través de un proceso de abstracción, donde se debe asumir la existencia de una variabilidad inherente en el proceso. Con el objetivo de que, a partir de un conjunto de covariables, se logre explicar el comportamiento de una característica particular de los individuos bajo estudio, conocida como variable respuesta o variable dependiente, se busca encontrar un modelo adecuado que describa la relación entre la variable respuesta y las covariables. Para esta elección se debe distinguir el tipo de variables que se desea considerar en el modelo (continuas, de conteo, categóricas, entre otros.) y si la relación de la variable dependiente con las covariables es consistente en la interpretación del análisis del fenómeno, para el cual existe un conjunto de posibles modelos adecuados capaz de explicar la realidad de forma correcta. Colectivamente el modelo más utilizado es el modelo de tipo lineal, donde se expresa la relación de la variable dependiente con las covariables, bajo la certeza de que existe una relación lineal entre ellas; este modelo consiste en expresar la esperanza condicionada de la variable respuesta como una combinación lineal de las covariables bajo el supuesto de homocedasticidad (Stigler, 1981 y Alcaide, 2015).

Los modelos de regresión estándar, como el modelo de regresión lineal, presentan claras deficiencias al momento de ser utilizados analizando distintos tipos de

variables, ya que ignoran la naturaleza de la variable dependiente. Como consecuencia de ello, a través de la literatura estadística propuesta en Nelder (1972) y ampliada en McCullagh y Nelder (1989) que presentan una extensión de la teoría de los modelos lineales, donde se agrega la posibilidad de modelar variables respuestas continuas o categóricas, no exigiéndose necesariamente normalidad en la variable respuesta, sino que se generaliza a cualquier distribución que pertenezca a la familia exponencial. Esto permite considerar distribuciones como Binomial, Poisson, Hipergeométrica, Binomial Negativa entre otras. A esta nueva forma de modelar, se denomina modelo lineal generalizado y se mencionará como MLG, donde se supone que la esperanza de la variable respuesta está relacionada con las covariables a través de una función denominada función de enlace (López, 2016).

## 1.1 Componentes del modelo (Alcaide, 2015)

Son tres componentes importantes en la modelación lineal estadística generalizada, siendo ellas, la componente aleatoria, la componente sistemática y la función de enlace que en conjunto forman el Modelo Lineal Generalizado (MLG).

### 1.1.1 Componente aleatoria

Sea  $Y$ , variable dependiente, que se define como la componente aleatoria del modelo y cuya distribución pertenece a la familia exponencial. Se considera como una variable aleatoria perteneciente a la familia exponencial, definida por los parámetros  $\theta$  y  $\phi$ , cuando se puede presentar de la siguiente forma:

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad ; \quad (1.1)$$

donde  $f()$  hace referencia a la función de probabilidad en el caso de que  $Y$  sea una variable discreta, o función de densidad en el caso de que  $Y$  sea una variable continua;  $\theta$  es el parámetro de forma, denominado canónico y  $\phi$  es el parámetro de escala. Entonces, se definen  $a(\phi)$ ,  $b(\theta)$  y  $c(y, \phi)$  como funciones específicas de cada elemento de la familia exponencial.

En un MLG se asume que la variable dependiente  $Y$ , como se mencionó anteriormente, está generada a partir de una función de distribución de la familia exponencial, donde la media ( $\mu$ ) existe como se expresa a continuación:

$$E(Y) = \mu \quad . \quad (1.2)$$

Luego, la varianza se denota como se expresa en la siguiente formula, siendo ésta no más que una función de la media, se denomina función de varianza denotada por función  $V$ , capturando la relación entre  $E(Y)$  y la  $\text{Var}(Y)$ .

$$\text{Var}(Y) = V(\mu) \quad . \quad (1.3)$$

### 1.1.2 Componente sistemática (López, 2016 y Contreras, 2012)

La componente sistemática de un modelo lineal generalizado describe a las covariables que ingresan en forma de efecto fijo al modelo, capturando la variabilidad de  $Y$  a través de las  $p$  covariables  $(X_1, \dots, X_p)$  denotado por  $X$  y sus correspondientes parámetros  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^\top$ , expresándose de manera generalizada de la siguiente forma.

$$\eta_i = X\beta \quad . \quad (1.4)$$

Cabe mencionar que la componente sistemática del MLG puede incluir términos tales como las covariables originales, potencias y transformaciones de estas variables, considerando la función de enlace que se utilizará.

### 1.1.3 Función de enlace

Hay una necesidad de encontrar una función que relacione el valor esperado ( $\mu$ ) con la componente sistemática ( $X\beta$ ). Y de esta forma nace la función de enlace y se simboliza como  $g(\mu_i)$ .

En un modelo con distribución gaussiana se tiene lo siguiente:

$$\mu = \eta \quad . \quad (1.5)$$

Para llegar a la expresión anterior de forma general, se introduce la tercera componente del MLG, la función de enlace escrita de la siguiente manera:

$$\eta_i = g(\mu_i) \quad . \quad (1.6)$$

Es importante mencionar que para cada distribución de la variable dependiente existe una función de enlace canónica que relaciona el parámetro canónico directamente con el predictor lineal (componente sistemática) cumpliendo con las restricciones naturales que posee la variable respuesta. En este caso, siempre positiva (McCullagh y Nelder, 1989).

Así, para las distribuciones de la familia exponencial de interés en este trabajo, se tiene una función de enlace canónica como se expresa a continuación:

### Binomial negativa

$$\theta(\mu_i) = \ln\left(\frac{\alpha\mu_i}{1-\alpha\mu_i}\right) \quad , \quad \theta^{-1}(\eta_i) = \left(\frac{1}{\alpha(\exp(-\eta_i)-1)}\right) \quad ;$$

siendo  $\alpha = 1/r$ .

### Poisson

$$\theta(\mu_i) = \log(\mu_i) \quad , \quad \theta^{-1}(\eta_i) = \exp(\eta_i).$$

Los respectivos MLG bajo una variable aleatoria con distribución Poisson o con distribución binomial negativa al aplicarle una función de enlace logarítmica quedan expresados respectivamente de la siguiente manera:

- **MLG Poisson:** al considerar la función de enlace como el canónico se obtiene lo siguiente:

$$\log(\mu_i) = \eta_i = X_i\beta \quad \text{ó} \quad \mu_i = \exp(X_i\beta) \quad .$$

- **MLG Binomial Negativo:** al considerar la función logarítmica como función de enlace se obtiene lo siguiente:

$$\begin{aligned} &\text{Logaritmo } g(\mu_i) = \log(\mu_i) \quad ; \\ &\log(\mu_i) = \eta_i = X_i\beta \quad \text{ó} \quad \mu_i = \exp(X_i\beta) \quad . \end{aligned}$$

Luego, es necesario realizar el cálculo de los estimadores desconocidos de  $\beta$  para cada modelo mencionado anteriormente, existiendo diferentes técnicas para este proceso. Pero se debe tener en cuenta que existen técnicas con mayor eficiencia que otras; es por esto, que a continuación se explica una técnica en específico para el cálculo de los estimadores.

## 1.2 Estimación de los parámetros del modelo (López, 2016)

Para cada uno de los dos modelos especificados anteriormente, es necesario estimar los parámetros de la componente sistemática y posteriormente se valora la precisión de esas estimaciones a través del cálculo de la discrepancia entre pares de modelos, con el fin de seleccionar el modelo óptimo.

El método mejor de estimación estadística de parámetros para el MLG según Vasconcellos, Cordeiro y Barroso en 2000, es el método de Máxima Verosimilitud (MV), ya que es el más adecuado para encontrar estimadores que posean las propiedades de consistencia y eficiencia asintótica exigibles (Alcaide, 2015).

Entonces, si la función de distribución de la variable respuesta  $Y$  corresponde a una familia de distribuciones  $H$  conocida y si se tiene un vector de observaciones dado por  $Y^\top = (Y_1, \dots, Y_n)$ , la función de verosimilitud considera la posibilidad de que un vector  $\beta \in R^p$  otorgue el vector de respuestas observado (González, 2001).

La función de verosimilitud expresada por la función de densidad conjunta de las variables aleatorias independientes  $Y_1, \dots, Y_n$  se reduce a la forma:

$$L(\beta) = h(y_1, \dots, y_n) = h(y_1, \beta)h(y_2, \beta)\dots h(y_k, \beta) = \prod_{i=1}^n h(y_i; \beta) \quad . \quad (1.7)$$

El estimador de máxima verosimilitud de  $\beta$ , corresponde al vector  $b$  el cual maximiza  $L(\beta)$  en el espacio paramétrico  $\Omega = \{(\beta, \sigma^2) : \beta \in R^p, \sigma^2 > 0\}$ ; esto es  $L(b) \geq L(\beta), \forall \beta \in \Omega$ .

Para poder obtener el EMV, se maximiza  $L(\beta)$  para  $\beta \in R^p$ .

Dado que la función logaritmo es monótona, aplicando logaritmo a la expresión (1.7) se obtiene lo siguiente:

$$l(\beta) = \log L(\beta) = \sum_{i=1}^n \log h(y_i; \beta) \quad . \quad (1.8)$$

Entonces, si la función  $l(\beta)$  es derivable y continua, maximizar  $L(\beta)$  o  $l(\beta)$  es

equivalente.

Luego, es necesario mencionar que en algunos casos, cuando se estiman parámetros de una familia conocida de distribuciones con máxima verosimilitud, se producen resultados más confiables en comparación a los obtenidos con otros métodos de estimación (método de los momentos), siendo sus resultados más precisos y menos sesgados. Además, en algunas oportunidades los resultados de las estimaciones están por fuera del espacio paramétrico, lo que nunca sucede con el método de MV.

### 1.3 Adecuación del modelo (Alcaide, 2015).

Una vez calculados los parámetros, se debe observar el grado de discrepancia entre los datos observados y los datos esperados del modelo. Una diferencia pequeña entre los datos observados y el conjunto de valores estimados  $\hat{\mu}$ , puede ser tolerable, pero en cuanto a una discrepancia grande no lo es.

De esta forma, si se admite una combinación satisfactoria para la distribución de la componente aleatoria y la función de enlace, es necesario determinar la cantidad de términos que son necesarios en esta estructura lineal para una descripción óptima de los datos observados. Un número amplio de variables explicativas puede llevar a que un modelo explique razonablemente los datos, pero con un aumento considerable en la complejidad de su interpretación. Por otro lado, un número pequeño de variables explicativas puede llevar a un modelo de fácil interpretación, pero con un ajuste escaso de los datos; es por esto que, se busca un modelo intermedio.

Para determinar un buen modelo se debe siempre tener presente la finalidad que desea el investigador. Si la finalidad del modelo es del tipo predictivo, se seleccionan las variables que expliquen el mayor grado de variabilidad de la variable respuesta, aplicándose los criterios estadísticos que serán evaluados más adelante en este trabajo. Y, cuando la finalidad del modelo es explicativa, son los argumentos teóricos los que deben tomar un mayor protagonismo, guiado por el investigador basándose en una especificación de un modelo máximo inicial y de un conjunto sucesivo de modelos restringidos que se comparan mediante un ajuste condicional (Alcaide, 2015).

Al tener en cuenta lo que se ha mencionado anteriormente, el objetivo del proceso de modelamiento es la obtención de un modelo que sea capaz de representar los

datos  $y$ , al mismo tiempo reducir la complejidad. En otras palabras, atender los criterios de **bondad de ajuste**.

### 1.3.1 Bondad de ajuste (Alcaide 2015).

Un modelo lineal generalizado puede evaluar de distintas formas la bondad de ajuste, entre las que destacan las siguientes:

- **Devianza residual:**

$$D(y, \hat{\mu}) = 2 \{l(y; y) - l(\hat{\mu}; y)\} \quad . \quad (1.9)$$

Es simplemente la distancia entre el logaritmo de la función de verosimilitud del modelo saturado (con  $n$  parámetros, uno por observación) y el modelo que se está implementando (con  $p$  parámetros). Si la diferencia es pequeña, se indica que un número menor de parámetros es igual de bueno el ajuste que el modelo saturado.

Luego, para examinar la adecuación de un modelo lineal generalizado, el valor de la diferencia se compara con el percentil de alguna distribución de probabilidad referente. Al ser correcto el modelo el estadístico se distribuye asintóticamente según una  $\chi_{n-p}^2$ , con  $n - p$  grados de libertad (McCullagh y Nelder, 1989).

$$D(y, \hat{\mu}) \sim \chi_{n-p}^2 \quad , \text{ cuando } \mu \rightarrow \infty \quad . \quad (1.10)$$

- **Coefficiente de determinación  $R^2$ :**

La medida  $R^2$  se define como la reducción proporcional en la incertidumbre debido a la inclusión de las covariables, que explica la proporción de variación de la variable respuesta al ser explicada por la covariables.

Viene expresado de la siguiente forma:

$$R^2 = 1 - \frac{D(y; \hat{\mu})}{D(y; \hat{\mu}_0)} \quad ; \quad (1.11)$$

donde,  $D(y; \hat{\mu})$  es la devianza residual del modelo ajustado y  $D(y; \hat{\mu}_0)$  es la función de desviación del modelo simple que se usa como referencia, y se verifica que  $0 \leq R^2 \leq 1$ .

- **Estadístico chi-cuadrado de Pearson:**

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad ; \quad (1.12)$$

donde,  $V(\hat{\mu}_i)$  es una función de la varianza estimada para la distribución de la variable respuesta.

Se han mencionado criterios de comparación que se basan en la máxima verosimilitud que permiten obtener un mejor modelo que ajusta de mejor manera los datos, pero estos no sancionan su complejidad en el aumento de la cantidad de parámetros en el modelo, lo que sí sucede cuando se utilizan medidas de contraste como lo son el criterio *AIC* y *BCI*, que serán explicados a continuación:

- **Criterio de información de Akaike (AIC):**

Este criterio fue planteado por Akaike en 1974, haciendo referencia a la mezcla entre la teoría de máxima verosimilitud, con la información teórica y la entropía de información, tomando en cuenta los distintos cambios en la bondad de ajuste (López, 2016).

En el caso general el AIC es:

$$AIC = k - 2\ln(\hat{L}) \quad ; \quad (1.13)$$

siendo,  $k$  el número de parámetros del modelo estadístico estimado y  $\hat{L}$  es el máximo valor de la función de verosimilitud del modelo considerado. Luego, de un conjunto de modelos candidatos para los datos, se escoge el modelo que posea el menor AIC, dado que, éste criterio posee un componente

penalizador a través de una función creciente, el cual, penaliza el aumento del número de parámetros estimados, aumentando el valor que arroja como resultado.

- **Criterio de información bayesiano (BIC):**

$$BIC = k \ln(n) - 2 \ln \hat{L} \quad ; \quad (1.14)$$

siendo  $k$  el número de parámetros libres a ser estimados,  $n$  es el tamaño de la muestra y  $\hat{L}$  es el máximo valor de la función de verosimilitud. Es posible aumentar el ajuste del modelo con la adición de parámetros, pero si se hace puede resultar en sobreajuste. Tanto el BIC y AIC resuelven este problema mediante la introducción de un término de penalización para el número de parámetros en el modelo, este término es mayor en el BIC que en el AIC.

