Universidad de Valparaíso
CHILE
Facultad de Ciencias
Instituto de Estadística

Estimación del tonelaje de alimentación de mineral en molino SAG 1011 y molinos de Bolas 1012 y 1013, mediante modelos lineales generalizados

Trabajo presentado por:
Katalina Belén Lizana Rosales

Trabajo de titulación para optar al título de:
Ingeniero en Estadística

Supervisado por:
Claudia Navarro Villarroel, Ph.D.

Enero de 2019, Valparaíso, Chile

Agradecimientos

En primer lugar quiero agradecer a Juan Madrid Varas (Q.E.P.D), quién fue un pilar fundamental en mi formación y una de las personas más importantes en mi vida, ya que, siempre me apoyó y alentó a esforzarme en los momentos más difíciles.

También debo agradecer a mi madre y hermano por el apoyo incondicional durante mi vida y sobre todo en esta etapa.

Agradezco a los profesores del Instituto de Estadística de la Universidad de Valparaíso, debido a su dedicación con los estudiantes, en especial a la profesora Claudia Navarro Villarroel por aceptar guiarme en esta culminación de mi vida académica.

Debo agradecer a los profesionales pertenecientes a la Compañía Minera Doña Inés de Collahuasi, por brindarme la oportunidad de trabajar en conjunto y aportar diversos conocimientos del área de la minería, también por la entrega y disposición de material ofrecido para poder realizar este trabajo de titulación.

Finalmente debo agradecer a mis amigos con quienes finalizamos esta etapa de Universidad, ya que estuvieron siempre disponible cuando los necesité, me brindaron el apoyo y alentaron a continuar con este desafío.

Abreviaturas

A lo largo de este trabajo de título, se utilizaron diferentes abreviaturas para referirse a distintos conceptos de interés, las cuales se presentan a continuación:

- CMDIC: Compañía Minera Doña Inés de Collahuasi.
- SAG: Molino semi-autógeno.
- MB: Molino de Bolas.
- MLG: Modelos Lineales Generalizados.
- MV: Máxima Verosimilitud.
- M.s.n.m: Metros sobre el nivel del mar.
- T/d: Toneladas diarias.
- SCM: Sociedad Contractual Minera.

Resumen

Hoy en día, la estadística ha tomado una gran relevancia en la minería, debido a que se necesita obtener una mayor producción, confiabilidad y optimización en los diferentes procesos.

Para llevar a cabo este trabajo de titulación se realizó un análisis de datos a partir de los Modelos Lineales Generalizados (MLG). Los MLG se utilizan cuando la variable respuesta sigue una distribución perteneciente a la familia exponencial, en donde las observaciones en estudio deben ser independientes entre sí.

Con el objetivo de optimizar el proceso de la molienda en la Compañía Minera Doña Inés de Collahuasi, se propuso estimar el tonelaje de alimentación de mineral en molinos SAG 1011, Bolas 1012 y 1013, mediante modelos lineales generalizados.

Esto se logró mediante un análisis estadístico utilizando datos reales obtenidos en la planta concentradora de la Compañía Minera Doña Inés de Collahuasi, el cual se efectúo mediante la utilización de el software estadístico Stata SE Versión 15.

Abstract

Nowadays, statistics has taken a great importance in mining, because it is necessary to obtain greater production, reliability and optimization in the different processes.

To carry out this study, a data analysis was done out using Generalized Linear Models (GLM). GLM are used when the response variable follows a distribution belonging to the exponential family, where the observations under study, must be independent of each other.

With the objective of optimizing the grinding process at the Doña Inés de Collahuasi Mining Company, it was proposed to estimate the tonnage of mineral feed in SAG 1011, Balls 1012 and 1013 mills, using generalized linear models.

This was achieved through a statistical analysis using real data obtained in the concentrator plant of the Doña Inés de Collahuasi Mining Company, for this Stata SE statistical software Version 15 was used.

Índice general

Resumen **IV**

Abstract **V**

Objetivos **1**

Hipótesis **2**

1. Introducción **3**

2. Modelos lineales generalizados **9**
2.1. Componente aleatoria 10
2.2. Componente sistemática 11
2.3. Función de enlace 11
2.4. Bondad de ajuste 13
2.5. Análisis de residuos 16

3. Modelos lineales generalizados Gaussiano y Gamma **18**
3.1. Modelo lineal generalizado Gaussiano 18
3.1.1. Componente aleatoria 19
3.1.2. Componente sistemática 19
3.1.3. Función de enlace 19
3.2. Modelo lineal generalizado Gamma 20
3.2.1. Componente aleatoria 20
3.2.2. Componente sistemática 20
3.2.3. Función de enlace 21

4. Aplicación **22**
4.1. Análisis de datos 22
4.1.1. Análisis exploratorio 24
4.1.2. Aplicación del modelo lineal generalizado Gaussiano 28
4.1.3. Aplicación del modelo lineal generalizado Gamma 31
4.2. Comparación de modelos 34

5. Conclusiones **35**

Índice de figuras

1.1. Ubicación CMDIC. 4

4.1. Histograma de la variable respuesta. 25
4.2. Matriz de correlación. 27
4.3. Curva de ajuste a la variable respuesta con datos Gaussianos. 28
4.4. Residuos de Anscombe v/s predicción de la media para el Tonelaje de alimentación. 30
4.5. Curva de ajuste a la variable respuesta con datos Gamma. 31
4.6. Residuos de Anscombe v/s predicción de la media para el Tonelaje de alimentación. 33

Índice de tablas

4.1. Esquema del conjunto de datos. 24
4.2. Estadística descriptiva del Tonelaje de alimentación. 25
4.3. Estadística descriptiva del conjunto de datos. 26
4.4. Correlación entre Tonelaje de alimentación y variables independientes. 27
4.5. Estimación de parámetros distribución Gaussiana. 28
4.6. Estimación de los coeficientes del MLG Gaussiano con función de enlace identidad. 29
4.7. Coeficientes e IC del 95 % MLG Gaussiano. 29
4.8. Estimación de parámetros distribución Gamma. 31
4.9. Estimación de los coeficientes del MLG Gamma con función de enlace recíproca. 32
4.10. Coeficientes e IC del 95 % MLG Gamma. 32
4.11. Comparación de medidas de bondad de ajuste. 34

Objetivos

El objetivo principal a abarcar en el presente trabajo de titulación es: estudiar los modelos lineales generalizados, para así aplicarlos en datos reales relacionados a la molienda de la Compañía Minera Doña Inés de Collahuasi SCM.

De este objetivo general, se desprenden los siguientes objetivos específicos:

i. Analizar las propiedades de los modelos lineales generalizados.

ii. Aplicar los modelos lineales generalizados a un conjunto de datos reales.

iii. Proponer el modelo lineal generalizado definitivo para los datos utilizados.

Hipótesis

A través de un modelo lineal generalizado, es posible estimar el tonelaje de alimentación del mineral en molinos SAG 1011, Bolas 1012 y 1013, para así lograr obtener un resultado óptimo en la producción minera.

Debido a que el flujo de cobre producido es una relación directa entre el tonelaje procesado, el contenido del metal en el mineral y las distintas características metalúrgicas que influyen en la producción del mineral, se tiene que si el tonelaje de alimentación de mineral es bajo, se dañan los revestimientos del molino y por el contrario, si el tonelaje de alimentación es alto, existe una sobrecarga del molino, lo cual demanda una mayor potencia del motor y por ende, un mayor consumo de energía. Es por esto que para este trabajo de título se propuso estimar el tonelaje de alimentación de mineral en molinos SAG 1011, Bolas 1012 y 1013 (pertenecientes a la Línea 3), para optimizar el proceso de la molienda en la Compañía Minera Doña Inés de Collahuasi SCM.

Capítulo 1

Introducción

La estadística impacta prácticamente en todos los aspectos de la vida cotidiana, debido a que a partir de las diferentes actividades que se realizan día a día es posible recopilar datos, los cuales una vez siendo analizados, permiten tomar decisiones.

La estadística se puede definir como aquella ciencia que estudia los fenómenos inciertos o las situaciones que no se pueden predecir con certeza, pero de la que sí se puede recabar información. En consecuencia, cada vez se hace más necesario conocer el comportamiento de aquellas variables que aportan información relevante al momento de realizar la toma de decisiones ante algún estudio de interés.

Se suelen considerar diferentes técnicas estadísticas, las cuales permiten obtener la información requerida, una de las técnicas más utilizadas es el análisis de regresión lineal, la cual sirve para investigar y modelar la relación entre variables.

Son numerosas las aplicaciones de la regresión, y las hay en casi todos los campos profesionales, incluyendo en ingeniería, ciencias físicas y químicas, economía, administración, ciencias biológicas, ciencias sociales, entre otras (Montgomery, Peck, y Vining, 2006). Sin embargo, para la utilización de esta técnica se deben considerar los supuestos respectivos para llevar a cabo este análisis, los cuales son: la no colinealidad entre las variables independientes, la homocedasticidad de la varianza, la linealidad entre las variables y la independencia entre los residuos. En el caso de la regresión lineal ordinaria se requiere además la normalidad en la variable respuesta tanto como en los residuos.

En el análisis univariado, el modelo lineal general surge de la necesidad de cuantificar las relaciones entre un conjunto de variables, en la que una de ella se denomina variable respuesta o dependiente y las restantes son las variables explicativas o independientes. Este modelo asume que la variable dependiente sigue una distribución normal y con varianza homocedástica (McCulloch y Searle, 2001).

También es muy común que a medida que aumenta la media muestral, aumente también su varianza, estos problemas se pueden llegar a solucionar mediante la transformación de la variable respuesta (por ejemplo, tomando logaritmos). Sin embargo, estas transformaciones no siempre consiguen corregir la falta de normalidad, la heterocedasticidad o la no linealidad de los datos (Cayuela, 2009).

Debido a que no siempre es posible cumplir con dichos supuestos correspondientes para un análisis de regresión, principalmente debido a la normalidad de los datos a analizar, surge la necesidad de estudiar otros modelos, es aquí donde se le da lugar a los Modelos Lineales Generalizados (MLG), los cuales no necesariamente se basan en considerar una variable respuesta que siga una distribución de tipo normal, si no, que para utilizar dichos modelos solo se necesita que la variable respuesta provenga de una distribución perteneciente a la familia exponencial, la estimación de los parámetros de los MLG se lleva a cabo a través de estimadores máximo verosímil.