En ambos criterios, el modelo que resulte con el menor valor de AIC y BIC será el más adecuado para explicar de mejor manera los datos, utilizando la menor cantidad de posibles parámetros.

#### 1.4 **Análisis de los residuos** (Osses, 2013)

En la práctica, puede ocurrir que aun escogiendo cuidadosamente un modelo y luego al ajustarlo a un conjunto de datos, entregue resultados insatisfactorios. Esto genera desviaciones sistemáticas que se originan por haber escogido una función de enlace inadecuada o incluir covariables erróneas en el modelo. Las discrepancias aisladas pueden ocurrir debido a datos extremos, que pueden ser recolecciones de los datos de forma errónea o factores externos que no fueron considerados en la recolección de los datos. Es un requisito verificar la adecuación del modelo para analizar posibles desviaciones de las suposiciones hechas para el modelo (varianza constante, linealidad, entre otros), así como la existencia de observaciones extremas con alguna interferencia en los resultados del ajuste (Alcaide, 2015).

Por otro lado, generalmente si la componente sistemática tiene alguna interpretación, como en el caso de la distribución Gamma o de la distribución Poisson,

se puede reducir devianza residual a lo siguiente:

$$D(\beta) = \sum_{i=1}^n \left[ y_i \ln \left( \frac{y_i}{\mu_i} \right) \right] . \quad (1.15)$$

Existen cuatro maneras apropiadas para trabajar los residuos de los modelos lineales generalizados, expresados a continuación:

- **Residuos de Pearson:**

$$R_L(y, \theta) = \frac{y - E_\theta(y)}{SD_\theta(y)} ; \quad (1.16)$$

donde,  $E_\theta$  es la media y  $SD_\theta$  es la desviación estándar de los residuos.

- **Transformación específica:**

$$R_T(y, \theta) = \frac{t(y) - E_\theta[t(y)]}{SD_\theta[t(y)]} ; \quad (1.17)$$

donde, la función  $t(\cdot)$  es una transformación específica, normalmente elegida dependiendo de la distribución particular de  $y$ .

- **Devianza residual:**

$$R_D(y, \theta) = sgm(\hat{\theta} - \theta) 2 \left[ l(\hat{\theta}, y) - l(\theta, y) \right]^{\frac{1}{2}} ; \quad (1.18)$$

donde,  $\hat{\theta}$  es el estimador de máxima verosimilitud de  $\theta$ , sobre la base de  $y$ , sin restricción alguna para el modelo  $\theta_i = g(x, \beta)$  y,  $l(\theta, y) = \log f(y, \theta)$  es el registro de la función de verosimilitud.

La desvianza residual es de gran importancia para poder realizar un análisis residual, ya que entrega una idea de la variabilidad de los datos y como

ésta es explicada por el modelo. Además, Pierce y Shafer en 1986, citado en Osses (2013), sugieren trabajar con la desviación residual, cuando se trabaje con distribuciones de la familia exponencial, debido que el resultado es muy similar al que genera una excelente transformación para conseguir la normalidad.

- **Residuo de Anscombe:**

El residuo de Anscombe descrito en Anscombe en 1972 citado por Alcaide (2015) toma valores cercanos a los de la desviación estandarizada. Pero, el residuo de Anscombe realiza una mejor interpretación que la devianza residual. Éste trata de normalizar la diferencia entre los valores observados y ajustados de manera que la heterogeneidad en los datos es identificable.

Los residuos de Anscombe usan la función de varianza del modelo, definiéndose de la siguiente forma:

$$r_i^A = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}} \quad ; \quad (1.19)$$

donde,

$$A(\cdot) = \int d\mu_i / V^{\frac{1}{3}}(\mu_i) \quad . \quad (1.20)$$

Además, se debe tener en cuenta que si se escoge mal la función de varianza puede producir una tendencia errónea en los residuos.

## 1.5 Interpretación (Alcaide, 2015)

Una vez obtenido el modelo adecuado para los datos en estudio, haciendo uso de los criterios antes descritos en este capítulo, el proceso finaliza con la interpretación del modelo.

Al momento de interpretar el modelo se debe tener en cuenta la transformación que se realiza al aplicar una función de enlace, que provoca en la mayoría de los casos una ecuación del modelo expresada en términos multiplicativos. Por esto, hay que tener cuidado en la interpretación del modelo que resulte, debido a que ésta no es directa por la transformación que realiza la función de enlace.

Se estudiará en el siguiente capítulo los modelos para variables de conteo. Los más utilizados en la práctica son el modelo de Poisson y el modelo Binomial Negativo.

---

# MODELOS LINEALES GENERALIZADOS POISSON Y BINOMIAL NEGATIVO

---

Las variables de recuento cuentan con una amplia presencia en diversos ámbitos de la investigación tanto en las ciencias sociales, descrito en Merkleesson y Roth en 2000, como en las ciencias de la salud, descrito en Lindsey en 2001 pero, existiendo un aporte con mayor significancia para el tratamiento estadístico que se realiza para este tipo de variables en el área de la medicina, siendo uno de ellos Böhning en 1994, también, las ciencias políticas descrito en King en 1989 y ciencias económicas realizado por Meliciani en 2000, mencionados en Alcaide (2015).

Este capítulo se centrará en dos modelos lineales generalizados para variables de recuento o conteo; el modelo lineal generalizado Poisson y el modelo lineal Binomial Negativo.

## 2.1 Datos de recuento (Alcaide, 2015)

Se denominan variables de recuento a aquellas variables que determinan el número de sucesos o eventos que ocurren en una misma unidad de observación, en un intervalo temporal o espacial definido.

A partir de esta definición, se derivan dos características principales de una variable de conteo: poseen una naturaleza discreta y no negatividad en sus resultados.

La variable  $Y$ , que toma los valores  $0,1,2,\dots$ , se caracteriza por tomar valores numéricos infinitos que se pueden ordenar de manera creciente, y cuya probabilidad de ocurrencia va en descenso a medida que el valor de la variable aumenta.

No todos los modelos son aplicables para las variables de conteo, pues pueden originarse ciertos problemas como pérdida de información o problemas de inconsistencia en sus resultados, es por ello, que se estudiarán a continuación dos modelos lineales generalizados, que explican de forma adecuada este tipo de variables.

## 2.2 Modelo lineal generalizado Poisson (MLGP) (López, 2016)

La distribución de Poisson es la distribución habitualmente utilizada en procesos de conteo y el modelo lineal generalizado Poisson es un modelo ideal para modelar valores enteros no negativos, principalmente cuando la frecuencia de que ocurra un determinado proceso es baja.

Se dice que una variable  $Y_i$  sigue el MLGP si cumple lo siguiente:

$$Y \sim P(\mu_i), \quad i = 0, 1, 2, \dots \quad ; \quad (2.1)$$

con función de probabilidad dada por:

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots \quad ; \quad (2.2)$$

de parámetro  $\mu > 0$ .

Y, la media con la varianza, están dadas por:

$$E(Y) = \mu \quad ;$$

$$V(Y) = \mu \quad .$$

Esta igualdad de media y varianza se nombra como la equidispersión de la distribución de Poisson.

Así, el modelo se define como:

$$g(\mu_i) = \eta_i = X_i^\top \beta \quad ; \quad (2.3)$$

donde,

- $X_i = (X_{i1}, \dots, X_{ip})^\top$  corresponde al vector de covariables explicativas.
- $\beta = (\beta_0, \dots, \beta_p)^\top$  corresponde al vector de parámetros desconocidos.

Las componentes en las que se divide el MLGP son:

### 2.2.1 Componente aleatoria

Dado  $Y_i, \dots, Y_n$  un vector de la variable respuesta positiva y sea  $X_i = (X_{i1}, \dots, X_{ip})^\top$  un vector de covariables explicativas con parámetro  $\mu_i$ , se define lo siguiente:

$$Y \sim P(\mu_i), \quad i = 0, 1, 2, \dots, \quad . \quad (2.4)$$

### 2.2.2 Componente sistemática

Dado  $\mu_i$  y un predictor lineal representado a continuación:

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} \\ &= \beta_0 + \sum_{j=1}^p \beta_j X_{ij} \\ &= X_i^\top \beta \end{aligned} \quad (2.5)$$

### 2.2.3 Función de enlace

Mediante la elección de una función de enlace, ambos componentes mencionados se combinan en el modelo. expresado de la siguiente forma:

$$g(\mu_i) = \eta_i \quad . \quad (2.6)$$

La función de enlace canónico para el modelo MLGP es:

$$\log(\mu_i) = \eta_i \quad . \quad (2.7)$$

## 2.2.4 Estimación de máxima verosimilitud

Sea  $Y_1, \dots, Y_n$  un conjunto con  $n$  observaciones aleatorias e independientes donde el predictor es  $X$ , entonces la función de verosimilitud es expresada como sigue:

$$\frac{\prod_{i=1}^n \mu_i^{y_i} \exp(-\sum_{i=1}^n \mu_i)}{\prod_{i=1}^n y_i!} \quad ; \quad (2.8)$$

donde  $\mu_i = g^{-1}(X^\top \beta)$

- **Devianza residual:**

$$\begin{aligned} D(y; \hat{\mu}) &= 2\{l(y; y) - l(\hat{\mu}; y)\} \quad , \\ &= 2 \sum_{i=1}^n \{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\} \quad . \end{aligned} \quad (2.9)$$

- **Coeficiente de determinación  $R^2$ :**

$$\begin{aligned} R^2 &= 1 - \frac{D(y; \hat{\mu}_i)}{D(y; \hat{\mu}_0)} \quad ; \\ &= \frac{\sum_{i=1}^n \{y_i \log(\hat{\mu}_i / y_i) - (\hat{\mu}_i - y_i)\}}{\sum_{i=1}^n \{y_i \log(y_i / \bar{y}_i)\}} \quad . \end{aligned} \quad (2.10)$$

- **Estadístico chi-cuadrado de Pearson:**

$$\begin{aligned} \chi^2 &= \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad ; \\ &= \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \quad . \end{aligned} \quad (2.11)$$

- **Residuo de Pearson:**

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \quad \text{con } i = 1, 2, \dots, n \quad . \quad (2.12)$$

- **Residuo de Pearson studentizado:**

$$r_i^p = \frac{y_i - \hat{\mu}_i}{(\sqrt{\hat{\mu}_i})(1 - h_i)} \quad \text{con } i = 1, 2, \dots, n \quad ; \quad (2.13)$$

donde,  $h_i$  es el  $i$ -ésimo elemento de la matriz de proyección.

- **Residuo de Anscombe:**

Los residuos Anscombe usan la función de varianza del MLGP.

$$Poisson : V(\mu) = \mu \quad . \quad (2.14)$$

Este residuo se denota de la siguiente forma :

$$r_i^A = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}} \quad ;$$

donde,

$$A(\cdot) = \int d\mu_i / V^{\frac{1}{3}}(\mu_i) \quad .$$

El MLGP cumple la característica de modelo heterocedástico, en donde la varianza de las perturbaciones no es constante, por lo tanto la variabilidad es diferente para cada observación.

Sin duda, es cierto que el MLGP presenta mejoras con respecto a los modelos ordinarios, pero puede resultar inapropiado en otros aspectos, debido a la violación de ciertos supuestos importantes, siendo uno de ellos la sobredispersión descrita en Winkelmann en 2000, citado en Alcaide (2015) y López (2016), en donde se especifica que se rechazan sistemáticamente ciertas admisiones consideradas en el MLGP.

**La sobredispersión** acontece cuando  $Var(Y) > E(Y)$ , es decir cuando el cociente entre estos dos, es distinto de 1. Si se está en el caso de un exceso de dispersión en los datos, podrían resultar sesgadas las estimaciones de los errores estándar, lo que podría causar errores en las inferencias de los parámetros del modelo de regresión (Alcaide, 2015).

Entre las diversas causas de sobredispersión destacan (Alcaide, 2015):

- Alta variabilidad en los datos.
- Los datos no provienen de una distribución Poisson.
- Los eventos no ocurren independientemente a través del tiempo.
- Falta de estabilidad, es decir, la probabilidad de ocurrencia de un evento puede ser independiente de la ocurrencia de un evento previo, pero no es constante.
- Errores de especificación de la media  $\mu$ , como por ejemplo omitir variables explicativas o que entran al modelo a través de alguna transformación en lugar de linealmente explicado en Winkelmann en 2000 (Alcaide, 2015).
- Errores al elegir la función de enlace, es decir, tal vez no fue apropiado elegir el enlace log-lineal.

Existen diferentes métodos para descubrir una sobredispersión en los datos, siendo uno de ellos el coeficiente de variación (CV) explicado en Lindsey en 1995,

citado por López (2016) y Alcaide (2015), y se plantea como sigue:

$$CV = \frac{V(\mu_i)}{\mu_i} . \quad (2.15)$$

donde,  $V(\cdot)$  hace referencia a la función de varianza de la distribución evaluada.

Este  $CV$ , debiese ser igual a 1, siempre y cuando se cumpla el supuesto de equidispersión.

Generalmente, la sobredispersión se evalúa mediante la relación entre el estadístico de Pearson  $\chi^2$  o la función desviación  $D$  y sus respectivos grados de libertad ( $gl$ ), es decir calcular.

$$\frac{\chi^2}{gl} \quad o \quad \frac{D}{gl}; \quad (2.16)$$

si estos valores son mayores que 1, indican sobredispersión.

También, existe un diagnóstico basado en una prueba de Razón de Verosimilitud (RV) apoyada en las distribuciones de Poisson y binomial negativa.

- En el caso de la distribución de Poisson  $Var(Y) = \mu$ .
- En el caso de la distribución Binomial Negativa  $Var(Y) = \mu + k\mu^2$ .

Si el parámetro  $k = 0$ , es el caso donde una distribución Binomial Negativa se reducirá a una distribución Poisson.

Entonces la hipótesis que se plantean son:

$$H_0 : k = 0 \quad v/s \quad H_1 : k > 0 .$$

Para llevar a cabo esta prueba, se ajustan los 2 modelos: el Poisson y Binomial Negativo (BN). Para cada modelo se deberá obtener su respectiva función de log-verosimilitud, denotada por  $l$ .

El estadístico propuesto es:

$$RV = -2(l(Poisson) - l(BN)) \quad . \quad (2.17)$$

Cameron y Triveli citado por (Alcaide,2015), dicen que esta prueba tiene una distribución asintótica  $\chi^2_{(1-2\alpha,1)}$ . Por lo tanto, se rechazará si el estadístico es mayor que  $\chi^2_{(1-2\alpha,1)}$ . De resultar esta situación es más conveniente modelar el número de ocurrencias a través de una distribución binomial negativa.

Luego, a pesar de que el MLGP es el modelo de referencia en los estudios de variables de recuento, resultando adecuado para modelar valores enteros no negativos, especialmente cuando la frecuencia de ocurrencia es baja, presentan varios problemas al momento de trabajar con datos en donde la media y la varianza no coinciden. De darse esta situación, es más apropiado ajustar el número de ocurrencias por medio del modelo lineal generalizado Binomial Negativo, donde la interpretación es igual que en el caso del modelo lineal generalizado Poisson.

### 2.3 Modelo lineal generalizado Binomial Negativo (MLGBN) (López, 2016)

Una de las razones de porqué el modelo de Poisson falla, es debido a la heterogeneidad no observada. Es decir, hay factores no observados en los datos, que son característicos de los sujetos bajo estudio, y éstos causan cierta influencia sobre la variabilidad que se asocia a la variable dependiente.

La dificultad que existe, es que la heterogeneidad no observada podría tener ciertas consecuencias en los procesos de inferencia estadística. Existiendo sobredispersión y un número amplio de sucesos registrados con ceros. Entonces la heterogeneidad no observada con la sobredispersión existente no capturada por el modelo lineal generalizado Poisson, puede modelarse a través de un modelo más flexible, siendo el modelo paramétrico estándar para datos de recuento con presencia de estos dos factores mencionados, MLGBN.

Se define la distribución binomial negativa como aquella variable que estudia la probabilidad de observar un número determinado de fracasos, hasta el  $r$ -ésimo éxito en una serie de experimentos Bernoulli independientes. La descripción de  $r$ , es que es un entero positivo. Además, existen al menos dos formas que permiten llegar a la distribución binomial negativa: la más utilizada asume que se está en presencia de una mezcla de distribuciones, en donde las observaciones se distribuyen como una distribución de Poisson, pero se presume un elemento de heterogeneidad individual no observado (que sigue una distribución Gamma en su fórmula tradicional), manifestando que la verdadera media no ha sido medida correctamente. La otra forma, asume que hay una manera específica de dependencia entre sucesos, de modo que la ocurrencia de un evento aumenta la probabilidad de ocurrencia de eventos posteriores, pero también podría tratarse de datos longitudinales.

Elementos que conforman el MLGBN son:

### 2.3.1 Componente aleatoria

Sea  $Y_i, \dots, Y_n$  una variable aleatoria independiente la cual hace referencia al número de sucesos necesarios para poder obtener  $r$ -éxitos. Es decir, el número de éxitos está establecido y la aleatoriedad es el número de eventos de interés, de modo que la componente aleatoria establecida es:

$$Y_i \sim BN(\mu_i, \phi), \quad i = 0, 1, 2, 3, \dots \quad .$$

### 2.3.2 Componente sistemática

Dado  $\mu_i$ , con el predictor lineal representado por:

$$\eta_i = X_i^\top \beta \quad .$$

### 2.3.3 Función de enlace

Ambas componentes presentadas anteriormente se combinan en el modelo, a través de la elección de la función de enlace:

$$g(\mu_i) = \eta_i \quad ;$$

donde,  $g(\cdot)$  es una función de enlace.

La función de enlace canónica para este modelo es la logarítmica, ya que cumple con las restricciones de la naturaleza de la variable respuesta:

$$\log(\mu_i) = \eta_i \quad . \quad (2.18)$$

Es importante tener presente que, este modelo puede ser usado en presencia de sobredispersión, es decir, en casos en los que  $Var(Y) \geq E(Y)$ .

Su representación tradicional es obtenida mediante la incorporación de un término de perturbación en la medida del modelo Poisson, una aleatoriedad en el parámetro  $\mu_i$ .

$$\mu_i^* = \exp(x_i\beta + \varepsilon_i) = \mu_i \exp(\varepsilon_i) \quad (2.19)$$

donde el término de perturbación  $\varepsilon_i$  sigue una distribución Gamma. Siendo su función de probabilidad:

$$P(Y = y_i/x_i) = \frac{\Gamma(y_i + v_i)}{\Gamma(y_i + 1)\Gamma(v_i)} \left( \frac{v_i}{v_i + \mu_i} \right)^{v_i} \left( \frac{\mu_i}{v_i + \mu_i} \right)^{y_i} \quad (2.20)$$

con  $\mu_i = E(Y) = \exp(X_i\beta)$  y  $v_i = (1/\alpha)\mu_i^t$  donde  $t = 0, 1$

La especificación final depende de cómo se define  $v$ .

A continuación, como se describe en el primer capítulo, se debe hacer el cálculo de los estimadores de los parámetros de MLGBN.

### 2.3.4 Estimación de máxima verosimilitud

Se considera la siguiente partición  $\theta = (\beta^t, \phi)^t$ , la cual denota el logaritmo de la función de verosimilitud (López 2016).

Si se denota la función de probabilidad como sigue:

$$P(Y = y) = \frac{\Gamma(\phi + y)}{\Gamma(y + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi}\right)^y \left(\frac{\phi}{\mu + \phi}\right)^\phi, \quad y = 0, 1, 2, \dots ;$$

se obtiene el logaritmo de la función de verosimilitud de la siguiente forma:

$$L(\theta) = \sum_{i=1}^n \left[ \log \left\{ \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \right\} + \phi \log \phi + y_i \log \mu_i - (\phi + y_i) \log(\mu_i + \phi) \right]; \quad (2.21)$$

donde  $\mu_i = g^{-1}(X_i^\top \beta)$ , es una función *score* para  $\beta$ .

Se calculan primeramente las derivadas para la función *score* de  $\beta$ .

$$\begin{aligned} \frac{\partial L(\theta)}{\partial \beta_j} &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{\beta_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} x_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} x_i \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\phi \left( \frac{d\mu_i}{d\eta_i} \right)}{\mu_i(\phi + \mu_i)} (y_i - \mu_i) x_{ij} \right\} \\ &= \sum_{i=1}^n w_i f_i^{-1}(y_i - \mu_i) x_{ij} \quad ; \end{aligned} \quad (2.22)$$

donde,

$$w_i = \frac{(d\mu_i/d\eta_i)^2}{(\mu_i^2 \phi^{-1} + \mu_i)} \quad . \quad (2.23)$$

Luego se expresa la función *score* en forma matricial para  $\beta$  de la siguiente forma:

$$U_\beta(\theta) = X^t W F^{-1}(y - \mu) \quad ; \quad (2.24)$$

con,

- $X$ : matriz del modelo lineal  $x_i^2$  con  $i = 1, \dots, n$ .
- $W = \text{diag}(w_1, \dots, w_n)$  con  $w_i = \frac{(d\mu_i/d\eta_i)^2}{(\mu_i^2 \phi^{-1} + \mu_i)}$  .
- $F = \text{diag}(f_1, \dots, f_n)$  con  $f_i = \frac{d\mu_i}{d\eta_i}$  .
- $y = (y_1, \dots, y_n)^t$  .
- $\mu_i = (\mu_1, \dots, \mu_n)^t$  .

Lo mismo se puede expresar para la función *score* de  $\phi$ , de la siguiente manera.

$$U_\phi(\theta) = \sum_{i=1}^n \left[ \psi(\phi + y_i) - \psi(\phi) - \frac{(y_i + \phi)}{(\phi + \mu_i)} + \log \left\{ \frac{\phi}{(\phi + \mu_i)} \right\} + 1 \right] \quad ; \quad (2.25)$$

donde,  $\psi(\cdot)$  en una función digama (derivada logarítmica de la función gama).

Para poder obtener la matriz de información de Fisher se debe calcular las derivadas.

$$\frac{\partial^2 L(\theta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n \left\{ \frac{(\phi + y_i)}{(\phi + \mu_i)^2} - \frac{y_i}{\mu_i} \right\} \left( \frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{il} + \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \right\} \frac{d^2 \mu_i}{d\eta_i^2} x_{ij} x_{il};$$

donde, los valores esperados están dados por:

$$\begin{aligned} E \left\{ \frac{\partial^2 L(\theta)}{\partial \beta_j \partial \beta_l} \right\} &= - \sum_{i=1}^n \frac{\phi (d\mu_i/d\eta_i)^2}{(\phi + \mu_i)} x_{ij} x_{il} \\ &= - \sum_{i=1}^n w_i x_{ij} x_{il} \quad . \end{aligned} \quad (2.26)$$

Luego se puede expresar la información de Fisher para  $\beta$ , en forma matricial de la siguiente manera:

$$k_{\beta\beta}(\theta) = E \left\{ \frac{\partial^2 L(\theta)}{\partial \beta \partial \beta^t} \right\} = X^t W X \quad . \quad (2.27)$$

Lawless en 1982 señaló que la información de Fisher para  $\phi$  se puede expresar como sigue (López, 2016):

$$k_{\beta\beta}(\theta) = \sum_{i=1}^n \left\{ \sum_{j=1}^n (\phi + j)^2 Pr(Y \geq j) - \phi^{-1} \mu_i / (\mu_i + \phi) \right\} \quad ; \quad (2.28)$$

donde,  $\beta$  y  $\phi$  corresponden a dos parámetros ortogonales. Por lo tanto, la matriz de información de Fisher para  $\theta$  asume una forma de bloque diagonal.

$$k_{\theta\theta} = \begin{bmatrix} k_{\beta\beta} & 0 \\ 0 & k_{\phi\phi} \end{bmatrix} \quad . \quad (2.29)$$

La estimación de máxima verosimilitud para  $\theta$  y  $\phi$  puede ser conseguida mediante un algoritmo de mínimos cuadrados ponderados para obtener  $\hat{\theta}$  desarrollado a partir del punto (2.24) y el método de Newton-Raphson para obtener  $\hat{\phi}$ , desarrollado a partir del punto (2.25) el cual se expresa como sigue:

$$\beta^{m+1} = (X^t W^m X)^{-1} X^t W^m y_*^{(m)} \quad ; \quad (2.30)$$

y,

$$\phi^{m+1} = \phi^m - \left\{ \frac{U_{\phi}^m}{L_{\phi\phi}} \right\} \quad . \quad (2.31)$$

Para  $m = 0, 1, 2, 3, \dots$ , en la que:

$$y_* = X\beta + F^{-1}(y - \mu) \quad . \quad (2.32)$$

### 2.3.5 Método Newton Raphson (López, 2016).

Este método permite encontrar aproximaciones de los ceros o las raíces de una función real, aplicado por Verdin en 2015 (López, 2016):

Entonces, sea  $\bar{x} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n]^t$  lo cual corresponde a una raíz del sistema no lineal de tamaño  $n \times n$ , de la ecuación  $f(x) = 0$ , cuya  $i$ -ésima ecuación está dada por:

$$f_i(x) = f_i(x_1, \dots, x_n) = 0, \quad i = 1, \dots, n \quad . \quad (2.33)$$

Se supone que  $x_k$  corresponde a una aproximación presente de  $\bar{x}$ . La manera en que se puede obtener una aproximación corregida  $x_{k+1}$  será resolviendo un sistema lineal que se aproxime al sistema (2.33) para que  $x$  esté cerca de  $x_k$ . Puntualmente, si  $x = x_k + dx$ , donde  $dx = [dx_1, dx_2, \dots, dx_n]^t$  se podrá conseguir una aproximación a la ecuación  $f_i(x_k + dx) = 0$  en (2.33) utilizando la diferencia completa.

Obteniéndose:

$$f_i(x_k) = \frac{\partial f_i(x_k)}{\partial x_1} dx_1 + \frac{\partial f_i(x_k)}{\partial x_2} dx_2 + \dots + \frac{\partial f_i(x_k)}{\partial x_n} dx_n = 0 \quad , \quad (2.34)$$

$$i = 1, \dots, n \quad .$$

El sistema anterior es lineal en  $dx_1, dx_2, \dots, dx_n$  y su forma matricial está dada de la siguiente manera:

$$\begin{bmatrix} \frac{\partial f_1(x_k)}{\partial x_1} & \frac{\partial f_1(x_k)}{\partial x_2} & \dots & \frac{\partial f_1(x_k)}{\partial x_n} \\ \frac{\partial f_2(x_k)}{\partial x_1} & \frac{\partial f_2(x_k)}{\partial x_2} & \dots & \frac{\partial f_2(x_k)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(x_k)}{\partial x_1} & \frac{\partial f_n(x_k)}{\partial x_2} & \dots & \frac{\partial f_n(x_k)}{\partial x_n} \end{bmatrix}^{-1} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix} = - \begin{bmatrix} f_1(x_k) \\ f_2(x_k) \\ \vdots \\ f_n(x_k) \end{bmatrix} = J^{-1} dx = -f(x_k) \quad ; \quad (2.35)$$

donde,  $J = f'(x_k)$ .

Por consiguiente, se puede obtener  $x_{k+1}$  a partir de  $x_k$  como  $x_{k+1} = x_k + dx_k$  donde  $dx_k$  corresponde a la solución de  $f'(x_k)^{-1}df = -f(x_k)$ .

La matriz  $J = f'(x_k)$  en (2.37), corresponde a la *matriz jacobiana no lineal* que se relaciona con la ecuación  $f(x)$  (la  $i$ -ésima ecuación). Por otro lado, la columna  $j$  de  $J$  posee todas las derivadas parciales con respecto a  $x_j$  (la  $j$ -ésima variable).

De esta forma:

$$f'(x_k) = \left[ \frac{\partial f_i(x_k)}{\partial x_j} \right]_{n \times n} . \quad (2.36)$$

Pero, si se desea estimar un vector de parámetros  $\theta$  a través de máxima verosimilitud, se deberán solucionar las ecuaciones de verosimilitud, en las cuales:

$$\theta_{(n+1)} = \theta_n - F'(\theta_n)^{-1}f(\theta_n) \quad ; \quad (2.37)$$

donde, se obtiene el logaritmo de la función de verosimilitud. Entonces:

$$f(\theta_n) = \frac{d \ln(L)}{d\theta} = S(\theta) \quad y \quad (2.38)$$

$$F'(\theta_n) = -I(\theta) = \begin{bmatrix} \frac{\partial^2 \ln(L)}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 \ln(L)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ln(L)}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 \ln(L)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln(L)}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial^2 \ln(L)}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 \ln(L)}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 \ln(L)}{\partial \theta_k \partial \theta_2} & \cdots & \frac{\partial^2 \ln(L)}{\partial \theta_k \partial \theta_k} \end{bmatrix} . \quad (2.39)$$

A la matriz anterior se le conoce como matriz de información de Fisher. Por consiguiente, su estimación mediante el método de Newton Raphson será:

$$\theta_{(n+1)} = \theta_n - F'(\theta_n)^{-1}f(\theta_n) .$$

El método de Newton Raphson posee la ventaja en la resolución de sistemas de ecuaciones no lineales, que corresponde a su rapidez de convergencia, en el momento en que se tiene una aproximación lo bastante precisa. Por lo contrario, una de sus desventajas es que se necesita una aproximación inicial exacta de la solución para así lograr certificar la convergencia.

Se debe tener presente que no siempre será sencillo poder establecer valores iniciales que logren dar lugar a una solución. Por lo general, los estimadores de momentos son considerados como valores iniciales para poder aplicar el método, no siendo sencillo obtener los estimadores mediante el método de los momentos ya que existen cálculos de por medio que dificultan encontrarlos.