Antecedentes de la empresa

Ubicación

La Compañía Minera Doña Inés de Collahuasi SCM, se encuentra ubicada en la comuna de Pica, región de Tarapacá a 4400 m.s.n.m. en el extremo norte de Chile. Su explotación se realiza a rajo abierto y posee tres yacimientos de mineral de cobre. En la Figura 1.1 se pueden apreciar los yacimientos de la compañía. Esta zona altiplánica se caracteriza por tener un clima lluvioso en verano y nevadas ocasionales en invierno.

Figura 1.1: Ubicación CMDIC.
Fuente: Google Maps.

La Compañía Minera Doña Inés de Collahuasi SCM es dedicada fuertemente a la extracción y producción de concentrado de cobre y molibdeno, logrando ser la segunda mayor operación de cobre en el mundo y junto con eso un de los mayores depósitos de recursos minerales de cobre del planeta. Existen tres yacimientos (Rosario, Ujina y Huinquintipa), e instalaciones industriales en el área cordillera y en el área puerto (filtrado y embarque) en Punta Patache (Iquique), sin embargo actualmente solo se explota el yacimiento ubicado en Rosario (Collahuasi, 2018).

Historia

Mediante la explotación de diversos sistemas de vetas los cuales están compuestos por cobre y plata de alta ley, en 1880 la Compañía Minera Doña Inés de Collahuasi llevó a cabo su actividad comercial, sin embargo, debido a una gran crisis económica esta operación finalizó en el año 1930, luego de este suceso las actividades en el sector se reiniciaron en el año 1978, en donde se identificaron distintos componentes en el yacimiento de Rosario (Collahuasi, 2018).

En 1991 se descubrió el yacimiento Ujina, a través de varios estudios de levantamientos de tipO aerofotogramétricos, imágenes satelitales, actividades de perforación de sondajes, entre otros. Posteriormente, en el año 1995 mediante ciertos estudios de factibilidad e impacto ambiental sobre el proyecto Collahuasi, este fue aprobado. Luego, a fines del año 1996, se comenzó con una etapa de desarrollo y construcción, la cual requirió una inversión de US$ 1.800 millones (Collahuasi, 2018).

Finalmente el día 7 de Abril de 1999 la Compañía abrió su operación comercial, y desde esa fecha se dedica a desarrollar sus actividades basándose en buenas prácticas, las cuales colaboran con el desarrollo económico, bienestar en la sociedad y cuidado de el medio ambiente en la región de Tarapacá (Collahuasi, 2018).

Posteriormente, se aprobó la ejecución del proyecto de fusión Ujina-Rosario, el cual se basó en el traslado de las actividades desde el yacimiento de Ujina hacia el de Rosario, se produjo un aumento de capacidad de tratamiento de la planta concentradora, el cual fue de 60.000 a 110.000 toneladas diarias (t/d). La inversión para el desarrollo de este proyecto fue de US$ 624 millones, el cual se inició en el año 2002 y finalizó el año 2004 (Collahuasi, 2018).

Por otra parte, en el año 2005 comenzó la construcción del proyecto planta de molibdeno, el cual consistía en la construcción de una planta de flotación selectiva de molibdeno en el sector Puerto Patache (Collahuasi, 2018).

El año 2016 Collahuasi generó 506.500 toneladas de cobre fino con ingresos de US$ 2.286 millones, con un costo operacional de US$ 1,14 la libra. Donde el 70 % de las exportaciones de concentrado de cobre tiene como destino China y otros países asiáticos, y el 24 % se comercializó en el mercado nacional (Collahuasi, 2018).

En la actualidad la compañía produce y comercializa concentrado de cobre y concentrado de molibdeno (Collahuasi, 2018).

Potencial minero

Define el valor actual de una minera y se compone por los recursos y reservas que tiene un yacimiento. La compañía minera Collahuasi posee una reserva minera de 3.088 millones de toneladas con una ley promedio de 0,9 %. Esto implica la presencia de 27.741.541 toneladas de cobre fino. Con respecto a los recursos minerales de la empresa, se contabilizan 9.975 millones de toneladas con una ley promedio de 0.78 % y 77.677.945 toneladas de cobre fino (Collahuasi, 2018).

Descripción de la molienda

Se da comienzo a el proceso de molienda una vez que el mineral chancado acoplado en la pila de almacenamiento es entregado al molino SAG. Existen 8 puntos de transferencia de material chancado llamados chutes, los cuales descargan bajo la extensión de la pila de almacenamiento, cada uno de estos descargan a un alimentador que saca el mineral desde la pila. En operación normal, 4 de los 8 alimentadores operan al mismo tiempo para asegurar el tonelaje de alimentación nominal al molino SAG y para conseguir una distribución adecuada de mineral en términos de granulometría, que en forma natural se segrega en la pila de almacenamiento. Cuando los alimentadores extraen mineral en una forma en que la pila de almacenamiento se va bajando en forma pareja, el mineral tiende a tener una distribución de tamaños constante. Un mezclado apropiado reduce las fluctuaciones rápidas en las características del mineral (dureza, tamaño, ley, humedad, entre otras) y produce una operación más estable en la planta (Collahuasi, 2018).

Abajo de la pila de almacenamiento existe un túnel, el cual conduce a la superficie. Existen 8 chutes de descarga de la pila de almacenamiento sobre la techumbre del túnel. El mineral se descarga a través de los chutes sobre los alimentadores, los cuales entregan la carga bajo un tonelaje controlado a una correa transportadora que alimenta al molino SAG. Cierta cantidad del mineral de la pila de almacenamiento se encuentra sobre los chutes de descarga de los alimentadores. Cuando los alimentadores están en operación, este mineral (la carga viva) fluye libremente dentro de los chutes (Collahuasi, 2018).

En dirección a el molino SAG, el mineral chancado pasa sobre un determinado pesómetro, el cual se encuentra instalado en la correa que alimenta a el molino SAG. Este pesómetro es de gran relavancia, no solo por los balances metalúrgicos, sino que porque este controla el tonelaje de alimentación hacia los molinos. Existe una señal de el tonelaje de pesaje de el

pesómetro, la que se usa para el controlador de peso pa así determinar la velocidad de los alimentadores de la pila de almacenamiento (Collahuasi, 2018).

Antes de que la correa de alimentación del molino SAG salga del túnel y después del paso por el pesómetro, pasa por debajo del sistema de alimentación de bolas al molino SAG, que consiste de 2 silos de 350 toneladas de capacidad cada uno, ambos contienen bolas de 5 pulgadas. Cada silo se encuentra equipado con una válvula rotatoria de bola. Se debe destacar que las bolas de molienda son agregadas periódicamente sobre el material. Posteriormente, el mineral y las bolas de molienda son descargadas en el chute de alimentación del molino SAG, en donde se procede a la agregación de agua de proceso y lechada de cal. Debido a esto, la velocidad regular del molino SAG puede variar, ya que esto regula la acción de cascada en el molino y logra proporcionar un medio de control al tonelaje de molienda (Collahuasi, 2018).

El mineral, los reactivos, las bolas de molienda, y el agua se mezclan y actúan en forma de cascada dentro de el molino, en donde, el tamaño del mineral es reducido hasta que fluya a través de las parrillas descarga del molino. Luego, el agua de este proceso se agrega a el chute de alimentación en proporción al peso de mineral fresco que se alimenta desde la pila de almacenamiento. Un controlador ajusta el flujo de agua automáticamente para igualar los cambios de tonelaje de mineral fresco y produce una densidad de pulpa especificada en el molino SAG (Collahuasi, 2018).

La carga de un molino SAG consta de mineral fresco, bolas de molienda de acero, retorno de mineral no procesado y agua. Posteriormente el molino realiza fuertes giros, con el fin de volcar todo su contenido para así provocar la acción de ruptura del material. La carga total del molino ocupa cerca del 22 % del volumen del molino, este está diseñado para lograr manipular el total de carga de hasta alcanzar un 30 % del volumen del molino. Las bolas de molienda por sí solas normalmente ocupan casi el 15 % del volumen en el molino. Los molinos están diseñados para contener un volumen máximo de bolas de un 15 % (estos volúmenes de llenado son aproximados) (Collahuasi, 2018).

El molino se encuentra recubierto de revestimientos de acero con cromo de molibdeno el cual resiste a el desgaste, con el fin de proteger la carcasa. Los revestimientos de los molinos están acondicionados con ciertos levantadores, los que sirven para ayudar a subir la carga durante la rotación del molino (Collahuasi, 2018).

El proceso de molienda de mineral se encuentra diseñado para poder disminuir el tamaño original del mineral chancado el cual se encuentra en la etapa de preparación para el posterior proceso de concentración de mineral mediante la flotación. El molino SAG es considerado una etapa en el proceso de conminución (reducción de tamaño de mineral), mientras que los dos molinos de Bolas (1012 y 1013) constituyen otra etapa, justo después del molino SAG. El objetivo del proceso en un molino de bolas es lograr reducir el mineral de un tamaño de alimentación de 20 % más grande hacia un tamaño de producto de 80 % más fino. Se ha encontrado que este tamaño obtenido por la reducción del mineral, es el óptimo como alimentación para el proceso de flotación (Collahuasi, 2018).

El molino de bolas está diseñado como un sistema de molienda en circuito cerrado. Esto quiere decir que las diferentes partículas de mineral no deben salir del circuito de molino de bolas, estas salen únicamente cuando el tamaño de estas partículas son suficientemente finas, con el fin de pasar por el rebose del sistema de clasificación de ciclones. Finalmente la pulpa del rebose alimentará los procesos de flotación (Collahuasi, 2018).

Área de trabajo

La compañía se divide en diferentes áreas, sin embargo este trabajo de título está centrado en la Vicepresidencia de Procesos de la compañía, específicamente en la Gerencia Metalurgia y Control de Procesos, con la cual se está trabajando en conjunto para dar solución a este análisis (Collahuasi, 2018).

En la minería, la gerencia de metalurgia es la encargada de definir, planificar, ejecutar y controlar el cumplimiento de las metas de producción y costos definidos para el proceso. Por otra parte, también debe gestionar la implementación de optimizaciones de procesos, desarrollar e incorporar nuevas tecnologías y metodologías para el procesamiento de minerales (Collahuasi, 2018).

Debido a que el flujo de cobre producido es una relación directa entre el tonelaje procesado, el contenido del metal en el mineral y las distintas características metalúrgicas que influyen en la producción del mineral. Si el tonelaje de alimentación de mineral es bajo, se dañan los revestimientos del molino y por el contrario, si este es alto, existe una sobrecarga del molino, lo cual demanda una mayor potencia del motor y un mayor consumo de energía. Es por esto que el objetivo principal de este trabajo de título es estimar el tonelaje de alimentación de mineral en molinos SAG 1011, Bolas 1012 y 1013 (pertenecientes a la Línea 3), para así optimizar el proceso de la molienda en la Compañía Minera Doña Inés de Collahuasi.