### 2.3.6 Adecuación del modelo (Alcaide, 2015)

Esta sección se centrará en la adecuación del modelo, concretamente, en la adecuación del modelo tradicional Binomial Negativo. Este modelo destaca porque se puede considerar como un modelo lineal generalizado y como un modelo compuesto Poisson-Gamma. El valor del enfoque del modelo lineal generalizado Binomial Negativo (MLGBN), está en la capacidad de evaluar el modelo usando los diferentes estadísticos de bondad de ajuste y los residuos asociados a los modelos lineales generalizados descritos en el primer capítulo. Al adecuar estos resultados al MLGBN se obtiene lo siguiente:

- **Devianza residual:**

$$\begin{aligned}
 D &= 2 \{l(y; y) - l(\hat{\mu}; y)\} \\
 &= 2 \sum_{i=1}^n \{y_i \ln(y_i / \hat{\mu}_i) - (y_i + 1/\hat{\alpha}) / \hat{\alpha} \cdot \ln [(1 + \hat{\alpha} y_i) / (1 + \hat{\alpha} \hat{\mu}_i)]\} \quad .
 \end{aligned}
 \tag{2.40}$$

- **Estadístico chi-cuadrado de Pearson:**

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{(\hat{\mu}_i + \hat{\alpha} \hat{\mu}_i)^2} \quad .
 \tag{2.41}$$

En cuanto al análisis del MLGBN, se siguen generalmente las variables como miembros de la familia de modelos lineales generalizados. Entre estas variedades se incluyen el residuo básico, los residuos de Pearson y desviación estandarizados y no estandarizados, así como los estudentizados de ambos.

■ **Residuo de Pearson:**

$$r_i^p = \frac{(y_i - \hat{\mu}_i)}{\sqrt{\hat{\mu}_i + \hat{\alpha}\hat{\mu}_i^2}} . \quad (2.42)$$

Siendo su versión estudentizado:

$$r_i^{pt} = \frac{(y_i - \hat{\mu}_i)}{\hat{\phi}\sqrt{\hat{\mu}_i + \hat{\alpha}\hat{\mu}_i^2}} ; \quad (2.43)$$

donde,  $\hat{\phi}$  es un estimador consistente del parámetro escala  $\phi$ . Y, su versión estandarizada, viene dada por:

$$r_i^{pst} = \frac{(y_i - \hat{\mu}_i)}{\sqrt{(1 - h_i)(\hat{\mu}_i + \hat{\alpha}\hat{\mu}_i^2)}} ; \quad (2.44)$$

siendo,  $h_i$  el i-ésimo elemento de la matriz de proyección.

■ **Residuo desviación:**

$$r_i^d = sgm(y_i - \hat{\mu}_i)\sqrt{D} ; \quad (2.45)$$

donde,  $D$  es la desviación residual para el MLGBN señalado en 2.40. Su versión estudentizado viene dada por:

$$r_i^{dt} = \frac{sgm(y_i - \hat{\mu}_i)\sqrt{D}}{\hat{\phi}} \quad . \quad (2.46)$$

Su versión estandarizada está dada por:

$$r_i^{dst} = \frac{sgm(y_i - \hat{\mu}_i)\sqrt{D}}{\sqrt{(1 - h_i)}} \quad . \quad (2.47)$$

▪ **Residuo de Anscombe:**

Los residuos Anscombe usan la función de varianza del modelo.

$$Binomial \quad negativa = V(\mu) = \mu(1 + \alpha\mu) \quad . \quad (2.48)$$

El residual Anscombe calculado para un MLGBN viene dado por:

$$r^A = \frac{\left\{ 3/\hat{\alpha} \left\{ (1 + \hat{\alpha}y_i)^{\frac{2}{3}} - (1 + \hat{\alpha}\hat{\mu}_i)^{\frac{2}{3}} \right\} + 3(y_i^{\frac{2}{3}} - \hat{\mu}_i^{\frac{2}{3}}) \right\}}{2(\hat{\alpha}\hat{\mu}_i^2 + \mu_i)^{\frac{1}{6}}} \quad . \quad (2.49)$$

### 2.3.7 Interpretación del modelo (Alcaide, 2015)

Cuando se piensa que se ha obtenido el modelo óptimo de MLGBN se debe aplicar los criterios de la bondad de ajuste, estudiar los residuos, finalizando el proceso de modelar con la interpretación del modelo.

Entonces resulta importante recordar que la transformación provocada por la aplicación de una función de enlace da lugar, en la mayoría de los casos, a una ecuación del modelo expresada en términos multiplicativos, en la que, como se señaló en el primer capítulo, la interpretación de los parámetros se realiza en términos del factor de cambio en el valor esperado para un incremento unitario de las variables explicativas.

De forma particular el modelo tradicional Binomial Negativo viene dado por:

$$\ln(E(Y)) = \widehat{\beta}_0 + X_1\widehat{\beta}_1 + X_2\widehat{\beta}_2 + \dots + X_p\widehat{\beta}_p \quad . \quad (2.50)$$

Siendo la función de enlace la función logarítmica. Además, cada variable explicativa puede influir de forma positiva o negativa a la variable de interés, dependiendo del valor de  $\beta$  estimado.

Al igual que en la regresión logística si una variable explicativa es cualitativa, se deben introducir las variables *dummy*. En este caso, si la variable explicativa  $X_k$  es una variable *dummy*, entonces se tiene la siguiente relación:

$$\frac{E(Y_i/X_{ik} = 1)}{E(Y_i/X_{ik} = 0)} = \exp(\widehat{\beta}_k) \quad . \quad (2.51)$$

La media condicional es  $\exp(\widehat{\beta}_k)$  veces mayor si  $X_k$  toma el valor 1 en vez del valor 0.

Si el caso es que sea una variable explicativa  $X_k$  con  $k = 1, \dots, p$  una variable continua, cada unidad de aumento de esta variable produce un aumento de  $\widehat{\beta}_j$  del log valor esperado de la variable respuesta. Esta interpretación proporciona una restricción a valores positivos de la variable respuesta, adecuándose al tipo de los datos de recuento.

---

# APLICACIÓN

---

## 3.1 Análisis de datos

En este capítulo se llevará a cabo la aplicación de los modelos mencionados, los modelos lineales generalizados de Poisson y Binomial Negativo. Pero para llevar a cabo este análisis es necesario realizar previamente una presentación de los datos a utilizar y un análisis exploratorio respectivamente, cabe mencionar que todos los análisis fueron realizados por el software estadístico Stata, versión 14.

Se utilizó la base de datos de egresos hospitalarios de 2016 de Chile y estas cifras se obtienen a partir del Informe Estadístico de Egreso Hospitalario (Decreto N° 1671/2010) cuyo reporte es obligatorio para todos los establecimientos de salud del territorio nacional para el año respectivo. Luego, a esta base de datos se le realizaron algunos filtros en donde; se consideró solo observaciones de la quinta región, se consideró observaciones con una edad mayor que 60 años y los días de estadía hospitalaria del paciente no sobrepasaron el año de hospitalización. De esta forma se logró analizar solo a los adultos mayores de la región de Valparaíso del año mencionado.

## 3.2 Variables

Las variables registradas en la base de datos de interés son las siguientes (DEIS, 2017).

- **ESTAB:** Código de establecimiento, correspondiente al establecimiento de ocurrencia.
- **ServicioSalud:** Código servicio de salud, correspondiente al servicio de salud de ocurrencia.
- **SEXO:** Sexo del paciente (1 = Hombre, 2 = Mujer, 3 = Indeterminado, 9 = Desconocido).
- **EDAD:** Edad del paciente expresada en años.
- **PREVI:** Previsión de salud (1 = Fonasa, 2 = Isapre, 3 = No Tiene (cancela), 4 = Cajas de Previsión FFAA, 5 = capredena, 6 = dipreca, 7 = otra).
- **BENEF:** Clase de beneficio, corresponde a los tramos de Fonasa (1 = A, 2 = B, 3 = C, 4 = D).
- **MOD:** Modalidad de atención de los beneficiarios Fonasa (1 = MAI (Modalidad Atención Institucional), 2 = MLE (Modalidad Libre Elección)).
- **COMUNA:** Código Comuna.
- **REGION:** Región de residencia.
- **SERC\_EGR:** Código del servicio clínico de egreso.
- **DIAS\_ESTAD:** Días de estada total.

- **COND\_EGR:** Condición al egreso (1 = Vivo, 2 = Fallecido).
- **INTERV\_Q:** Intervención quirúrgica (1 = Sí, 2 = No).

Se debe destacar que el fin de este análisis es determinar un modelo adecuado para estimar la variable de interés Días de estada hospitalaria a través de las variables otorgadas por la base de datos de Egresos Hospitalarios 2016 y determinar si existe algún grado de dependencia entre las variables mencionadas y la variable de interés.

### 3.3 Análisis exploratorio

Se visualizará el comportamiento de las variables con datos cualitativos nominales.

- Adultos mayores de la región de Valparaíso clasificados por servicio de salud en el que se atienden de forma recurrente:

En la Tabla 3.1 se aprecia el servicio de salud de recurrencia del adulto mayor que fue dado de alta, siendo mayor los sectores de Coquimbo, Valparaíso San antonio y Viña del mar Quillota, siendo cerca del 80 % del total de datos y aportando con mayor frecuencia los adultos mayores que se atienden de forma reiterativa en los centros de salud ubicados en la región de Coquimbo con 16.298 registro, informando esto que la mayoría de las personas hospitalizadas en la región de Valparaíso son de la región de Coquimbo.

Tabla 3.1: Frecuencia de egresos hospitalarios de adultos mayores por servicio de salud de recurrencia.

Sector del servicio de salud	Frec.	Porcentaje
Arica	1	0,002
Iquique	2	0,004
Antofagasta	2	0,004
Atacama	9	0,020
Coquimbo	16.298	28,920
Valparaíso San Antonio	13.311	23,620
Viña Del Mar Quillota	15.808	28,050
Aconcagua	7.259	12,880
Metropolitano Norte	23	0,040
Metropolitano Occidente	55	0,100
Metropolitano Central	12	0,020
Metropolitano Oriente	371	0,660
Metropolitano Sur	3.107	5,510
Metropolitano Suroriente	19	0,030
Libertador B. O'Higgins	5	0,010
Del Maule	8	0,010
Ñuble	1	0,002
Concepción	4	0,010
Talcahuano	6	0,010
Bíobío	2	0,004
Araucanía Sur	6	0,010
Valdivia	14	0,020
Del Reloncaví	20	0,040
Aisén	2	0,004
Arauco	2	0,004
Chiloé	2	0,004
Total	56.349	100 %

- Adultos mayores de la región de Valparaíso clasificados por sexo:

Se visualiza en la Tabla 3.2, que existen en los registros mayor cantidad de *Mujeres* que *Hombres*, correspondiendo cerca del 52 % de los datos a *Mujeres*. Esto resulta semejante a la realidad a nivel país, que indica que existe mayor cantidad de mujeres que hombres en el territorio nacional como de mencionó en las nociones preliminares.

Tabla 3.2: Frecuencia de los egresos hospitalarios de adultos mayores por sexo.

Sexo	Frec.	Porcentaje
Hombre	27.266	48,39
Mujer	29.083	51,61
Total	56.349	100,00

- Clasificación los adultos mayores de la región de Valparaíso por previsión de salud con la que se atienden:

De la Tabla 3.3 se aprecia que la mayor parte de los adultos mayores de la región, pertenecen a la previsión de salud Fonasa con 42.533 registros que corresponden cerca del 75 %.

Tabla 3.3: Frecuencia de egresos hospitalarios de adultos mayores por previsión de salud.

Previsión de salud	Frec.	Porcentaje
Fonasa	42.533	75,48
Isapre	5.516	9,79
No tiene	494	0,88
Capredena	1.013	1,80
Dipreca	1.436	2,55
Otra	5.171	9,18
Sin información	186	0,33
Total	56.349	100,00

- Modalidad de atención de los beneficiarios de Fonasa:

Se extrae de la Tabla 3.4 que la mayor parte de los adultos mayores de la región que se atienden con la previsión de salud Fonasa poseen la modalidad de atención MAI (Modalidad Atención Institucional); en esta modalidad las prestaciones médicas que se otorgan a los beneficiarios es por los organismos que integran el Sistema Nacional de Servicios de Salud, sean dependientes del Ministerio de Salud o entidades públicas o privadas con las cuales los servicios de salud hayan realizado algún convenio para estos efectos.

Tabla 3.4: Frecuencia de egresos hospitalarios de adultos mayores por modalidad de atención de los beneficiarios de Fonasa.

Modalidad de atención B.F.	Frec.	Porcentaje
<i>MAI</i> (Modalidad Atención Institucional)	35.337	83,07
<i>MLE</i> (Modalidad Libre Elección)	7.204	16,93
Total	42.541	100

- Comuna de residencia de los adultos mayores de la región de Valparaíso:

En la tabla 3.5 se observa a los adultos mayores de la región clasificados por su comuna de residencia, obteniendo la comuna de Valparaíso y la comuna de Viña del Mar la mayor cantidad de observaciones con 12.105 registros (21,48 %) y 11.105 registros (19,71 %), respectivamente. Por otro lado, las comunas con menor cantidad de observaciones fueron la comuna de Juan Fernández y la comuna de Papudo con 7 registros (0,01 %) y 102 registros (0,18 %), respectivamente.

Tabla 3.5: Frecuencia de egresos hospitalarios de adultos mayores por comuna de residencia.

Código comuna	Frec,	Porcentaje
Algarrobo	340	0,60
Cabildo	574	1,02
Calera	1.385	2,46
Calle Larga	326	0,58
Cartagena	569	1,01
Casablanca	789	1,40
Catemu	587	1,04
Concón	851	1,51
El Quisco	316	0,56
El Tabo	275	0,49
Hijuelas	447	0,79
Isla de Pascua	213	0,38
Juan Fernández	7	0,01
La Cruz	340	0,60
La Ligua	1.003	1,78
Limache	506	0,90
Limache	1.607	2,85
Llaillay	1.088	1,93
Los Andes	2.241	3,98
Nogales	2.519	4,47
Olmué	491	0,87
Panquehue	194	0,34
Papudo	102	0,18
Petorca	298	0,53
Puchuncaví	401	0,71
Putendo	832	1,48
Quillota	2.115	3,75
Quilpué	4.577	8,12
Quintero	889	1,58
Rinconada	202	0,36
San Esteban	440	0,78
San Felipe	2.284	4,05
Santa María	428	0,76
Santo Domingo	247	0,44
Valparaíso	12.105	21,48
Villa Alemana	349	6,19
Viña del Mar	11.105	19,71
Zapallar	166	0,10
Total	56.349	57100

- Condición en la que se encuentra al momento de egresar el adulto mayor:

De la Tabla 3.6 se aprecia que la mayor parte de los adultos mayores de la región de Valparaíso, egresan vivos con 52,605 registros, correspondiente al 93,36 %.

Tabla 3.6: Frecuencia de la condición en la que egresan los adultos mayores hospitalizados.

Condición del egresado	Frec.	Porcentaje
Vivo	52.605	93,36
Muerto	3.744	6,64
Total	56.349	100

- Servicio clínico de egreso o nivel de cuidado que tuvo el adulto mayor de la región de Valparaíso antes de egresar:

Se observa en la Tabla 3.7 que el Área Médico-Quirúrgico de Cuidados Básicos es la que posee mayor frecuencia dentro de los adultos mayores de esta región con 21.257 registros que corresponden al 37,72 % y la sigue el Área Médica Adulto Cuidados Básicos con 9.969 registros que corresponden a un 17,69 %.

Tabla 3.7: Frecuencia del servicio clínico que recibió el adulto mayor antes de egresar del hospital

Código servicio salud	Frec.	Porcentaje
Medicina básica	1.723	3,060
Infeccioso adulto	4	0,010
Tratamiento antialcohólico	1	0,002
Medicina agudo	6	0,010
Área méd. quirúrg. indiferenc.	5.518	9,790
Cirugía básica	19	0,030
Traumatología-Ortopedia adulto	9	0,020
Traumatología agudo	2	0,004
Área quirúrg. infantil indif.	1	0,002
Pediatría indiferenciada	32	0,060
Obstetricia y ginecología	118	0,210
Neurología adulto	2	0,004
Psiquiatría corta estadia	3	0,010
Desintoxicación alcohol y drogas	1	0,002
Oncología indiferenciado	113	0,200
Urología	3	0,010
Medicina Física-Rehabilitación	233	0,410
Derivación médico quirúrgico	190	0,340
Tisiología crónico	4	0,010
Geriatría	527	0,940
U. Emergencia adulto	7	0,010
U.C.I. adulto	359	0,640
U.C.I. coronaria	12	0,020
U.T.I. indiferenciado	140	0,250
U.T.I. medicina	1	0,002
Pensionado	3.579	6,350
Á. médica adulto cuidados básicos	9.969	17,690
Á. médica adulto cuidados medios	2.075	3,680
Á. médico-quirúrgico cuidados básicos	21.257	37,720
Á. médico-quirúrgico cuidados medios	6.926	12,290
Á. cuidados intensivos adultos	809	1,440
Á. cuidados intermedios adultos	958	1,700
Á. médica pediátrica cuidados básicos	16	0,030
Á. obstetricia	618	1,100
Á. psiq. adulto corta estadía	145	0,260
Á. psiq. adulto mediana estadía	12	0,020
Á. psiq. adulto larga estadía	72	0,130
Á. psiq. forense adulto evaluación	59 1	0,002
Indiferenciado	884	1,544
Total	56.349	100,00

- Intervención quirúrgica por parte del egresado:

En la tabla 3.8 se extrae que la mayoría de las personas que fueron hospitalizadas y luego dados de alta, no se realizó ninguna intervención quirúrgica en el periodo de esta estadía con una frecuencia de 35.747 registros que corresponden al 63,44 %.

Tabla 3.8: Intervención quirúrgica

Intervención quirúrgica	Frec,	Porcentaje
Sí	20.602	36,56
No	35.747	63,44
Total	56.349	100

Ahora se visualizará las variables con datos cuantitativos.

- Edad del egresado hospitalario:

La Tabla 3.9 ilustra algunos estadísticos descriptivos de la variable edad, siendo la media igual a 74 años aproximadamente, indicando que en promedio la edad de los adultos mayores de la región de Valparaíso registrados en 2016 fue de 74 años. Además, se observa el valor mínimo igual a 60 que previamente se asignó como tal ya que la OMS (Organización Mundial de la Salud) define esta edad como el comienzo de la etapa etaria de la “tercera edad” y al observar el máximo, éste indica que existen registros de adultos mayores que llegan hasta los 115 años de edad en la región.

Tabla 3.9: descripción de la edad de los adultos mayores egresados

Variable	Obs	Media	Dev. Estand.	Min	Max
Edad	56.349	73,52959	8,984865	60	115

- Cantidad de días de estadía hospitalaria del egresado:

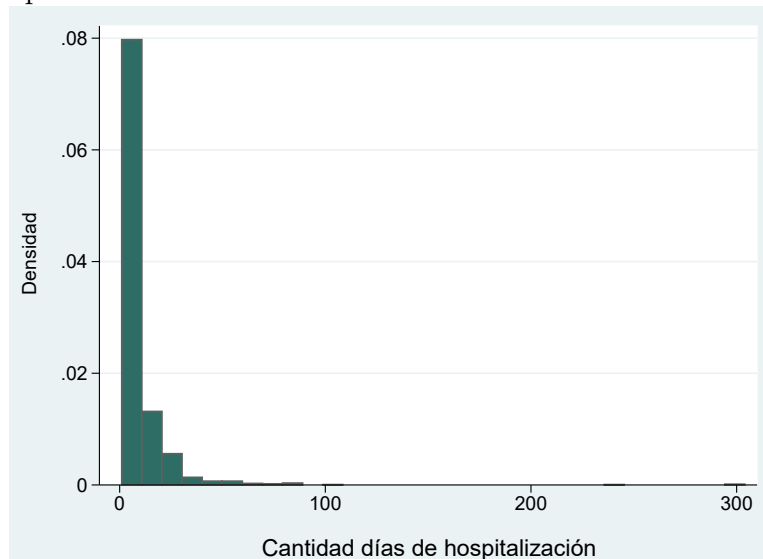
Por último, pero no menos importante, la Tabla 3.10 muestra el análisis descriptivo de la variable de interés del estudio “Días de estadía hospitalaria”. Que posee una media aproximadamente de 8,4 días, en otras palabras, el promedio de los días de estadía del adulto mayor de la región de Valparaíso corresponde aproximadamente a 8 días y además posee una desviación estándar igual a 13,2 que indica que existe una variación a considerar dentro de los datos.

Tabla 3.10: Descripción de días de estadía hospitalaria del adulto mayor

Variable	Obs	Media	Dev. Estand	Min	Max
Días de estadía	56.349	8,379865	13,16379	1	355

En la Figura 3.1 se ilustra el comportamiento de la variable de interés en el estudio siendo esta, días de estadía hospitalaria del adulto mayor en la región de Valparaíso, y se aprecia que cerca del 80 % de los adultos mayores es hospitalizado al menos 1 día. Haciendo considerar que se está en presencia de una variable que se podría distribuir Poisson o Binomial Negativa, dado que es una variable de conteo discreta y la mayor concentración de ocurrencia es en una frecuencia baja.

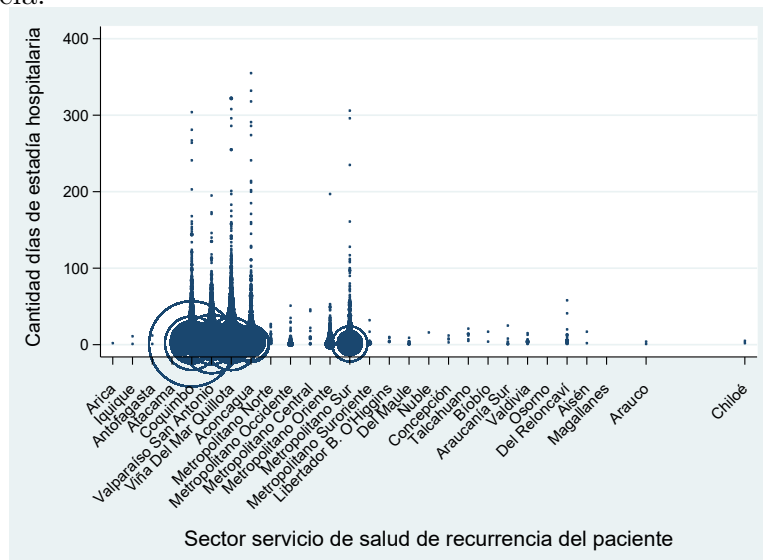
Figura 3.1: Frecuencia de días de estadía hospitalaria del adulto mayor en la región de Valparaíso.



Para continuar, es necesario visualizar antes de cualquier aplicación de un modelo lineal el comportamiento de las covariables definidas anteriormente con la variable respuesta, dado que se debe considerar al momento de modelar una variable si existe alguna asociación lineal de las covariables con la variable respuesta. Por ello, se verá el comportamiento de días de estadía hospitalaria con las demás variables a través de un gráfico de dispersión.

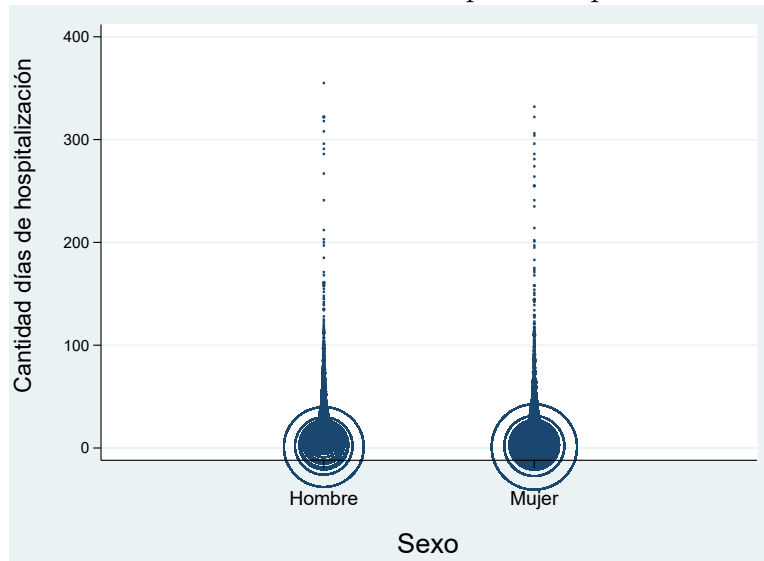
En los gráficos ilustrados se aplicó una variable *peso* que considera los valores repetidos para luego ampliar el círculo que hace referencia en el gráfico al dato puntual, lo que permite observar el comportamiento de la variable respuesta con las covariables definidas, de las cuales se aprecia que existe una asociación leve.

Figura 3.2: Gráfico de días de estadía hospitalaria por sector del servicio de salud de recurrencia.



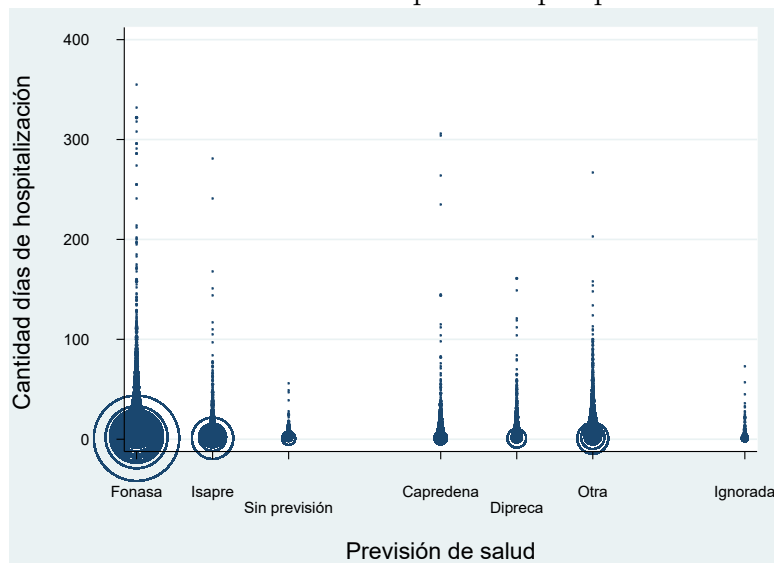
En el gráfico anterior se observa la cantidad de días de hospitalización del adulto mayor de la región de Valparaíso frente al sector de ubicación del servicio de salud que se atiende de forma recurrente el paciente, donde, se aprecia que gran parte de los datos son hospitalizados cerca de 1 o 2 días y provienen en los servicios de salud de los sectores de Coquimbo, de Valparaíso San Antonio y Viña del Mar Quillota en gran mayoría, pero siendo mayor el aporte de el sector de Coquimbo.

Figura 3.3: Gráfico de días de estadía hospitalaria por sexo del paciente.



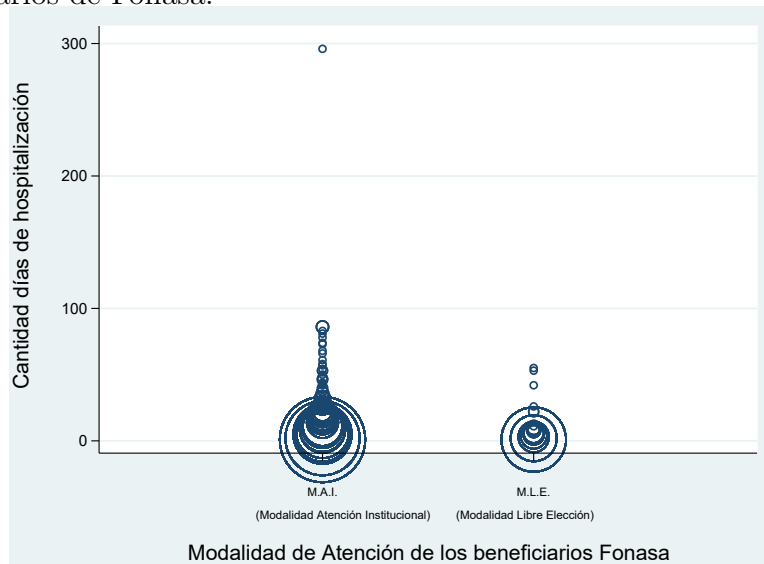
Se observa en el gráfico anterior la cantidad de días de hospitalización frente al sexo del paciente. Donde existe una leve cantidad superior de mujeres en comparación a cantidad de hombres y la mayoría de forma general cuenta con pocos días de hospitalización.

Figura 3.4: Gráfico días de estadía hospitalaria por previsión del paciente.



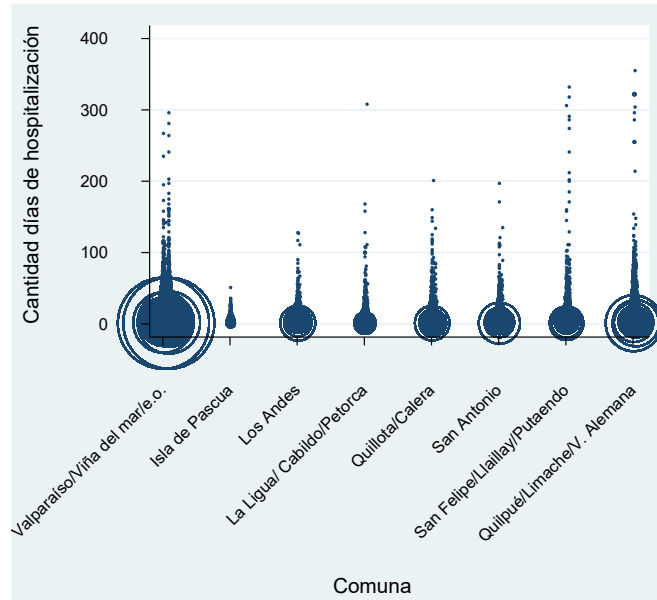
En el gráfico de dispersión de la cantidad de días de hospitalización frente a la previsión de salud que poseen los adultos mayores hospitalizados, se aprecia que la mayor parte de los sujetos observados se atienden por Fonasa, luego le sigue la atención Isapre.

Figura 3.5: Gráfico de días de estadía hospitalaria por modalidad de atención de los beneficiarios de Fonasa.



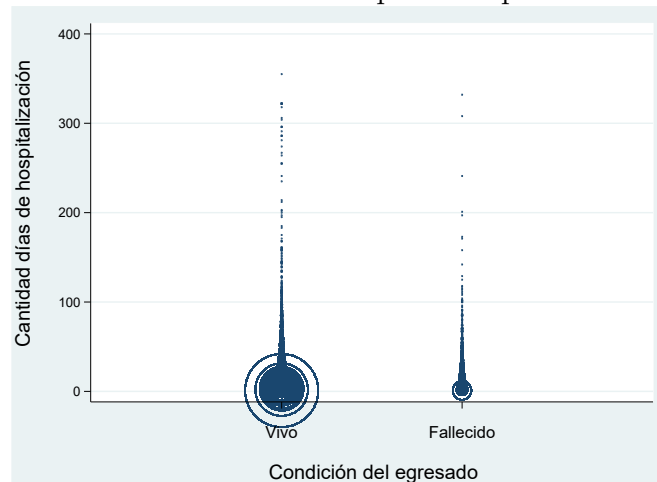
En el gráfico anterior que ilustra la modalidad de atención de los beneficiarios Fonasa frente a la cantidad de días de estadía hospitalaria, donde se aprecia que la mayor cantidad de observaciones se atienden por la Modalidad de Atención Institucional (M.A.I), pero también presentan mayor días de hospitalización en comparación a los que se atienden por la Modalidad de Libre Elección (M.L.E).