En el siguiente capítulo se dará a conocer la teoría de los modelos lineales generalizados, la cual se debe considerar para un correcto análisis.

Capítulo 2

Modelos lineales generalizados

En los diversos campos de estudio, es muy relevante la importancia que tiene la estadística aplicada para el desarrollo de investigaciones, ya que cada vez son más los profesionales de diferentes áreas que requieren de técnicas estadísticas para dar validez a sus estudios, tales como: simulación, muestreo, diseño de experimentos, modelamiento estadístico, inferencia estadística, entre otros. Para así poder llevar a cabo la recolección, el análisis de datos y posteriormente, llegar a una interpretación de estos.

Un modelo es una descripción matemática de los procesos que se creen dan lugar a las observaciones de un conjunto de datos. Los modelos estadísticos incluyen una ecuación que describe el presunto impacto de las variables explicativas y una descripción de las distribuciones de probabilidad asociadas con aspectos del proceso que se asumen caracterizan la variación aleatoria. Estos modelos incluyen una parte sistemática y una aleatoria (Stroup, 2012).

Para llevar a cabo un correcto análisis de los datos, se debe distinguir el tipo de variables que se desea a modelar (variables de conteo, continuas, categóricas, etc) y la relación existente entre las covariables y la variable respuesta, para así lograr distinguir entre los distintos tipos de modelos existentes, cual es el adecuado para de explicar la variable de interés de forma correcta.

Debido a que los modelos de regresión tradicionales, tales como los modelos de regresión lineal (simple o múltiple) se rigen bajo los supuestos ya mencionados anteriormente, limita a modelar aquellas variables que no cumplan con estos supuestos. Como consecuencia de esto, se presenta una extensión de la teoría de los modelos lineales, los cuales fueron estudiados por Nelder y Wedderburgn (1972), McCullagh y Nelder (1989), en donde se considera la posibilidad de modelar variables continuas o categóricas, sin regirse por la normalidad de la variable dependiente, sino que se permite modelar cualquier variable perteneciente a la familia exponencial, tales como las distribuciones Weibull, Beta, Bernoulli, Binomial, Geométrica, Poisson, Gamma, Hipergeométrica, Binomial negativa, entre otras. Esta técnica estadística de modelamiento es denominada como modelo lineal generalizado, en donde existen tres componentes importantes: 1)componente aleatoria, correspondiente a la variable respuesta Y, con función de probabilidad perteneciente a la familia exponencial, 2)componente sistemática,

correspondiente a la combinación lineal de las variables regresoras y 3) función de enlace, la cual cumple con enlazar la componente aleatoria con la sistemática (McCullagh y Nelder, 1989).

A continuación se dan a conocer detalladamente los componentes pertenecientes a los modelos lineales generalizados:

2.1. Componente aleatoria

La componente aleatoria del modelo corresponde a la variable respuesta Y que tiene una función de densidad asignada perteneciente a la familia exponencial.

Esta consiste en un variable aleatoria Y con observaciones independientes $(y_1, \ldots, y_n)$. En muchas aplicaciones, las observaciones Y son binarias y se identifican como éxito y fracaso. Aunque de modo más general, cada Y_i, $i = 1, 2, ..., n$ indicaría el número de éxitos entre un número fijo de ensayos y se modelaría como una distribución binomial. En otras ocasiones, cada observación es un recuento, con lo que se puede asignar a Y una distribución Poisson o una distribución Binomial Negativa. Finalmente, si las observaciones son continuas se puede asumir para Y una distribución Normal o alguna otra distribución, dependiendo del recorrido de la variable. Todos estos modelos se pueden incluir dentro de la llamada familia exponencial, definida por los parámetros θ y ϕ lo que se representa se la siguiente forma:

$$f(y; \theta, \phi) = exp\left\{\frac{y\theta - b(\theta)}{\alpha(\phi)} + c(y, \phi)\right\} \quad ; \tag{2.1}$$

donde, $f(\cdot)$ representa la función de probabilidad en el caso de que Y fuese una variable discreta, o una función de densidad en el caso de que Y fuese una variable de tipo continua, θ es el parámetro canónico de la función y ϕ es un parámetro de escala. Se definen como funciones específicas de los elementos de la familia exponencial a las expresiones $\alpha(\phi)$, $b(\theta)$ y $c(y,\phi)$.

En los Modelos Lineales Generalizados (MLG) la media (μ) se expresa de tal manera que:

$$E(Y) = \mu. \tag{2.2}$$

Mientras que la varianza de Y es denotada por una función V, captando la relación entre $E(Y)$ y la $Var(Y)$, en donde:

$$Var(Y) = V(\mu). \tag{2.3}$$

2.2. Componente sistemática

La componente sistemática del modelo corresponde a la combinación lineal de las variables regresoras, las cuales ingresan en forma de efecto fijo al modelo, así captando la variabilidad de Y a través de las covariables $(x_1,\ x_2,\ldots,\ x_p)$, las cuales se denotan por X junto con sus parámetros correspondientes, $\beta = (\beta_0,\ \beta_1,\ldots,\ \beta_p)$. En donde generalizando la expresión, queda de la siguiente manera:

$$\eta_i = X\beta. \tag{2.4}$$

2.3. Función de enlace

La función de enlace del modelo corresponde a la función del valor esperado de Y, $E(Y)$, denotada en la componente aleatoria, como una componente lineal de las variables predictoras, la cual cumple la función de enlazar la componente aleatoria con la sistemática.

Se obtiene la siguiente expresión de la función enlace:

$$\mu = \eta \ ;$$

en donde,

$$\eta = g(\mu_i) \ , \tag{2.5}$$

donde $g(\cdot\,)$ corresponde a la función enlace.

Existen diferentes tipos de funciones enlace, las cuales se pueden observar en el siguiente cuadro:

Distribución	Nombre	Función Enlace
Poisson	Logarítmico	$\log \mu_i = \eta_i$
Normal	Identidad	$\mu_i = \eta_i$
Binomial	Logit	$\log(\frac{\mu_i}{n-\mu_i}) = \eta_i$
Multinomial	Logit	$\log(\frac{\mu_i}{n-\mu_i}) = \eta_i$
Exponencial	Inversa	$\mu^{-1} = \eta_i$
Gamma	Inversa	$\mu^{-1} = \eta_i$

Estimación

Habiendo seleccionado un modelo en particular, se requiere estimar los parámetros y evaluar la precisión de las estimaciones. En el caso de los MLG (Modelos lineales generalizados), la estimación se lleva a cabo definiendo una medida de bondad de ajuste entre los datos observados y los valores ajustados por el modelo. Las estimaciones de los parámetros son aquellos valores que minimizan el criterio de bondad de ajuste. Si $f(y;\theta)$ es la función de densidad o la distribución de probabilidad para la variable Y, se obtiene el parámetro θ, entonces la log-verosimilitud se expresa como una función del parámetro de valor medio, $\mu = E(Y)$, es:

$$l(\mu; y) = log f(y;\theta).$$

La log-verosimilitud basada en un conjunto de observaciones independientes $y_1,...,y_n$, es solo la suma de las contribuciones individuales, de modo que:

$$l(\mu; y) = \sum_i log f_i(y_i;\theta_i) \qquad con, \;\; \mu = (\mu_1,...,\mu_n).$$

Nótese que la función de densidad $f(y;\theta)$ se considera como una función de Y para θ fija, mientras que la log-verosimilitud se considera principalmente como una función de θ para los datos particulares Y observados.
Hay ventajas en usar como criterio de bondad de ajuste, la log-verosimilitud $l(\mu; y)$, sino una función lineal particular:

$$D^*(y;\mu) = 2l(y;y) - 2l(\mu;y) \;\; ,$$

la cual se nombra como desviación escalada. Téngase en cuenta que, para los modelos pertenecientes a la familia exponencial considerados aquí, $l(y;y)$ la máxima verosimilitud que se puede obtener para un ajuste exacto en el que los valores ajustados son iguales a los datos observados. Porque $l(y;y)$ no depende de los parámetros maximizando $l(\mu;y)$ es equivalente a minimizar $D^*(y;\mu)$ con respecto a μ, sujeto a las restricciones impuestas por el modelo.

Para los modelos de regresión lineal de teoría normal con varianza σ^2 conocida, se tiene para una sola observación.

$$f(y;\mu) = \frac{1}{\sqrt{(2\pi\sigma^2)}}\, exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \;\; ,$$

así que, la log-verosimilitud es:

$$l(\mu; y) = -\frac{1}{2}\, log(2\pi\sigma^2) - \frac{(y-\mu)^2}{(2\sigma^2)}. \tag{2.6}$$

El ajuste de $\mu = y$ proporciona la máxima log-verosimilitud, es decir,

$$l(y; y) = -\frac{1}{2}\, log(2\pi\sigma^2) \ \ ,$$

de modo que la función de desviación escalada es:

$$D^*(y; \mu) = 2l(y; y) - l(\mu; y) = (y - \mu)^2/\sigma^2 \ \ , \tag{2.7}$$

por lo tanto, del factor conocido sigma, la desviación en este caso es idéntica a la suma de residuos al cuadrado y la desviación mínima es sinónimo de los últimos cuadrados.

Predicción

Para que la predicción sea útil, las cantidades pronosticadas deben ir acompañadas de medidas de precisión. Estas generalmente son calculadas suponiendo que la estructuración que produjeron los datos permanece constante, y que el modelo utilizado en el análisis sea sustancialmente correcto. Para una cuenta de predicción como una idea de unificación que conecta el análisis de covarianza y varios tipos de estandarización.

2.4. Bondad de ajuste

El proceso de ajustar un modelo a los datos se puede considerar como una forma de reemplazar un conjunto de valores de datos y por un conjunto de valores ajustados μ derivados de un modelo que implica, generalmente, un número relativamente pequeño de parámetros. En general, μs no serán exactamente igual a los ys, y entonces surge la pregunta de cuán discrepantes son, porque si bien una pequeña discrepancia es tolerable, una gran discrepancia no lo es. Las medidas de discrepancia o bondad de ajuste pueden formarse de varias maneras, pero interesa principalmente el formado a partir del logaritmo de una proporción de verosimilitud, para llamar a la desviación (López, 2004).

Dada n observaciones, se pueden ajustar modelos que contengan hasta n parámetros. El modelo más simple, el modelo nulo, tiene un parámetro, que representa una μ común para todos los ys; el modelo nulo considera toda la variación entre el ys y el componente aleatorio.