Figura 3.6: Gráfico de días de estadía hospitalaria por comuna de residencia del paciente.



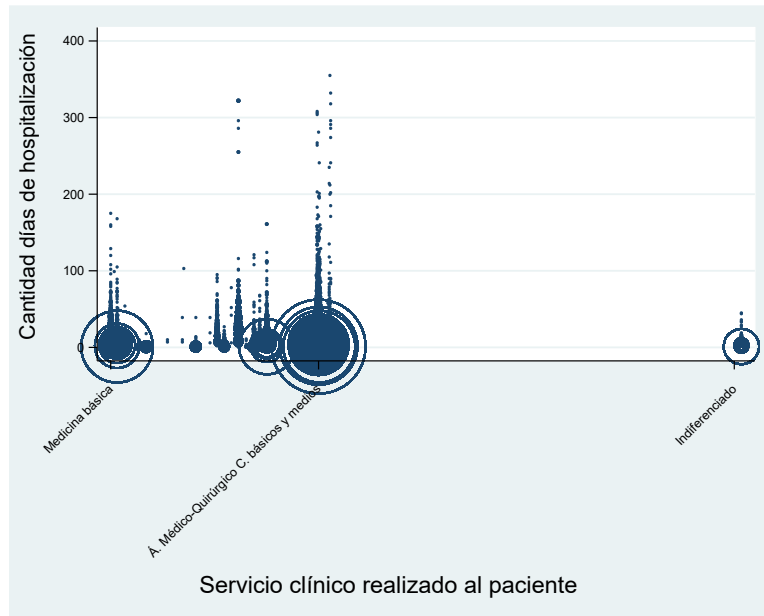
Se observa el gráfico de cantidad de días de hospitalización frente a la comuna de residencia del adulto mayor, que la mayor concentración de las observaciones reside en las comunas de Valparaíso y Viña del Mar. Seguido luego de las comunas de Quilpué, Limache y Villa Alemana.

Figura 3.7: Gráfico de días de estadía hospitalaria por condición del egresado.



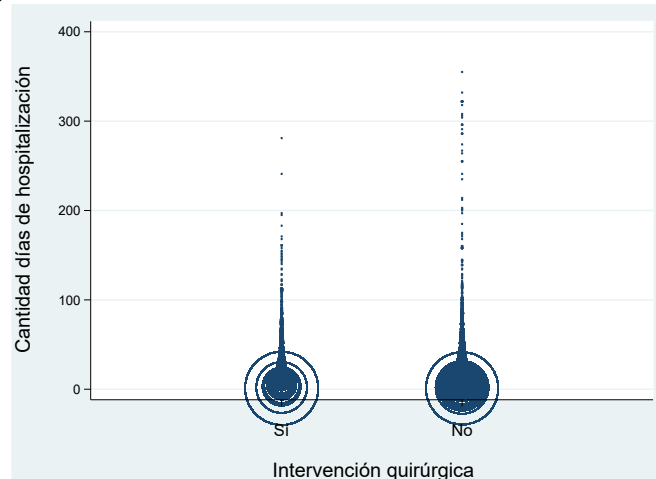
Del gráfico entre condición del egresado y días de estadía hospitalaria, se extrae que la mayor parte de las observaciones egresan vivos, pero con un mayor número de días de estadía.

Figura 3.8: Gráfico de días de estadía hospitalaria por servicio clínico realizado al paciente.



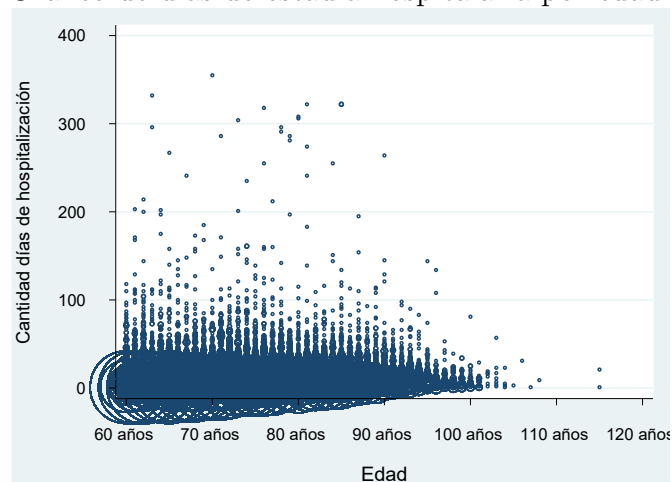
En el gráfico entre servicio clínico de cuidado que se realizó al paciente y días de estadía hospitalaria, se observa que la mayor parte de observaciones se presentan en los servicios clínicos que entreguen el área quirúrgica de cuidados básicos y medios y el área psiquiátrica del adulto mayor, aportando una mayor cantidad de días de estadía hospitalaria del paciente.

Figura 3.9: Gráfico de días de estadía hospitalaria por sí presenta intervención quirúrgica el egresado.



Se aprecia en el gráfico entre intervención quirúrgica del egresado y días de estadía que la concentración de adultos mayores hospitalizados no presentaron alguna intervención quirúrgica durante su estadía.

Figura 3.10: Gráfico de días de estadía hospitalaria por edad del egresado.



En el gráfico entre edad del egresado y días de estadía hospitalaria se concluye que existe una leve tendencia para afirmar que al aumentar la edad del egresado, aumenta los días de estadía hospitalaria que éste presente.

De igual manera es necesario calcular un coeficiente de asociación específico para el tipo de variable que se busque asociar. Es por ello que se utilizó el Coeficiente Lambda ( $\lambda$ ) para variables nominales, el Coeficiente Polychoric que permite relacionar variables dicotómicas, el Coeficiente tau de Kendall para variables ordinales y por último el Coeficiente de Correlación de Pearson para variables continuas como se verán a continuación en la Tabla 3.11 y además es necesario mencionar que al realizar una prueba de independencia a cada correlación, resultaron que no existía suficiente información para determinar que fueran independientes las covariables con la variable respuesta, entonces, existe alguna dependencia entre ellas, excepto días de estadía con previsión de salud:

Tabla 3.11: Correlaciones

Variables	Coeficiente	Correlación	Independencia
Días estadía - establecimiento	Lambda	0,0003	No
Días estadía - servicio de salud	Lambda	0,0248	No
Días estadía - sexo	Polychoric	-0,0176	No
Días estadía - edad	Spearman	0.0545	No
Días estadía - previsión	Tau Kendall	-0,0021	Sí
Días estadía - comuna	Lambda	0,0028	No
Días estadía - servicio egreso	Lambda	0,0041	No
Días estadía - condición del egresado	Polychoric	0,0646	No
Días estadía - Intervención quirúrgica	Polychoric	0,1178	No
Días estadía - Modalidad de atención <sup>a</sup>	Lambda	0,0003	No

<sup>a</sup>los coeficientes recorren de -1 a 1 excepto el Coeficiente Lambda que recorre de 0 a 1.

### 3.4 Aplicación del modelo lineal generalizado Poisson (MLGP)

Se cuenta con registros hospitalarios de 2016, éstos no son aleatorios, entonces fue necesario considerar una muestra aleatoria de tamaño 2500 de los 56.349 registros. Se escogió este número de datos ya que al simular esta cantidad fue la que obtuvo parámetros similares a la población. Luego ahora se supone que la cantidad de días de estadía hospitalaria del adulto mayor de la región de Valparaíso de 2016 es independiente con otros para un modelo lineal generalizado Poisson con parámetro  $\mu_i$ . Entonces  $Y$  sea cantidad de días de estadía:

$$Y \sim P(\mu_i)$$

Se consideró al momento de aplicar el MLGP las variables que resultaron dependientes de la variable respuesta en la Tabla 3.11, para dar como resultado después de ir quitando una por una todas las variables que no resultaron significativas en el modelo. Donde, se llegó a los resultados que se muestran en la Tabla 3.12.

Tabla 3.12: Estimación de los coeficientes del MLGP con función de enlace logarítmica.

Coeficiente	Estimación	Error estándar	Z	P >  z
2.SEXO ( $X_1$ )	-0,0470	0,0139	-3,35	0,001
EDAD ( $X_2$ )	0,0127	0,0008	16,15	<0,001
2.COND_EGR ( $X_3$ )	0,3641	0,0243	14,99	<0,001
2.INTERV_Q ( $X_4$ )	0,1274	0,0152	8,33	<0,001
Intercepto	1,0908	0,0578	18,87	<0,001

Se observa en la Tabla 3.12 las variables que resultaron significativas en el modelo Poisson. Cabe mencionar que la mayoría de las variables son nominales por lo que se ingresaron al modelo como variables *dummy*. Se observan los resultados de los coeficiente de la estimación, pero al ser un modelo que posee una función de enlace logarítmica, los resultados se deben exponenciar para su correcta interpretación como se visualiza en la Tabla 3.13.

Tabla 3.13: Exponencial de los coeficientes del MLGP

Coefficiente	Estimación	Intervalo del 95 %
2.SEXO ( $X_1$ )	0,9544	0,9288 - 0,9808
EDAD ( $X_2$ )	1,0128	1,0112 - 1,0145
2.COND_EGR ( $X_3$ )	1,4393	1,3724 - 1,5094
2.INTERV_Q ( $X_4$ )	1,1343	1,1023 - 1,1704
Intercepto	2,9765	2,6577 - 3,3336

En la Tabla 3.13 a primera vista se aprecian las estimaciones interpretables del modelo lineal generalizado Poisson con su intervalo de confianza del 95 % pertinente. Todos los coeficientes resultan positivos en comparación con la Tabla 3.12, por lo que las covariables aportan de forma ascendente a la variable respuesta Días de estadía hospitalaria del adulto mayor.

Quedando el modelo de la siguiente manera:

$$\begin{aligned} \widehat{DIAS\_ESTAD} &= e^{\beta_0 + \beta_1(X_1=2) + \beta_2 X_2 + \beta_3(X_3=2) + \beta_4(X_4=2)} \\ &= e^{\beta_0} e^{\beta_1(X_1=2)} e^{\beta_2 X_2} e^{\beta_3(X_3=2)} e^{\beta_4(X_4=2)} \end{aligned}$$

Donde los coeficientes tienen un efecto multiplicativo.

Para realizar una comparación de modelos es necesario observar los estadísticos que aparecen como resultado al momento de implementar la regresión siendo estos el AIC, el BIC y la devianza ( $D$ ).

Para el MLGP se obtuvo:

$$\begin{aligned} AIC &= 13,434 \\ BIC &= 5599,854 \end{aligned}$$

$$\frac{D}{gl} = \frac{25,120,849}{2,495} = 10,068$$

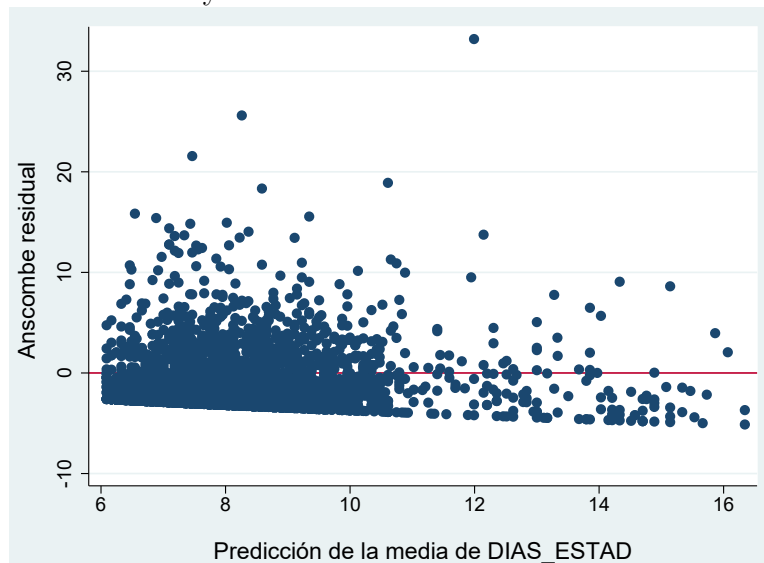
El valor del AIC y el BIC se comparará posteriormente con los entregados por el modelo lineal generalizado Binomial Negativo, además, al observar el resultado de la devianza dividido por los grados de libertad del MLGP,

indica que el modelo no logra capturar la sobredispersión existente en los datos, ya que éste valor debe ser lo más cercano a 1, lo que hace suponer que el MLGBN logrará mejores resultados.

A continuación, se sigue con el análisis de los residuos Anscombe del MLGP.

### 3.5 Diagnóstico del MLGP

Figura 3.11: Residuos Anscombe v/s predicción de la media para días de estadía hospitalaria del adulto mayor



De lo anterior, se observa que en el gráfico de los residuos de Anscombe frente a los pronósticos de la media de días de estadía hospitalaria, los residuos tienen una tendencia a la media con excepción de algunos datos atípicos que podrían ser influyentes.

### 3.6 Aplicación del modelo lineal generalizado Binomial Negativo (MLGBN)

Se trabajó con la misma muestra aleatoria de tamaño 2500 que se utilizó en la aplicación del modelo lineal generalizado Poisson, luego se puede suponer que la cantidad de días de estadía del adulto mayor de la región de Valparaíso de 2016 es independiente con otros para un modelo Binomial Negativo, entonces sea  $Y$  la cantidad de días de estadía de un adulto mayor que se distribuye como sigue:

$$Y \sim BN(\mu_i, \phi)$$

Se consideró al igual que en el modelo lineal generalizado Poisson todas las variables que resultaron con un grado de dependencia con la variable respuesta en la Tabla 3.11, lo cual se redujo luego de sacar una por una todas las variables que no resultaron significativas en el modelo. Dando como resultado los valores de la Tabla 3.14:

Tabla 3.14: Estimación de los coeficientes del MLGBN con función de enlace logarítmico.

Coeficiente	Estimación	Error estándar	Z	P >  z
EDAD ( $X_1$ )	0,0134	0,0024	5,57	<0,001
2.COND_EGR ( $X_2$ )	0,4135	0,0873	4,73	<0,001
2.INTERV_Q ( $X_3$ )	0,1673	0,0446	3,75	<0,001
Intercepto	0,9856	0,1769	5,57	<0,001

Se aprecia en la Tabla 3.14 las variables que resultaron significativas en el MLGBN, en comparación al MLGP éstas fueron mucho menos, debido a que al contener sobredispersión los datos, el primer modelo conduce a confusiones, ya que algunas variables contribuyen significativamente al modelo, cuando esto no ocurre. Además, todas las variables resultan con un coeficiente que aporta de forma positiva a la variable respuesta, pero se debe exponenciar los coeficientes de estimación para su interpretación como sigue en la Tabla 3.15.

Tabla 3.15: Exponencial de los coeficientes del MLGBN

Coeficiente	Estimación	Intervalo del 95 %
EDAD ( $X_1$ )	1,0135	1,0087 - 1,0183
2.COND_EGR ( $X_2$ )	1,5120	1,2742 - 1,7943
2.INTERV_Q ( $X_3$ )	1,1822	1,0831 - 1,2903
Intercepto	2,6795	1,8943 - 3,7900

El modelo queda como sigue:

$$\begin{aligned} \widehat{DIAS\_ESTAD} &= e^{\beta_0 + \beta_1(X_1) + \beta_2(X_2=2) + \beta_3(X_3=2)} \\ &= e^{\beta_0} e^{\beta_1(X_1)} e^{\beta_2(X_2=2)} e^{\beta_3(X_3=2)} \end{aligned}$$

Queda en evidencia que los coeficientes tienen un efecto multiplicativo. Para el modelo de Binomial Negativo resultaron los siguientes valores:

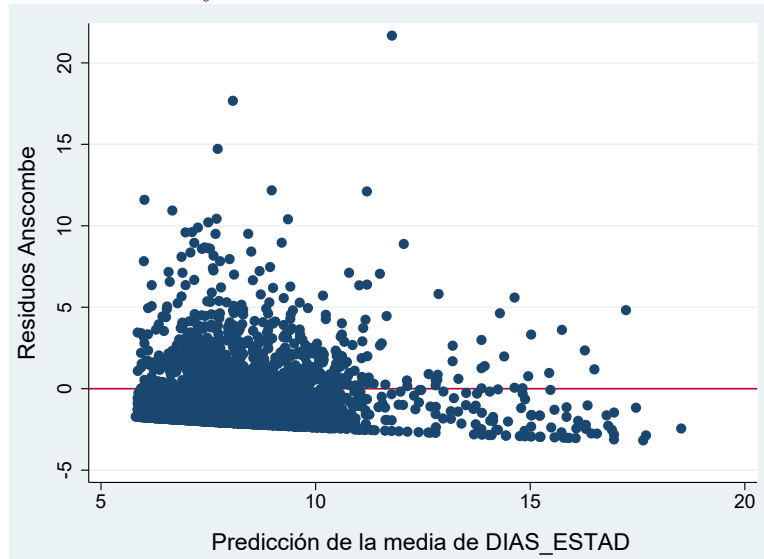
$$\begin{aligned} AIC &= 6,326 \\ BIC &= -16922,070 \end{aligned}$$

$$\frac{D}{gl} = \frac{2606,752}{2496} = 1,044$$

A continuación, se realiza el análisis de residuos del modelo de Binomial Negativo.

### 3.7 Diagnóstico del MLGBN

Figura 3.12: Residuos Anscombe v/s predicción de la media para días de estadía hospitalaria del adulto mayor.



De lo anterior, se observa que en el gráfico de los residuos de Anscombe frente a los pronósticos de la media de días de estadía hospitalaria, los residuos tienen una tendencia a la media con excepción de algunos datos atípicos que podrían ser influyentes.

Luego se comparan los resultados de cada modelo.

Tabla 3.16: Comparación entre ambos modelos

	MLGP	MLGBN
AIC	13,284	6,278
BIC	5279,377	-16966,290
$\frac{D}{gl}$	9,942	1,021

Al observar la Tabla 3.16 se comprueba la hipótesis que se tenía sobre que sería mejor la aplicación de un modelo Binomial Negativo dado que éste se ajustaría mejor a los datos y, esto se comprueba con el valor AIC y BIC ya que en el modelo mencionado resultó con menores valores. Además, la devianza dividida por los grados de libertad del modelo, resultó más cercano a uno, informando que el modelo está capturando la sobredispersión existente en los datos, esto sucede porque se agrega la estimación de un parámetro adicional que captura la sobredispersión que presentan los datos.

---

## CONCLUSIONES

---

- Al considerar los datos de conteo propuestos, se concluye de los modelos estudiados, que el mejor modelo que logra ajustarse a los datos es el MLG Binomial Negativo. Dado que, representa la sobredispersión de los datos respetando la naturaleza discreta de la variable respuesta y la no negatividad en sus resultados.
- Se determinó que existe una relación entre los días de estadía hospitalaria del adulto mayor de la región de Valparaíso de 2016 con la edad que presenta el adulto mayor, la condición de su egreso hospitalario y si durante su hospitalización tuvo alguna intervención quirúrgica, ya que fueron las variables que resultaron significativas en el modelo lineal generalizado Binomial Negativo.
- Se trató de evaluar la presencia de alguna interacción relevante de las variables contenidas en los modelos propuestos en este trabajo de investigación, pero las interacciones que se evaluaron no resultaron coherentes al momento de realizar la interpretación, dado su naturaleza nominal y la falta de datos en las posibles interacciones.

---

## RECOMENDACIONES

---

- Siempre al considerar cualquier conjunto de datos, es necesario conocer sus características, para realizar una correcta aplicación del modelo escogido.
- Al aplicar algún tipo de modelo es de gran importancia ver la naturaleza de la variable respuesta, su recorrido y posibles supuestos.
- Sería oportuno indagar con mayor profundidad en las variables nominales que poseen un gran número de categorías para determinar un modelo con mejor ajuste y extenderlo a nivel nacional, para determinar cuáles son las causas de incremento de días de estadía hospitalaria de un paciente.
- Se debe tener en cuenta al momento de interpretar los coeficientes estimados del modelo la función de enlace que se aplicó, para no caer en interpretaciones erróneas.

---

# APÉNDICE

---

## 3.8 Código en Stata

```
//lectura base de datos//
clear all
import excel "C:\Users\johuwsa\Documents\Mi tesis (
  Hospitalización del Adulto Mayor en Chile)\Proyecto Adulto
  Mayor-Egresos hospitalarios\Datos Egresos Hospitalarios
  Chile\Work in Stata\Egresos2016.xls", sheet("Sheet1")
firstrow clear
set more off , permanent

/* restricciones que se realizaron en la base de datos*/
keep if edad>=60
drop if region != 5
keep if dias_estad <365

//Etiquetas de las variables//

label variable ESTAB "Establecimiento por código"
label variable ServicioSalud "Servicio de salud Por código"
label variable SEXO "Sexo"
label variable EDAD "Edad"
label variable PREVI "Previsión de salud"
label variable BENEF "Clase de beneficiario"
label variable MOD "Modalidad de atención de los beneficiarios
  Fonasa"
label variable COMUNA "Código de comuna"
label variable REGION "Región de residencia"
label variable SERV_RES "Servicio de salud de referencia"
```

```

label variable FECHA_EGR "Fecha de alta del paciente"
label variable SERC_EGR "Código del servicio clínico de egreso"
label variable DIAG1 "Código diagnóstico principal"
label variable DIAG2 "Causa externa"
label variable DIAS_ESTAD "Cantidad días de hospitalización"
label variable CONDEGR "Condición del egresado"
label variable INTERV_Q "Intervención Quirúrgica"

// Análisis descriptivo , frecuencias de variables nominales//

hist DIAS_ESTAD
sum DIAS_ESTAD, detail
tabulate ESTAB
tabulate ServicioSalud
tabulate SEXO
tabulate PREVI
tabulate MOD
tabulate COMUNA
tabulate SERV_RES
tabulate SERC_EGR
tabulate CONDEGR
tabulate INTERV_Q
tabulate EDAD
sum EDAD
tabulate DIAS_ESTAD

//etiqueta la variable ServicioSalud//

# delimit ;
label def ServicioSalud_2 1 "Arica" 2 "Iquique" 3 "Antofagasta"
4 "Atacama"
5 "Coquimbo" 6 "Valparaíso San Antonio" 7 "Viña Del Mar
Quillota"
8 "Aconcagua" 9 "Metropolitano Norte" 10 "Metropolitano
Occidente"
11 "Metropolitano Central" 12 "Metropolitano Oriente" 13 "
Metropolitano Sur"
14 "Metropolitano Suroriente" 15 "Libertador B. O'Higgins"
16 "Del Maule"
17 "Ñuble" 18 "Concepción" 19 "Talcahuano" 20 "Bíobío" 21 "
Araucanía Sur"
22 "Valdivia" 23 "Osorno" 24 "Del Reloncaví" 25 "Aisén" 26 "
Magallanes"
28 "Arauco" 33 "Chiloé"
;
# delimit cr , replace

```

```

label val ServicioSalud ServicioSalud_2

// etiqueta la variable SERCEGR//

# delimit ;
label def ServicioEgreso 110 "Medicina básica" 112 "Infeccioso
adulto"
113 "Tratamiento antialcohólico" 115 "Medicina agudo"
119 "Á. méd. quirúrg. indife." 120 "Cirugía básica"
130 "Traumatología–Ortopedica" 131 "Traumatología agudo"
145 "Área quirúrg. infantil indif." 150 "Pediatria
indiferenciada"
160 "Obstetricia y ginecología" 190 "Neurología adulto"
211 "Psiquiatria corta estadia" 213 "Desintoxicación alcohol y
drogas"
230 "Oncología indiferenciado" 250 "Urología"
260 "Medicina Física–Rehab." 270 "Deriv. médico quirúr."
280 "Tisiología crónico" 290 "Geriatría" 301 "U. Emergencia
Adulto"
312 "U.C.I. Adulto" 314 "U.C.I. Coronaria" 320 "U.T.I.
Indiferenciado"
321 "U.T.I. Medicina" 330 "Pensionado" 401 "Á. Médica Adulto
Cuidados Básicos"
402 "Á. Médica Adulto Cuidados Medios" 403 "Á. Médico–
Quirúrgico Cuidados Básicos"
404 "Á. Médico–Quirúrgico Cuidados Medios" 405 "Á. Cuidados
Intensivos Adultos"
406 "Á. Cuidados Intermedios Adultos" 407 "Á. Médica Pediátrica
Cuidados Básicos"
416 "Á. Obstetricia" 418 "Á. Psiquiatria Adul. Corta estad."
419 "Á. Psiquiatria Adul. Mediana estad." 420 "Á. Psiquiatria
Adul. Larga estad."
423 "Á. Psiquiatria Forense Adulto" 999 "Indiferenciado"
;
# delimit cr , replace

label val SERCEGR ServicioEgreso

// Gráficos descriptivos //

sort DIAS_ESTAD ESTAB
by DIAS_ESTAD ESTAB: g ENE = _N
twoway (scatter DIAS_ESTAD ESTAB [w=ENE], msymbol(circle_hollow
))

```

```

sort DIAS_ESTAD EDAD
by DIAS_ESTAD EDAD: g ENE1 = _N
twoway (scatter DIAS_ESTAD EDAD [w=ENE1], msymbol(circle_hollow
))

sort DIAS_ESTAD PREVI
by DIAS_ESTAD PREVI: g ENE2 = _N
twoway (scatter DIAS_ESTAD PREVI [w=ENE2], msymbol(
circle_hollow))

sort DIAS_ESTAD COMUNA
by DIAS_ESTAD COMUNA: g ENE3 = _N
twoway (scatter DIAS_ESTAD COMUNA [w=ENE3], msymbol(
circle_hollow))

sort DIAS_ESTAD SERV_RES
by DIAS_ESTAD SERV_RES: g ENE4 = _N
twoway (scatter DIAS_ESTAD SERV_RES [w=ENE4], msymbol(
circle_hollow))

sort DIAS_ESTAD SEXO
by DIAS_ESTAD SEXO: g ENE5 = _N
twoway (scatter DIAS_ESTAD SEXO [w=ENE5], msymbol(circle_hollow
))

sort DIAS_ESTAD CONDEGR
by DIAS_ESTAD CONDEGR: g ENE6 = _N
twoway (scatter DIAS_ESTAD CONDEGR [w=ENE6], msymbol(
circle_hollow))

sort DIAS_ESTAD INTERV_Q
by DIAS_ESTAD INTERV_Q: g ENE7 = _N
twoway (scatter DIAS_ESTAD INTERV_Q [w=ENE7], msymbol(
circle_hollow))

sort DIAS_ESTAD SERCEGR
by DIAS_ESTAD SERCEGR: g ENE8 = _N
twoway (scatter DIAS_ESTAD SERCEGR [w=ENE8], msymbol(
circle_hollow))

sort DIAS_ESTAD ServicioSalud
by DIAS_ESTAD ServicioSalud: g ENE9 = _N
twoway (scatter DIAS_ESTAD ServicioSalud [w=ENE9], msymbol(
circle_hollow))

sort DIAS_ESTAD MOD
by DIAS_ESTAD MOD: g ENE10 = _N

```

```

twoway (scatter DIAS_ESTAD MOD [w=ENE10], msymbol(circle_hollow
))

//Correlaciones// \\

// se debe instalar el Coeficiente Polychoric\\
// se debe instalar el Coeficiente Lambda y guttman1 para
variables nominales\\

ssc install guttman1

guttman1 DIAS_ESTAD ESTAB
spearman DIAS_ESTAD ESTAB
//ktau DIAS_ESTAD ESTAB
lambda DIAS_ESTAD ServicioSalud
//spearman DIAS_ESTAD ServicioSalud
polychoric DIAS_ESTAD SEXO
//spearman DIAS_ESTAD SEXO
corr DIAS_ESTAD EDAD
//spearman DIAS_ESTAD EDAD
spearman DIAS_ESTAD PREVI
ktau DIAS_ESTAD PREVI
lambda DIAS_ESTAD COMUNA
spearman DIAS_ESTAD COMUNA
//lambda DIAS_ESTAD SERV_RES
lambda DIAS_ESTAD SERC_EGR
spearman DIAS_ESTAD SERC_EGR
polychoric DIAS_ESTAD CONDE_EGR
spearman DIAS_ESTAD CONDE_EGR
polychoric DIAS_ESTAD INTERV_Q
spearman DIAS_ESTAD INTERV_Q
lambda DIAS_ESTAD MOD
//ktau DIAS_ESTAD MOD
//spearman DIAS_ESTAD MOD

// Recolección de una muestra aleatoria //

sample 2500, count
save "muestra", replace
use muestra
sum DIAS_ESTAD, detail

//Modelo poisson //

```

```

glm DIAS_ESTAD i.SEXO EDAD i.CONDEGR i.INTERV_Q , family(
  poisson) link(log) nolog

// Análisis de los residuos Anscombe MLGP //

predict e, anscombe
predict y, mu
twoway (scatter e y), yline(0)

// cálculo de los coeficientes del modelo Poisson//

display exp(-.0466306) , exp(-.0738724) , exp(-.0193888)
display exp(.0127074) , exp(.0111648) , exp(.01425)
display exp(.3641415) , exp(.3165404) , exp(.4117425)
display exp(.1273875) , exp(.0974243) , exp(.1573508)
display exp(1.090755) , exp(.9774565) , exp(1.204053)

// Modelo binomial negativo //

glm DIAS_ESTAD EDAD i.CONDEGR i.INTERV_Q, family(nbinomial)
  link(log) nolog

// Análisis de los residuos Anscombe MLGBN //

predict er, anscombe
predict yu, mu
twoway (scatter er yu), yline(0)

//cálculo de los coeficientes binomial negativo
display exp( .013396) , exp(.0086823) , exp(.0181098)
display exp(.4134707) , exp(.2423146) , exp(.5846269)
display exp( .16735) , exp(.0798164) , exp(.2548837)
display exp(.9856338) , exp(.6388907) , exp(1.332377)