En un modelo lineal generalizado, se puede evaluar de diferentes formas la bondad de ajuste, entre las cuales se destacan las siguientes:

Devianza residual

Es la distancia entre el logaritmo de la función de verosimilitud del modelo saturado (con n parámetros). Si la diferencia existente es pequeña, indica que un número menor de parámetros es igual de bueno que el ajuste obtenido por el modelo saturado. La devianza residual está definida de la siguiente manera:

$$D(y, \hat{\mu}) = 2\{l(y; y) - l(\hat{\mu}; y)\}. \tag{2.8}$$

Para examinar la adecuación de un modelo lineal generalizado, el valor de la diferencia se compara con el percentil de la distribución de probabilidad. Si el modelo es correcto, el estadístico se distribuye asintóticamente según una χ^2_{n-p}, con $n - p$ grados de libertad (McCullagh y Nelder, 1989).

$$D(y, \hat{\mu}) \sim \chi^2_{n-p} \quad , cuando \;\; \mu \to \infty. \tag{2.9}$$

Estadístico Chi-cuadrado de Pearson

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \;\; ; \tag{2.10}$$

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

Coeficiente de determinación

El coeficiente de determinación (R^2) se define como la reducción proporcional en la incertidumbre debido a la inclusión de las covariables, que explica la proporción de variación respuesta al ser explicada por las covariables. Este viene dado por:

$$R^2 = 1 - \frac{D(y; \hat{\mu})}{D(y; \hat{\mu}_0)} \;\; ; \tag{2.11}$$

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 utiliza como referencia, se verifica que $0 \leq R^2 \leq 1$.

Los criterios que se mencionaron se basan en la estimación máxima verosimilitud, los cuales permiten obtener un mejor modelo, el cual se logra ajustar de mejor forma a los datos analizados, sin embargo, estos no sancionan su complejidad en el aumento de los parámetros en el modelo, pero esto sí sucede cuando se utilizan medidas de contraste, en donde, los criterios se utilizan para seleccionar la estructura de covariancia y media del modelo, son funciones del logaritmo de la verosimilitud y un término de penalidad basado en el número de parámetros del modelo. Estos criterios son, el criterio de información Akaike (AIC) (Akaike, 1973) y el criterio de información Bayesiano (BIC) (Schwarz, 1978). Ambos criterios se diferencian por el valor de la penalización. Se prefieren modelos con valores pequeños de ambos criterios.

Citerio de información Akaike (AIC)

El criterio de Akaike es una medida de la bondad de ajuste de un modelo estadístico. Se puede decir que describe la relación entre el sesgo y varianza en la construcción del modelo, o hablando de manera general acerca de la exactitud y complejidad del modelo. Este criterio fue planteado por Akaike en 1973 y proporciona un medio para la comparación entre los modelos para una posterior selección del modelo, en donde, el modelo que tiene el mínimo AIC es el mejor (Garcia, Castellana, Rapelli, Koegel y Catalano, 2014).

Finalmente el criterio de información Akaike está dado por:

$$AIC = 2k - 2ln(L) \ ; \tag{2.12}$$

donde, k es el número de parámetros del modelo y $ln(L)$ es la función de log-verosimilitud para el modelo. Luego, de un conjunto de modelos candidatos para los datos analizados, como se mencionó anteriormente, se escoge el modelo que posea el menor AIC. Este criterio posee un componente penalizados a través de una función creciente, el cual penaliza el aumento del número de parámetros estimados, aumentando el valor que se arroja como resultado.

Citerio de información Bayesiano (BIC)

El criterio Bayesiano es una medida de bondad de ajuste para un modelo estadístico, y es a menudo utilizado como un criterio para para la selección de modelos entre un conjunto finito de modelos. Este criterio esta estrechamente relacionado con el criterio de información de Akaike, introduce un término de penalización para el número de parámetros en el modelo (Gómez, Torres, García y Navarro, 2012).

Finalmente, el criterio de información Bayesiano está dado por:

$$BIC = kln(n) - 2ln(L) \ ; \tag{2.13}$$

donde, k es el número de parámetros del modelo y $ln(L)$ es la función de log-verosimilitud para el modelo, como se mencionó anteriormente, se escoge el modelo que posea el menor BIC. Un criterio Bayesiano bajo, implica un número menor de variables explicativas, mejor ajuste, o ambos.

2.5. Análisis de residuos

El residuo de cada dato mide la discrepancia entre el valor observado y el pronosticado por el modelo (López, 2004). Existen cuatro formas apropiadas para trabajar los residuos de los modelos lineales generalizados, los cuales se observan a continuación:

Residuos de Pearson

Estos son una generalización inmediata de los residuos habituales para datos normales, esta dado por:

$$R_L(y,\theta) = \frac{y - E_\theta(y)}{SD_\theta(y)} \ ; \tag{2.14}$$

donde, E_θ es la media y SD_θ pertenece a la desviación estándar de los residuos.

Devianza Residual

La devianza residual es de suma importancia para lograr realizar un análisis residual del modelo, ya que esta entrega una idea de la variabilidad de los datos y como esta es explicada por el modelo. La devianza esta dada por:

$$R_D(y,\theta) = sgm(\hat{\theta} - \theta)2[l(\hat{\theta}, y) - l(\theta, y)]^{1/2} \ ; \tag{2.15}$$

en donde, $\hat{\theta}$ es el estimador máximo verosímil de θ.

Según Pierce y Shafer en 1986, citado en Osses (2013), sugieren trabajar con la devianza residual, cuando se trabaje con distribuciones pertenecientes a la familia exponencial, esto se debe a que el resultado es muy similar al que genera una excelente transformación a la variable para lograr corregir la normalidad de esta.

Residuo de Anscombe

Anscombe en 1972, definió unos residuos basados en una función $A(y)$ en lugar de y, de manera que la distribución de $A(y)$ sea tan Normal como sea posible. El residuo de Anscombe toma valores cercanos a los de la desviación estandarizada, este residuo realiza una mejor interpretación que la devianza residual.

Los residuos de Anscombe utilizan la función de varianza del modelo, definiéndose de la siguiente manera:

$$r_i^A = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}} \ ; \tag{2.16}$$

en donde,

$$A(\cdot\,) = \int \frac{d\mu_i}{V^{1/3}(\mu_i)}. \tag{2.17}$$

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

Interpretación

Al momento de obtener el modelo adecuado para el análisis de los datos, este proceso finaliza con la interpretación de los resultados del modelo seleccionado, haciendo referencia a cada uno de los criterios descritos en el capítulo.

Cuando se va a interpretar el modelo obtenido, se debe considerar la transformación que se realiza al aplicar una función enlace, lo que provoca en la mayoría de los casos una ecuación del modelo expresada en términos multiplicativos, es por esto que, se debe tener cuidado en la interpretación del modelo que resulte, debido a que dicha interpretación no es directa.

En el siguiente capítulo se darán a conocer los modelos lineales generalizados utilizados para el análisis de los datos de la molienda de la compañía minera.

Capítulo 3

Modelos lineales generalizados Gaussiano y Gamma

En el capítulo anterior se profundizó sobre los modelos lineales generalizados, por lo que en el capítulo presente se dará a conocer las respectivas componentes de los modelos a aplicar (Gaussiano y Gamma).

3.1. Modelo lineal generalizado Gaussiano

La distribución Gaussiana es utilizada para variables de tipo continuas, permite modelar fenómenos naturales, sociales, psicológicos, entre otros.

Se dice que una variable Y_i sigue un modelo lineal generalizado Gaussiano, si cumple con lo siguiente:

$$Y \sim N(\mu; \sigma^2), \quad \mu \in \mathbb{R} \ \ y \ \ \sigma > 0 \tag{3.1}$$

con función de densidad dada por:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.2}$$

la media y varianza están dadas por:

$$E(Y) = \mu \tag{3.3}$$

$$V(Y) = \sigma^2 \tag{3.4}$$

Finalmente, el modelo se define como:

$$g(\mu_i) = \eta_i = X_i^T \beta \ ; \tag{3.5}$$

donde,

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

3.1.1. Componente aleatoria

Sea $Y_i, ...Y_n$ una variable aleatoria independiente continua y sea $(X_{i1}, ..., X_{ip})^T$ un vector de covariables explicativas, su componente aleatoria queda definida por:

$$Y \sim N(\mu, \sigma^2) \tag{3.6}$$

3.1.2. Componente sistemática

Dado μ_i y un predictor lineal representado a continuación:

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} \\ &= \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} \\ &= X_i^T \beta \end{aligned} \tag{3.7}$$

3.1.3. Función de enlace

Mediante la correcta elección de la función de enlace, la componente aleatoria y sistemática se combinan en el modelo, expresado de la siguiente forma:

$$g(\mu_i) = \eta_i \tag{3.8}$$

La función de enlace canónico para el modelo lineal generalizado Gaussiano se denomina función de vínculo identidad y está dada por:

$$\mu_i = \eta_i \tag{3.9}$$

Logverosimilitud

La función de verosimilitud se representa de la siguiente forma:

$$l(\mu; \sigma) = \sum_{i=1}^{n} \frac{y_i \mu_i - \frac{\mu^2}{2}}{\sigma^2} + \sum_{i=1}^{n} \frac{y^2}{(2\sigma^2)} - log\sigma - log\sqrt{2\pi} \tag{3.10}$$

3.2. Modelo lineal generalizado Gamma

La distribución Gamma, es utilizada para variables de tipo continuas y positivas. Se dice que una variable Y_i sigue un modelo lineal generalizado Gamma, si cumple con lo siguiente:

$$Y \sim \gamma(\nu;\mu), \quad \nu > 0 \;\; y \;\; \mu > 0 \tag{3.11}$$

con función de densidad dada por:

$$f(y) = \frac{1}{\Gamma(\nu)}(\frac{\nu}{\mu})^{\nu} y^{\nu-1} e^{-\frac{\nu}{\mu}y}, \;\; y > 0 \tag{3.12}$$

la media y varianza están dadas por:

$$E(Y) = \mu \tag{3.13}$$

$$V(Y) = \frac{\mu^2}{\nu} \tag{3.14}$$

donde μ es el parámetro de media y ν es el parámetro de forma de la distribución.

Finalmente, el modelo se define como:

$$g(\mu_i) = \eta_i = X_i^T \beta \;\; ; \tag{3.15}$$

donde,

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

3.2.1. Componente aleatoria

Sea $Y_i, ...Y_n$ una variable aleatoria independiente continua y sea $(X_{i1}, ..., X_{ip})^T$ un vector de covariables explicativas, su componente aleatoria queda definida por:

$$Y \sim \gamma(\nu;\mu), \quad con \;\; \nu y \mu > 0 \tag{3.16}$$

3.2.2. Componente sistemática

Dado μ_i y un predictor lineal representado a continuación:

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} \\ &= \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} \\ &= X_i^T \beta \end{aligned} \tag{3.17}$$

3.2.3. Función de enlace

Mediante la correcta elección de la función de enlace, la componente aleatoria y sistemática se combinan en el modelo, expresado de la siguiente forma:

$$g(\mu_i) = \eta_i \tag{3.18}$$

La función de enlace canónico para el MLG Gamma se denomina función de vínculo recíproca y está dada por:

$$\frac{1}{\mu_i} = \eta_i \tag{3.19}$$

Logverosimilitud

La función de verosimilitud se representa de la siguiente forma:

$$l(\mu, \nu) = \sum_{i=1}^{n} \nu n_i[\bar{y}_i(-\frac{1}{\mu_i}) - log(\mu_i)] + (\nu - 1)\sum_{ij} log(y_{ij}) + \nu n log(\nu) - n log\Gamma(\nu) \tag{3.20}$$

Devianza residual

La función de devianza del modelo Gamma, es fijando el parámetro de dispersión tal como:

$$\begin{aligned} D(\mu) &= 2[l(\hat{\mu}^{FM}, \nu) - l(\mu, \nu)] \\ &= 2\nu \sum_{i=1}^{n} n_i[\frac{\bar{y}_i}{\mu_i} - log\frac{\bar{y}_i}{\mu_i} - 1] \end{aligned} \tag{3.21}$$

donde, $\hat{\mu}^{FM} = \bar{y}_i$.

Si se representa $\hat{\mu}^M$ al estimador máximo verosímil (EMV) de la media en un modelo M, y se considera fija la dispersión, se verifica que bajo M:

$$D(\hat{\mu}^M) \overrightarrow{\alpha} \chi^2_{n-p} \tag{3.22}$$

siendo p el número de parámetros libres de beta del modelo.

En el siguiente capítulo se podrá observar la aplicación y análisis de estos modelos mencionados en el capítulo presente.

Capítulo 4

Aplicación

4.1. Análisis de datos

En este capítulo, se llevará a cabo la aplicación de los modelos lineales generalizados Gaussiano y Gamma. Para la obtención de este análisis es necesario realizar previamente una presentación de los datos utilizados, y en conjunto un análisis de tipo exploratorio de las variables. El proceso de administración de datos, el análisis exploratorio de estos y la aplicación de los modelos, se realizó utilizando el Software Stata SE Versión 15.

Se utilizó el conjunto de datos de la molienda de la Compañía Minera Doña Inés de Collahuasi, en donde las mediciones son diarias a partir desde Enero del 2015 a Mayo del 2018. Luego, como una primera etapa de administración de datos, se procedió a realizar una limpieza del conjunto de datos, eliminando aquellas variables que no tenían variabilidad y otras variables que poseían un gran número de observaciones faltantes. De esta forma se logró analizar aquellas variables de interés.

Descripción de las variables

A continuación se observa la descripción de la totalidad de variables que se obtuvieron:

- **TON_ALIM**: Tonelaje de alimentación.
- **VELOCIDAD_S**: Velocidad SAG 1011.
- **POTENCIA_S**: Potencia SAG 1011.
- **TORQUE_S**: Torque SAG 1011.
- **SOLIDO_S**: Sólido SAG 1011.
- **GRANULOMETRIA_S**: Granulometría SAG 1011.
- **PEBBLE_S**: Mineral no procesado SAG 1011.
- **AGUA_ALIM_S**: Agua de alimentación SAG 1011.
- **AGUA_TRATM_S**: Agua de tratamiento SAG 1011.
- **NIVEL_DE_CUBA_D**: Nivel de cuba SAG 1011.
- **CARGUIO_BOLAS_S**: Carguío de bolas SAG 1011.
- **VELOCIDAD_12**: Velocidad MB 1012.
- **POTENCIA_12**: Potencia MB 1012.
- **SOLALIMBAT_12**: Sólido de alimentación de batería MB 1012.
- **FLUJOSBAT_12**: Flujos de batería MB 1012.
- **PRESIONBAT_12**: Presión de batería MB 1012.
- **CARGUIO_BOLAS_12**: Carguío de bolas MB 1012.
- **VELOCIDAD_13**: Velocidad MB 1013.
- **POTENCIA_13**: Potencia MB 1013.
- **AGUA_CUBA_13**: Agua de la cuba MB 1013.
- **SOLALIMBAT_13**: Sólido de alimetación de batería MB 1013.
- **FLUJOSBAT_13**: Flujos de batería MB 1013.
- **PRESIONBAT_13**: Presión de batería MB 1013.
- **CARGUIO_BOLAS_13**: Carguío de bolas MB 1013.

Se debe destacar que el propósito de este análisis es determinar el modelo adecuado para estimar la variable de interés tonelaje de alimentación de mineral, a través de las variables entregadas por CMDIC.

4.1.1. Análisis exploratorio

A continuación, se observa un esquema de la tabla de datos.

FECHA	TON_ALIM	VELOCIDAD_S	POTENCIA_S	$\cdots$
01/Enero/2015	4118,32	9,16	20693,56	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Junio/2015	4824,72	9,76	21839,98	$\cdots$
02/Junio/2015	4839,51	9,75	22357,89	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Diciembre/2015	5373,17	9,00	20841,64	$\cdots$
02/Diciembre/2015	5439,06	9,38	21433,34	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Enero/2016	5029,10	9,12	20571,54	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Junio/2016	5160,64	9,61	22307,43	$\cdots$
02/Junio/2016	4178,19	8,95	20573,66	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Diciembre/2016	5714,37	8,00	20789,94	$\cdots$
02/Diciembre/2016	5768,17	7,52	18919,93	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Enero/2017	4570,67	9,37	22738,31	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Junio/2017	4880,43	8,37	15232,28	$\cdots$
02/Junio/2017	4871,69	9,45	16373,65	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Diciembre/2017	5444,81	9,03	19433,37	$\cdots$
02/Diciembre/2017	4483,35	8,35	17977,76	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
01/Enero/2018	4757,59	8,35	16836,56	$\cdots$
02/Enero/2018	4894,73	8,87	18313,20	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
31/Mayo/2018	2531,97	7,02	13873,07	$\cdots$

Tabla 4.1: Esquema del conjunto de datos.

En el Tabla 4.2 se puede visualizar un análisis descriptivo de la variable respuesta analizada, en donde, se observa la cantidad de observaciones, con una media de 4.428 toneladas diarias de mineral que ingresan al molino SAG 1011, una desviación estándar de 793, un mínimo de toneladas diarias de 982 y como máximo 7.084 toneladas diarias de mineral. El tonelaje de alimentación hace referencia a las toneladas de mineral que ingresa al Molino SAG 1011, medidas en toneladas/día.

Variable	Observaciones	Media	D.E	Mínimo	Máximo
Tonelaje de alimentación	1.247	4.428,23	793,31	981,60	7.084,24

Tabla 4.2: Estadística descriptiva del Tonelaje de alimentación.

A continuación, en la Figura 4.1 se puede observar un histograma de la variable respuesta, en donde se obtiene una visualización sobre el comportamiento de el Tonelaje de alimentación.

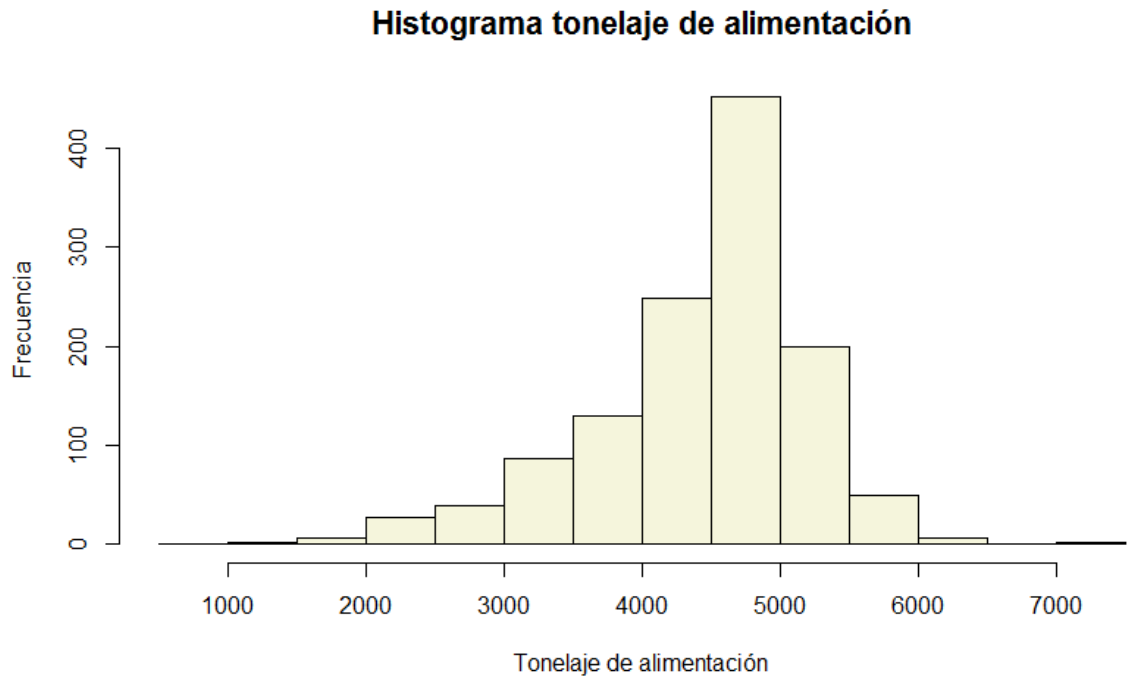

Figura 4.1: Histograma de la variable respuesta.

En la Tabla 4.3 se puede visualizar un análisis descriptivo de las variables en estudio, en donde, se observa la cantidad de observaciones, la media, la desviación estándar, el mínimo y máximo de cada variable.