```

### 3.9 Salidas computacionales

```
. glm DIAS_ESTAD i.SEXO EDAD i.COND_EGR i.INTERV_Q , family(poisson) link(log) nolog

Generalized linear models                No. of obs   =    2,500
Optimization      : ML                   Residual df   =    2,495
                                                Scale parameter =    1
Deviance          = 25120.84905           (1/df) Deviance = 10.06848
Pearson          = 42996.34405           (1/df) Pearson  = 17.233

Variance function: V(u) = u                [Poisson]
Link function     : g(u) = ln(u)           [Log]

Log likelihood    = -16786.87431           AIC          = 13.4335
                                                BIC          = 5599.854
```

DIAS_ESTAD	OIM				
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
2.SEXO	-.0466306	.0138991	-3.35	0.001	-.0738724 -.0193888
EDAD	.0127074	.0007871	16.15	0.000	.0111648 .01425
2.COND_EGR	.3641415	.0242867	14.99	0.000	.3165404 .4117425
2.INTERV_Q	.1273875	.0152877	8.33	0.000	.0974243 .1573508
_cons	1.090755	.0578063	18.87	0.000	.9774565 1.204053

```
. glm DIAS_ESTAD EDAD i.COND_EGR i.INTERV_Q , family(nbinomial) link(log) nolog

Generalized linear models                No. of obs   =    2,500
Optimization      : ML                   Residual df   =    2,496
                                                Scale parameter =    1
Deviance          = 2606.752447           (1/df) Deviance = 1.044372
Pearson          = 4594.943922           (1/df) Pearson  = 1.840923

Variance function: V(u) = u+(1)u^2        [Neg. Binomial]
Link function     : g(u) = ln(u)           [Log]

Log likelihood    = -7903.871395           AIC          = 6.326297
                                                BIC          = -16922.07
```

DIAS_ESTAD	OIM				
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
EDAD	.013396	.002405	5.57	0.000	.0086823 .0181098
2.COND_EGR	.4134707	.0873262	4.73	0.000	.2423146 .5846269
2.INTERV_Q	.16735	.0446608	3.75	0.000	.0798164 .2548837
_cons	.9856338	.176913	5.57	0.000	.6388907 1.332377

Las siguiente imagenes fueron obtenidas del sitio oficial del DEIS.



## DICCIONARIO DE BASE DE DATOS, EGRESOS HOSPITALARIOS 2016

Nombre del campo	Tipo de dato	Largo	Descripción
ServicioSalud	Numérico	2	Código Servicio de Salud, corresponde al Servicio de Salud de ocurrencia. ( <a href="#">Código de Servicio de Salud</a> )
Seremi	Numérico	2	Código SEREMI, corresponde a la SEREMI de ocurrencia. ( <a href="#">Código de SEREMI</a> )
ESTAB	Alfanumérico	6	Código de Establecimiento, corresponde al Establecimiento de ocurrencia. ( <a href="#">Código de Establecimientos de Salud</a> )
SEXO	Numérico	1	Sexo, los valores aceptados son: 1 = Hombre 2 = Mujer 3 = Indeterminado 9 = Desconocido
EDAD	Numérico	3	Edad en cantidad, expresada en año.
PREVI	Numérico	1	Previsión en salud, los valores aceptados son: 1 = FONASA 2 = ISAPRE 3 = SIN PREVISIÓN 5 = CAPREDENA 6 = DIPRECA 7 = OTRA 9= IGNORADA
BENEF	Numérico	1	Clase de Beneficiario, corresponde a los tramos de FONASA: A B C D
MOD	Numérico	1	Modalidad de Atención de los beneficiarios FONASA: 1 = MAI (Modalidad Atención Institucional) 2 = MLE (Modalidad Libre Elección)
COMUNA	Alfanumérico	5	Código Comuna, ( <a href="#">Código de comuna</a> )
FECHA_EGR	Fecha	10	Fecha de alta del paciente, formato fecha DD/MM/AAAA.
SERC_EGR	Alfanumérico	3	Servicio Clínico de Egreso o nivel de cuidado, ( <a href="#">Servicio Clínico o nivel de cuidado de Egreso</a> )
DIAS_ESTAD	Numérico	4	Días estada total.
DIAG1	Alfanumérico	4	Diagnóstico principal, corresponde al código de la CIE-10 ( <a href="#">Diagnóstico principal</a> )

(DEIS, 2017).



## DICCIONARIO DE BASE DE DATOS, EGRESOS HOSPITALARIOS 2016

DIAG2	Alfanumérico	4	Causa externa, corresponde al código de la CIE-10 cuando el diagnóstico principal corresponde a: "Traumatismos, envenenamientos y algunas otras consecuencias de causas externas". ( <a href="#">Causa externa</a> )
COND_EGR	Numérico	1	Condición al egreso, los valores aceptados son: 1 = Vivo 2 = Fallecido
INTERV_Q	Numérico	1	Intervención Quirúrgica, los valores aceptados son: 1 = Sí 2 = No
REGION	Numérico	2	Región de residencia, campo creado a partir de la comuna de residencia del paciente, de acuerdo a la División Político Administrativo vigente desde el año 2008 ( <a href="#">Código de Región</a> )
SERV_RES	Numérico	2	Servicio de Salud de referencia, campo creado a partir de la comuna de residencia del paciente, de acuerdo a la División Político Administrativo vigente desde el año 2008 ( <a href="#">Código de Servicio de Salud</a> )



(DEIS, 2017).



## EGRESOS HOSPITALARIOS 2011

### Servicios clínicos

Código	Descripción
010	Área médica indiferenciada
020	Área quirúrgica indiferenciada
110	Medicina básica
111	Neumología
112	Infeccioso adulto
113	Tratamiento antialcohólico
114	Nutrición
115	Medicina agudo
116	Agudo Indiferenciado
119	Área méd. quirúrg. indiferenc.
120	Cirugía básica
121	Cirugía tórax
122	Cirugía cardiovascular
123	Cirugía plástica-quemado adul.
124	Cirugía máxilo facial
125	Cirugía agudos
130	Traumatología-Ortopedia adulto
131	Traumatología agudo
132	Traumatología-Ortopedia Indif.
140	Cirugía infantil
141	Cirugía partes blandas Inf
142	Cirugía plástica-quemado Inf.
143	Traumatología y Ortopedia inf.
144	Cardiociugía infantil
145	Área quirúrg. infantil indif.
150	Pediatría indiferenciada
151	Neonatología incubadora
152	Neonatología cuna
153	Lactante
154	Segunda infancia
155	Infecciosos niño
156	Pediatría agudo
160	Obstetricia y ginecología
161	Obstetricia



(DEIS, 2017).



## EGRESOS HOSPITALARIOS 2011

### Servicios clínicos

162	Ginecología
170	Dermatología
180	Neuropsiquiatría infantil
190	Neurología adulto
191	Neurología adulto agudo
200	Neurocirugía indiferenciado
201	Neurocirugía adulto
202	Neurocirugía infantil
210	Psiquiatría
211	Psiquiatría corta estadia
212	Psiquiatría crónico
213	Desintoxicación alcohol y drogas
214	Psiquiatría forense mediana complejidad
215	Psiquiatría forense alta complejidad
216	Psiquiatría mediana estadia
217	Externalización Unidad Forense
218	U. Evaluac. Personas Imputadas
220	Oftalmología
230	Oncología indiferenciado
231	Oncología adulto
232	Oncología Infantil
240	Otorrinolaringología
245	Especialidad quirúrgica
250	Urología
260	Medicina Física-Rehabilitación
270	Derivación médico quirúrgico
271	Deriva. Méd.-Quirúrg. Respira.
280	Tisiología crónico
290	Geriatría
300	U. Emergencia Indiferenciado
301	U. Emergencia Adulto
302	U. Emergencia Niño
310	U.C.I. Indiferenciado
311	U.C.I. Neonatología
312	U.C.I. Adulto
313	U.C.I. Pediatría
314	U.C.I. Coronaria



## EGRESOS HOSPITALARIOS 2011

### Servicios clínicos

315	U.C.I. Quemado
320	U.T.I. Indiferenciado
321	U.T.I. Medicina
322	U.T.I. Cirugía
323	U.T.I. Pediatría
324	U.T.I. Neonatología
325	U.T.I. Quemado
326	U.T.I. Neurocirugía
327	Unidad coronaria intermedia
330	Pensionado
331	Pensionado oftalmología
999	Indiferenciado



(DEIS, 2017).



## EGRESOS HOSPITALARIOS 2011

### Comunas

04201	ILLAPEL	04	De Coquimbo
04104	LA HIGUERA	04	De Coquimbo
04101	LA SERENA	04	De Coquimbo
04203	LOS VILOS	04	De Coquimbo
04303	MONTE PATRIA	04	De Coquimbo
04301	OVALLE	04	De Coquimbo
04105	PAIGUANO	04	De Coquimbo
04304	PUNITAQUI	04	De Coquimbo
04305	RÍO HURTADO	04	De Coquimbo
04204	SALAMANCA	04	De Coquimbo
04106	VICUÑA	04	De Coquimbo
05602	ALGARROBO	05	De Valparaíso
05402	CABILDO	05	De Valparaíso
05502	CALERA	05	De Valparaíso
05302	CALLE LARGA	05	De Valparaíso
05603	CARTAGENA	05	De Valparaíso
05102	CASABLANCA	05	De Valparaíso
05702	CATEMU	05	De Valparaíso
05103	CON CON	05	De Valparaíso
05604	EL QUISCO	05	De Valparaíso
05605	EL TABO	05	De Valparaíso
05503	HIJUELAS	05	De Valparaíso
05201	ISLA DE PASCUA	05	De Valparaíso
05104	JUAN FERNÁNDEZ	05	De Valparaíso
05504	LA CRUZ	05	De Valparaíso
05401	LA LIGUA	05	De Valparaíso
05802	LIMACHE	05	De Valparaíso
05703	LLAILLAY	05	De Valparaíso
05301	LOS ANDES	05	De Valparaíso
05506	NOGALES	05	De Valparaíso
05803	OLMUÉ	05	De Valparaíso
05704	PANQUEHUE	05	De Valparaíso
05403	PAPUDO	05	De Valparaíso
05404	PETORCA	05	De Valparaíso
05105	PUCHUNCAVÍ	05	De Valparaíso
05705	PUTAENDO	05	De Valparaíso
05501	QUILLOTA	05	De Valparaíso

(DEIS, 2017).



## EGRESOS HOSPITALARIOS 2011

### Comunas

05801	QUILPUE	05	De Valparaíso
05107	QUINTERO	05	De Valparaíso
05303	RINCONADA	05	De Valparaíso
05601	SAN ANTONIO	05	De Valparaíso
05304	SAN ESTEBAN	05	De Valparaíso
05701	SAN FELIPE	05	De Valparaíso
05706	SANTA MARIA	05	De Valparaíso
05606	SANTO DOMINGO	05	De Valparaíso
05101	VALPARAISO	05	De Valparaíso
05804	VILLA ALEMANA	05	De Valparaíso
05109	VIÑA DEL MAR	05	De Valparaíso
05405	ZAPALLAR	05	De Valparaíso
13502	ALHUÉ	13	Metropolitana de Santiago
13402	BUIN	13	Metropolitana de Santiago
13403	CALERA DE TANGO	13	Metropolitana de Santiago
13102	CERRILLOS	13	Metropolitana de Santiago
13103	CERRO NAVIA	13	Metropolitana de Santiago
13301	COLINA	13	Metropolitana de Santiago
13104	CONCHALÍ	13	Metropolitana de Santiago
13503	CURACAVÍ	13	Metropolitana de Santiago
13105	EL BOSQUE	13	Metropolitana de Santiago
13602	EL MONTE	13	Metropolitana de Santiago
13106	ESTACIÓN CENTRAL	13	Metropolitana de Santiago
13107	HUECHURABA	13	Metropolitana de Santiago
13108	INDEPENDENCIA	13	Metropolitana de Santiago
13603	ISLA DE MAIPO	13	Metropolitana de Santiago
13109	LA CISTERNA	13	Metropolitana de Santiago
13110	LA FLORIDA	13	Metropolitana de Santiago
13111	LA GRANJA	13	Metropolitana de Santiago
13112	LA PINTANA	13	Metropolitana de Santiago
13113	LA REINA	13	Metropolitana de Santiago
13302	LAMPA	13	Metropolitana de Santiago
13114	LAS CONDES	13	Metropolitana de Santiago
13115	LO BARNECHEA	13	Metropolitana de Santiago
13116	LO ESPEJO	13	Metropolitana de Santiago
13117	LO PRADO	13	Metropolitana de Santiago
13118	MACUL	13	Metropolitana de Santiago

(DEIS, 2017).

---

## Referencias

---

- Alcaide Delgado, M. (2015). *Modelo de Regresión Binomial Negativa*. Universidad de Sevilla, Sevilla, España.
- Ceballos Acevedo, T. Velásquez Restrepo, P. Jaén Posada, J. (2014). *Duración de la estancia hospitalaria. Metodologías para su intervención*. Rev. Gerenc. Polít. Salud. 2014; 13(27): 274-295. <http://dx.doi.org/10.11144/Javeriana.rgyaps13-27.dehm>.
- Contreras Vilca, N. (2012). *Análisis de votos electorales usando modelos de regresión para datos de conteo* (tesis de pregrado). Universidad de Valparaíso, Valparaíso, Chile
- Goldstein, H. (2011). *Multilevel statistical models* – 4th ed. West Sussex, United Kingdom. John Wiley & Sons, Ltd.
- DEIS. (2017). Departamento de Estadística e Información de Salud. Chile: *deis.cl*: <http://www.deis.cl/bases-de-datos-egresos-hospitalarios/>
- INE. (2007). *Enfoque estadístico adulto mayor en Chile*. Boletín informativo del Instituto Nacional de Estadística.
- Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. New York, USA. Springer Science y Business Media LLC.
- López, P. A. (2016). *Modelos de regresión Poisson y Binomial Negativa en modelos lineales generalizados aplicados a datos correspondientes a accidentes de tránsito y lesionados entre los años 2011 y 2014 en la región de Valparaíso* (tesis de pregrado). Universidad de Valparaíso, Valparaíso, Chile.

- McCullagh, P. y Nelder, J.A. (1989). *An outline of generalized linear models in Generalized Linear Models second edition* (21-44). New York: Chapman and Hall.
- Navarro, A. Utzetb, F. Puigb, P. Caminala, J. Martína, M. (2001). *La distribución Binomial Negativa frente a la de Poisson en el análisis de fenómenos recurrentes*. Gac Sanit, 15, pp. 447-452.
- Osses, F. (2013). *Aplicación de los residuos en modelos lineales generalizados* (tesis de pregrado). Universidad de Valparaíso, Valparaíso, Chile.
- Sandoval, G. (2013). Hospitales reducen a seis días el plazo promedio de internación de pacientes. Chile: *La Tercera*. Extraído el Martes 18 de Julio del 2017 a las 15:45 desde <http://www.latercera.com/noticia/hospitales-reducen-a-seis-dias-el-plazo-promedio-de-internacion-de-pacientes/>