Variable	Observaciones	Media	D.E	Mínimo	Máximo
Velocidad SAG 1011	1.247	8,72	0,87	2,99	9,79
Potencia SAG 1011	1.247	19.134,20	2.772,40	5.704,09	23.715,09
Peso SAG 1011	1.246	876,76	75,75	513,31	1.086,57
Torque SAG 1011	1.246	20.948,12	1.856,71	12,84	26.202,80
Sólido SAG 1011	1.224	75,40	3,38	41,38	82,38
Granulometría SAG 1011	1.246	64,22	11,99	25,81	99,86
Mineral no procesado	1.240	452,66	265,589	20,06	1.980,31
Agua de alimentación SAG 1011	1.246	1.407,19	320,15	247,96	2.000
Agua de tratamiento SAG 1011	1.246	296,84	119,91	22,19	517,61
Nivel de cuba SAG 1011	1.246	69,94	6,81	30,99	100
Carguío de bolas SAG 1011	1.247	2.529,71	911,76	250	8.874,70
Velocidad MB 1012	1.245	11,69	0,32	8,00	12,01
Potencia MB 1012	1.246	14.747,19	990,18	2.829,58	16.420,27
Solido batería MB 1012	1.247	68,38	13,20	23,79	78,41
Flujos batería MB 1012	1.246	4.056,34	1.213,71	4,21	8.411,27
Presión batería MB 1012	1.246	9,69	0,70	6,59	12,72
Carguío de bolas MB 1012	1.247	10.144,64	4.065,65	227	34.572
Velocidad MB 1013	1.187	11,60	0,91	2,96	12,09
Potencia MB 1013	1.186	14.725,97	1.487,96	2.583,08	16.321,35
Agua cuba MB 1013	1.246	1.392,77	713,38	55,73	4.000
Solido batería MB 1013	1.235	63,64	20,28	1,57	78,40
Flujos batería MB 1013	1.246	4.787,85	1.210,02	372,74	8.404,85
Presión batería MB 1013	1.246	9,79	0,71	0,14	12,62
Carguío de bolas MB 1013	1.247	9.462,97	3.449,81	540	31.762

Tabla 4.3: Estadística descriptiva del conjunto de datos.

En la Figura 4.2 se observa una matriz de correlación de un subconjunto de variables en estudio.

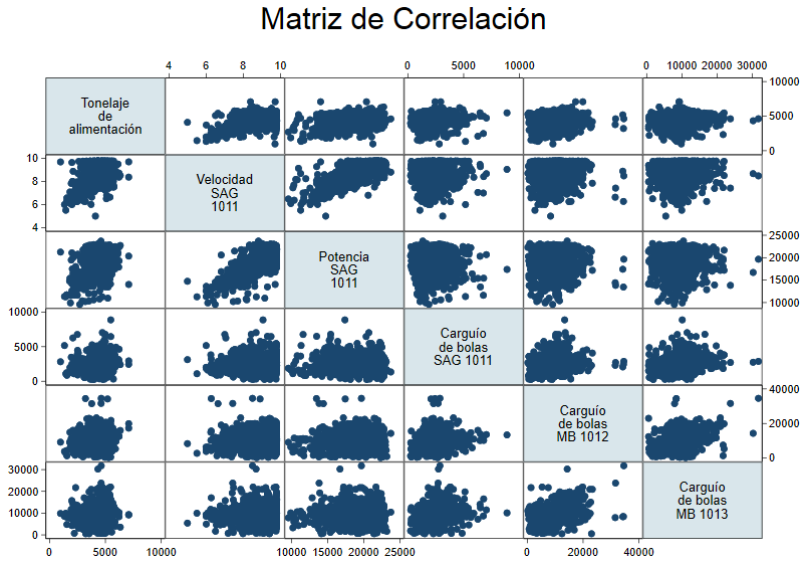

Figura 4.2: Matriz de correlación.

Se procedió a eliminar aquellas variables que no se encontraban correlacionadas con la variable respuesta, posteriormente se analizó el coeficiente de correlación entre el tonelaje de alimentación y las variables independientes a analizar, esto se puede visualizar a continuación.

Variables	Correlación
Tonelaje de alimentación - Velocidad SAG 1011	0,43
Tonelaje de alimentación - Potencia SAG 1011	0,37
Tonelaje de alimentación - Sólido SAG 1011	0,18
Tonelaje de alimentación - Granulometría SAG 1011	0,12
Tonelaje de alimentación - Pebble SAG 1011	0,41
Tonelaje de alimentación - Agua de alimentación SAG 1011	0,64
Tonelaje de alimentación - Carguío de bolas SAG 1011	0,29
Tonelaje de alimentación - Presión batería MB 1012	0,12
Tonelaje de alimentación - Carguío de bolas MB 1012	0,16
Tonelaje de alimentación - Agua cuba MB 1013	0,18
Tonelaje de alimentación - Sólido batería MB 1013	0,15
Tonejale de alimentación - Carguío de bolas MB 1013	0,13

Tabla 4.4: Correlación entre Tonelaje de alimentación y variables independientes.

4.1.2. Aplicación del modelo lineal generalizado Gaussiano

Se obtuvieron los parámetros de la distribución Gaussiana, en donde los resultados obtenidos, fueron los siguientes:

	Media(μ)	Desviación estándar(σ)
Estimación	4.428,233	793,051

Tabla 4.5: Estimación de parámetros distribución Gaussiana.

Una vez obtenidos los parámetros, se procedió a realizar una curva de ajuste de datos Gaussianos, para poder observar el comportamiento de la variable respuesta, en donde se obtuvo lo siguiente:

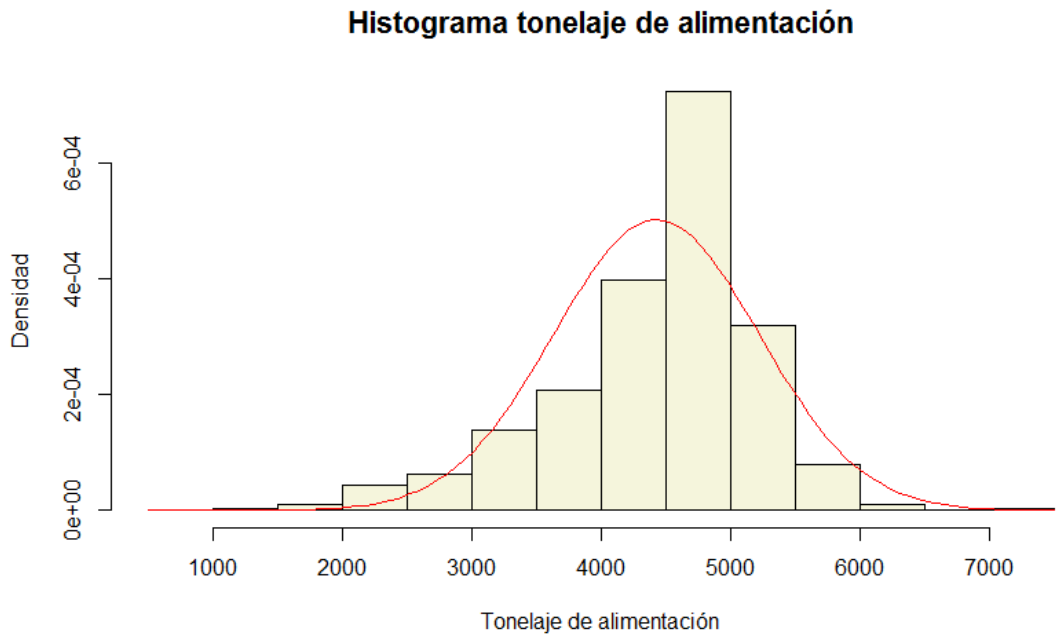

Figura 4.3: Curva de ajuste a la variable respuesta con datos Gaussianos.

En la Figura 4.3 se puede observar la curva de ajuste a el tonelaje de alimentación con datos Gaussianos.

Un vez realizado la correcta manipulación de los datos, eliminando aquellas variables que no se consideraron para este análisis, se procedió a realizar la aplicación del modelo lineal generalizado Gaussiano. Entonces sea Y el tonelaje de alimentación de mineral:

$$Y \sim N(\mu; \sigma^2)$$

Al momento de aplicar el modelo se consideraron aquellas variables correlacionadas con la variable respuesta, posterior a esto se procedió a ir eliminando aquellas variables que no eran significativas en el modelo, obteniendo el siguiente resultado de la estimación:

Coeficiente	Estimación	Error estándar	Z	P > |z|
Velocidad SAG 1011(X_1)	121,568	16,995	7,15	<0,001
Granulometría SAG 1011(X_2)	4,675	1,036	4,51	<0,001
Torque SAG 1011(X_3)	0,029	0,007	4,02	<0,001
Pebble SAG 1011(X_4)	0,433	0,052	8,40	<0,001
Sólido SAG 1011(X_5)	125,713	4,116	30,54	<0,001
Carguío de bolas SAG 1011(X_6)	0,052	0,013	3,93	<0,001
Presión batería MB 1012(X_7)	46,084	18,289	2,52	0,012
Agua cuba MB 1013(X_8)	0,078	0,017	4,62	<0,001
Agua alimentación SAG 1011(X_9)	1,873	0,0519	36,09	<0,001
Intercepto	10.558,82	401,865	26,27	<0,001

Tabla 4.6: Estimación de los coeficientes del MLG Gaussiano con función de enlace identidad.

De la Tabla 4.6 se puede observar que todos los coeficientes son positivos, por lo que estas covariables aportan de forma ascendente a la variable respuesta Tonelaje de alimentación. También se puede notar que todas las variables analizadas son altamente significativas en el modelo.

A continuación se puede observar la estimación de los parámetros con su respectivo intervalo de confianza (IC) del 95 %:

Variables	Estimación	Intervalo del 95 %
Velocidad SAG 1011	121,568	88,257−154,878
Granulometría SAG 1011	4,675	2,644−6,704
Torque SAG 1011	0,029	0,015−0,044
Pebble SAG 1011	0,433	0,332−0,533
Sólido SAG 1011	125,713	117,645−133,781
Carguío de bolas SAG 1011	0,052	0,025−0,077
Presión batería MB 1012	46,084	10,239−81,929
Agua cuba MB 1013	0,078	0,045−0,111
Agua alimentación SAG 1011	1,873	1,771−1,975
Intercepto	10.558,82	9,771,178−11.346,46

Tabla 4.7: Coeficientes e IC del 95 % MLG Gaussiano.

Finalmente, el modelo queda dado por:

$\widehat{Ton_Alim} = 10{,}558,82 + 121,568(X_1) + 4,675(X_2) + 0,029(X_3) + 0,433(X_4) + 125,713(X_5)$
$+ 0,052(X_6) + 46,084(X_7) + 0,078(X_8) + 1,873(X_9)$

Posterior a la estimación de el modelo, se procedió a calcular los estadísticos de bondad de ajuste del modelo, en donde se obtuvo:

$$AIC = 19{,}190,18$$

$$BIC = 19{,}246,59$$

$$Devianza = 196{,}638,1$$

A continuación, se realiza el análisis de residuos del modelo lineal generalizado Gaussiano.

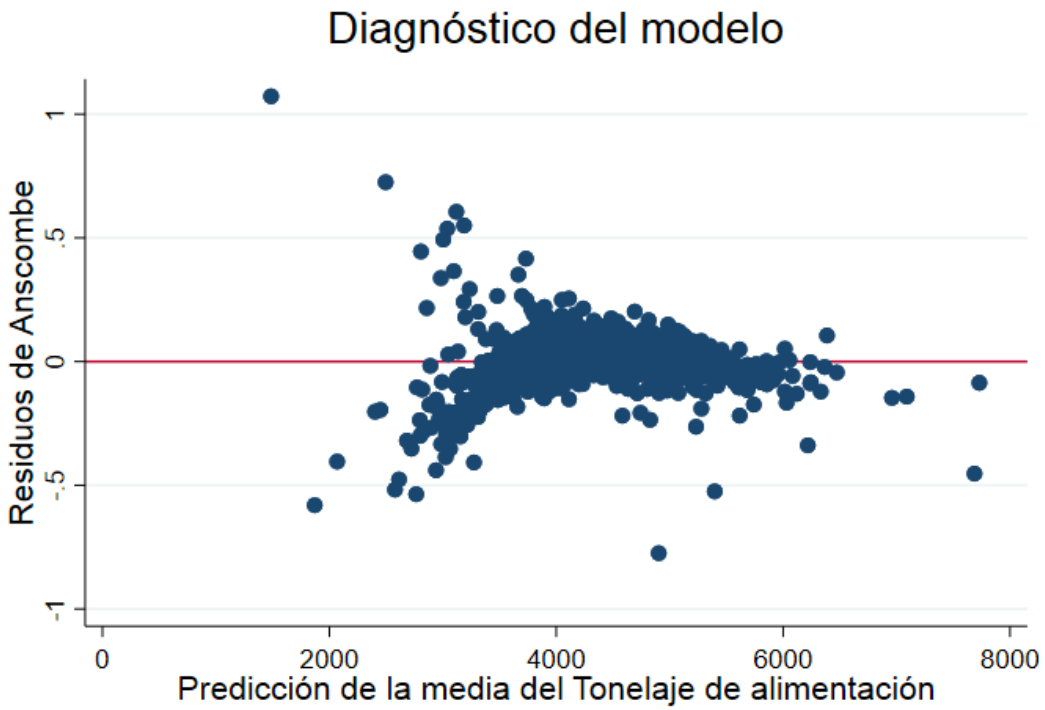

Figura 4.4: Residuos de Anscombe v/s predicción de la media para el Tonelaje de alimentación.

De la Figura 4.4, se observa el gráfico de los residuos de Anscombe frente a los pronósticos de la media del Tonelaje de alimentación del mineral, los residuos tienen una tendencia de acercarse en torno a la media, también se observan muchos datos atípicos y una varianza no constante.

4.1.3. Aplicación del modelo lineal generalizado Gamma

Se estimaron los parámetros de la distribución Gamma, mediante estimaciones de máxima verosimilitud (EMV), en donde los parámetros obtenidos, fueron los siguientes:

	Parámetro de forma(α)	Parámetro de escala(β)	Media(μ)
EMV	25,56	173,26	4.428,23

Tabla 4.8: Estimación de parámetros distribución Gamma.

Una vez obtenidos los parámetros, se procedió a realizar una curva de ajuste de datos Gamma, para poder observar el comportamiento de la variable respuesta, en donde se obtuvo lo siguiente:

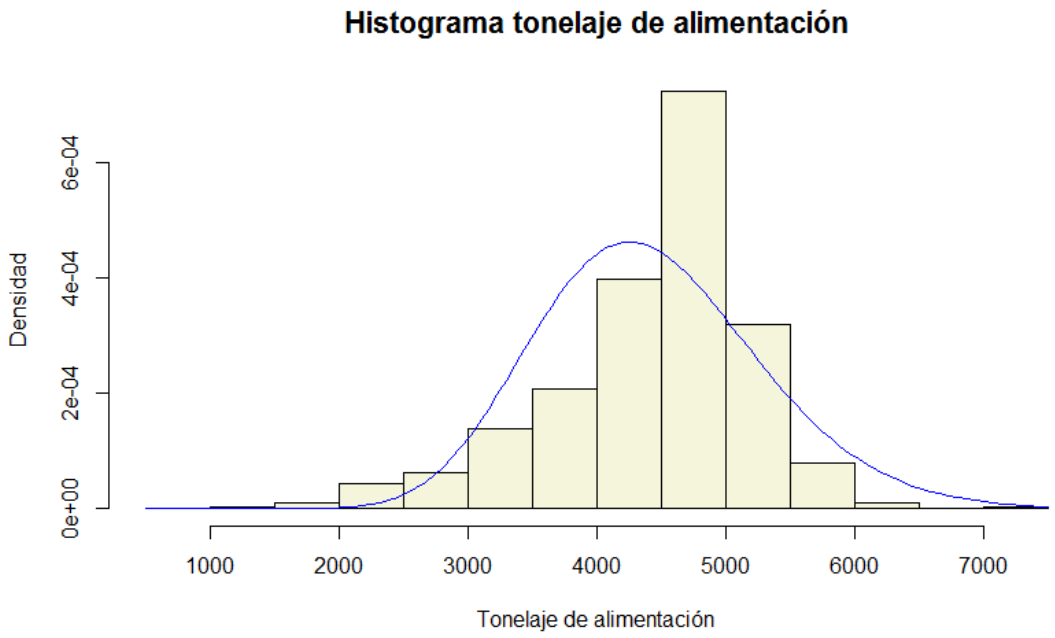

Figura 4.5: Curva de ajuste a la variable respuesta con datos Gamma.

En la Figura 4.5 se puede observar la curva de ajuste a el tonelaje de alimentación con datos Gamma.

Un vez realizado la correcta manipulación de los datos, eliminando aquellas variables que no sirven para este análisis, se procede a la aplicación del modelo lineal generalizado Gamma. Entonces sea Y el tonelaje de alimentación de mineral:

$$Y \sim \gamma(\nu; \mu)$$

Al momento de aplicar el modelo se consideraron aquellas variables correlacionadas con la variable respuesta, posterior a esto se procedió a ir eliminando aquellas variables que no eran significativas en el modelo, obteniendo el siguiente resultado de la estimación:

Coeficiente	Estimación	Error estándar	Z	P > |z|
Velocidad SAG 1011(X_1)	0,042	0,005	8,08	<0,001
Granulometría SAG 1011(X_2)	0,001	0,000	3,95	<0,001
Pebble SAG 1011(X_3)	0,000	0,000	6,04	<0,001
Sólido SAG 1011(X_4)	0,029	0,001	25,86	<0,001
Carguío de bolas SAG 1011(X_5)	0,002	0,000	4,57	<0,001
Presión batería MB 1012(X_6)	0,011	0,006	1,94	0,042
Agua cuba MB 1013(X_7)	0,004	0,001	3,24	0,001
Agua alimentación SAG 1011(X_8)	0,007	0,003	15,21	<0,001
Intercepto	4,943	0,109	45,33	<0,001

Tabla 4.9: Estimación de los coeficientes del MLG Gamma con función de enlace recíproca.

De la Tabla 4.9 se puede observar todos los coeficientes resultan ser positivos, por lo que al igual que el MLG Gaussiano, las covariables aportan de forma ascendente a la variable respuesta tonelaje de alimentación, también se puede notar que la gran mayoría de las variables analizadas son altamente significativas en el modelo.

A continuación, se puede observar la estimación de los parámetros con su respectivo intervalo de confianza (IC) del 95 %:

Variables	Estimación	Intervalo del 95 %
Velocidad SAG 1011	0,042	0,032−0,052
Granulometría SAG 1011	0,001	0,001−0,002
Pebble SAG 1011	0,000	0,000−0,000
Sólido SAG 1011	0,029	0,026−0,031
Carguío de bolas SAG 1011	0,002	0,000−0,000
Presión batería MB 1012	0,011	0,000−0,000
Agua cuba MB 1013	0,004	0,000−0,000
Agua alimentación SAG 1011	0,007	0,000−0,000
Intercepto	4,943	4,729−5,157

Tabla 4.10: Coeficientes e IC del 95 % MLG Gamma.

Finalmente el modelo queda dado por:

$\widehat{Ton_Alim} = \{4{,}434 + 0{,}042(X_1) + 0{,}001(X_2) + 0{,}000(X_3) + 0{,}029(X_4) + 0{,}002(X_5) + 0{,}011(X_6) + 0{,}004(X_7) + 0{,}007(X_8)\}^{-1}$

Al estimar el modelo, también se calcularon los estadísticos de bondad de ajuste del modelo, en donde se obtuvo:

$$AIC = 18{,}474,79$$

$$BIC = 18{,}525,84$$

$$Devianza = 14,32$$

A continuación se observa el análisis residual del modelo lineal generalizado Gamma:

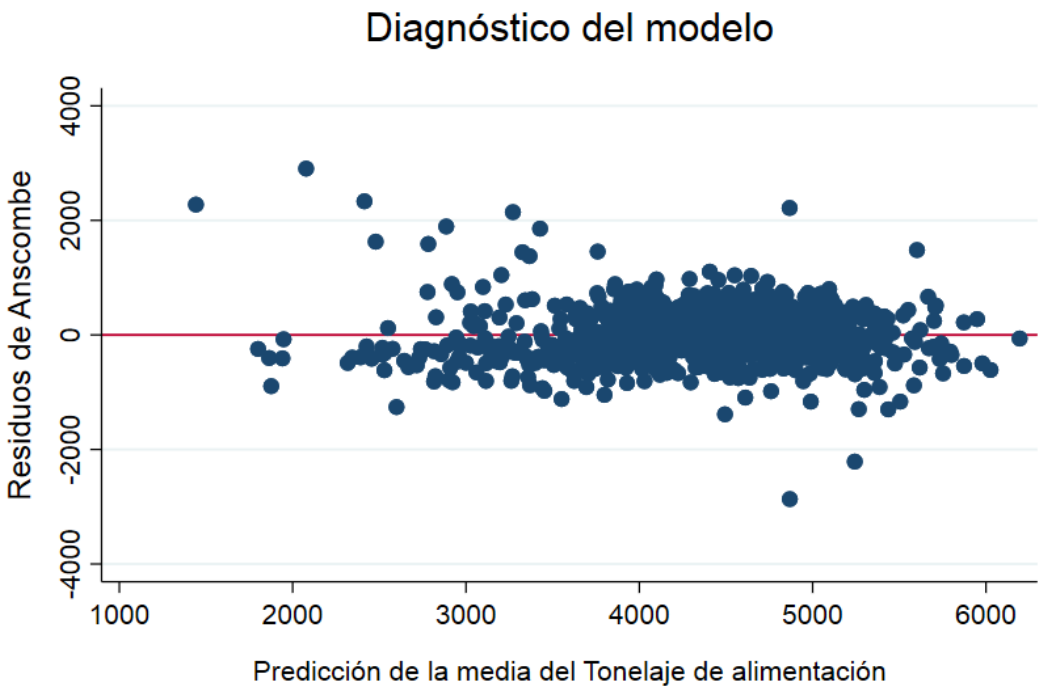

Figura 4.6: Residuos de Anscombe v/s predicción de la media para el Tonelaje de alimentación.

De la Figura 4.6, se observa que en el gráfico de los residuos de Anscombe frente a los pronósticos de la media del tonelaje de alimentación del mineral, los residuos tienen una tendencia en torno a la media con excepción de algunos datos atípicos que podrían ser influyentes, sin embargo estos se ajustan más notoriamente que los residuos del modelo lineal generalizado Gaussiano y es evidente la baja en la variación de los residuos.

4.2. Comparación de modelos

A continuación se observan los distintos estadísticos de bondad de ajuste obtenidos para los modelos analizados:

	MLG Gaussiano	MLG Gamma
AIC	19.190,18	**18.474,79**
BIC	19.246,59	**18.525,84**
Devianza	196.638,1	14,32

Tabla 4.11: Comparación de medidas de bondad de ajuste.

Como se puede observar, en cuanto al criterio de información Akaike y bayesiano entre los dos modelos, existen diferencias entre estos, sin embargo el modelo lineal generarizado Gamma resulta obtener coeficientes más bajos con respecto al MLG Gaussiano. En cuanto a la devianza residual, se puede notar que la diferencia es notoria, en donde, el MLG Gamma resulta obtener la menor devianza en comparación al modelo lineal generalizado Gaussiano.

Capítulo 5

Conclusiones

A partir de la revisión de la literatura en este presente trabajo de titulación, se logra concluir que los modelos lineales generalizados son aptos para poder estimar este tipo de datos, en donde, la variable no sigue una distribución normal, pero sí es perteneciente a la familia exponencial.

De los dos modelos estudiados, considerando los datos analizados, el mejor modelo que logra ajustarse a los datos, es el modelo lineal generalizado Gamma, debido a que este capta mayor variabilidad de la variable respuesta.

Se evaluó la presencia de alguna interacción relevante de aquellas variables contenidas en los modelos propuestos en este trabajo de titulación, pero las interacciones evaluadas no resultaron significativas.

Con el fin de optimizar el proceso de molienda en la Compañía Minera Doña Inés de Collahuasi, se concluye que del total de variables consideradas para este análisis, sólo 8 de estas permitieron obtener una estimación del tonelaje de alimentación de mineral en molinos SAG 1011 y Bolas 1012 y 1013.

Recomendaciones

Al momento de realizar un análisis a un conjunto de datos, es necesario conocer sus características e investigar tipos de modelos en los cuales estas apliquen.

Al aplicar algún tipo de modelo es de gran importancia conocer la naturaleza de la variable respuesta, su recorrido y sus posibles supuestos, para así asegurarse que el modelo elegido es el correcto.

Se debe considerar que para los modelos lineales generalizados se debe conocer la distribución de la variable respuesta, para así elegir el correcto modelo a aplicar.

Se debe tener en cuenta para algunos modelos lineales generalizados la función de enlace utilizada, al momento de realizar la correcta interpretación de los coeficientes estimados del modelo , para no caer en interpretaciones erróneas.

Referencias

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. B. N. Petrov y F. Csaki (Eds.). Second International Symposium on Information Theory (pp. 267–281).

Cayuela, L. (2009). *Modelos lineales generalizados (GLM)*. Universidad de Granada, Granada, España.

Compañía Minera Doña Inés de Collahuasi. (2018). *Collahuasi*.
Obtenido de: http://www.collahuasi.cl/

Gómez, L., Ávila, J., y Cayuela, L. (2013). Métodos de máxima verosimilitud en ecología y su aplicación en modelos de vecindad.*Revista científica de ecología y medio ambiente Ecosistemas*, 12-20. doi:10.7818/ECOS.2013.22-3.03.

Gómez, S., Torres, V., García, Y., y Navarro, J.A. (2012). Procedimientos estadísticos más utilizados en el análisis de medidas repetidas en el tiempo en el sector agropecuario. *Revista Cubana de Ciencias, Agricola, Tomo 46, Número 1.*

Laird, N., y Ware, J. (1982). *Random effects models for longitudinal data*. International Biometric Society, 963-974. doi:10.2307/2529876

López, A. (2004). *Modelos lineales generalizados*. Universidad de Valencia, Burjassot, Valencia.

McCullagh, P., y Nelder, J. (1989). *Generalized linear models*. New York: Chapman y Hall / CRC.

Montgomery, D., Peck, E., y Vining, G. (2006). *Introducción al análisis de regresión lineal (Primera en español ed.)*. México: Compañía Editorial Continental.

Osses, F. (2013). *Aplicación de los residuos en modelos lineales generalizados* (Tesis de pregrado). Universidad de Valparaíso, Valparaíso, Chile.

Schwarz, G. (1978). *Estimating the dimension of a model. Annals of Statistics, 6*, 461–464.

Apéndice

Código Stata

```
//Limpieza de la memoria//
clear all
version 15

//lectura de datos//
import excel "C:\Users\KatalinaBelen\Desktop\Universidad\Documentos Tesis
\Limpieza bd\BD_1.xlsx", sheet("Hoja1") firstrow

//Verificar si el identificador es único//
codebook Fecha

//Etiquetas de variables//
label variable TON_ALIM "Tonelaje de alimentación"
label variable VELOCIDAD_S "Velocidad SAG 1011"
label variable POTENCIA_S "Potencia SAG 1011"
label variable PESO_S "Peso SAG 1011"
label variable TORQUE_S "Torque SAG 1011"
label variable SOLIDO_S "Sólido SAG 1011"
label variable GRANULOMETRIA_S "Granulometría SAG 1011"
label variable PEBBLE_S "Mineral no procesado SAG 1011"
label variable RUIDO_NORTE_S "Ruido norte SAG 1011"
label variable RUIDOSUR_S "Ruido sur SAG 1011"
label variable AGUA_ALIM_S "Agua de alimentación SAG 1011"
label variable AGUA_TRATM_S "Agua de tratamiendo SAG 1011"
label variable NIVEL_DE_CUBA_S "Nivel de cuba SAG 1011"
label variable CARGUIO_BOLAS_S "Carguío de bolas SAG 1011"
label variable VELOCIDAD_12 "Velocidad MB 1012"
label variable POTENCIA_12 "Potencia MB 1012"
label variable AGUACUBA_12 "Agua de la cuba MB 1012"
label variable SOLALIMBAT_12 "Sólido de alimentación de batería MB 1012"
label variable FLUJOSBAT_12 "Flujos de batería MB 1012"
label variable PRESIONBAT_12 "Presión de batería MB 1012"
label variable CARGUIO_BOLAS_12 "Carguío de bolas MB 1012"
```

```
label variable VELOCIDAD_13 "Velocidad MB 1013"
label variable POTENCIA_13 "Potencia MB 1013"
label variable AGUA_CUBA_13 "Agua de la cuba MB 1013"
label variable SOLALIMBAT_13 "Sólido de alimentación de batería MB 1013"
label variable FLUJOSBAT_13 "Flujos de batería MB 1012"
label variable PRESIONBAT_13 "Presión de batería MB 1013"
label variable CARGUIO_BOLAS_13 "Carguío de bolas MB 1013"

//Reemplazar todos los valores "0" por NA//
mvdecode TON_ALIM VELOCIDAD_S POTENCIA_S PESO_S TORQUE_S SOLIDO_S GRANULOMETRIA_S
PEBBLE_S RUIDO_NORTE_S RUIDOSUR_S AGUA_ALIM_S AGUA_TRATM_S NIVEL_DE_CUBA_S
CARGUIO_BOLAS_S VELOCIDAD_12 POTENCIA_12 AGUACUBA_12 SOLALIMBAT_12 FLUJOSBAT1_12
PRESIONBAT_12 CARGUIO_BOLAS_12 VELOCIDAD_13 POTENCIA_13 AGUA_CUBA_13 SOLALIMBAT_13
FLUJOSBAT_13 PRESIONBAT_13 CARGUIO_BOLAS_13, mv(0)

//Eliminar variable Fecha//
drop Fecha

//Eliminar variables con exceso de datos faltantes//
drop RUIDO_NORTE_S
drop RUIDOSUR_S
drop AGUACUBA_12
drop SOLALIMBAT_12

//Resumen de datos a analizar//
sum

//Gráfica variable respuesta
//hist TON_ALIM
//gamma4 TON_ALIM
//qgamma TON_ALIM , param(25.556 173.278)
//gen t = rgamma(25.556, 173.278)
//gen x = (t/173.278)
//ksmirnov x = gammap(25.556, x)

//glm gamma original//
glm TON_ALIM VELOCIDAD_S POTENCIA_S PESO_S TORQUE_S SOLIDO_S GRANULOMETRIA_S
PEBBLE_S AGUA_ALIM_S AGUA_TRATM_S NIVEL_DE_CUBA_S CARGUIO_BOLAS_S VELOCIDAD_12
POTENCIA_12 SOLALIMBAT_12 FLUJOSBAT_12 PRESIONBAT_12 CARGUIO_BOLAS_12 VELOCIDAD_13
POTENCIA_13 AGUA_CUBA_13 SOLALIMBAT_13 FLUJOSBAT_13 PRESIONBAT_13 CARGUIO_BOLAS_13,
family(gamma) link(reciprocal)
\newpage

//Correlación de Spearman //
spearman TON_ALIM VELOCIDAD_S POTENCIA_S PESO_S TORQUE_S SOLIDO_S GRANULOMETRIA_S
```

```
PEBBLE_S AGUA_ALIM_S AGUA_TRATM_S NIVEL_DE_CUBA_S CARGUIO_BOLAS_S VELOCIDAD_12
POTENCIA_12 SOLALIMBAT_12 FLUJOSBAT_12 PRESIONBAT_12 CARGUIO_BOLAS_12 VELOCIDAD_13
POTENCIA_13 AGUA_CUBA_13 SOLALIMBAT_13 FLUJOSBAT_13 PRESIONBAT_13 CARGUIO_BOLAS_13

//Eliminar las variables que no estan correlacionadas con la variable respuesta//
drop NIVEL_DE_CUBA_S
drop VELOCIDAD_12
drop POTENCIA_12
drop SOLALIMBAT_12
drop FLUJOSBAT_12
drop VELOCIDAD_13
drop POTENCIA_13
drop FLUJOSBAT_13
drop PRESIONBAT_13

//Nueva correlación de Spearman sin variables no correlacionadas con la respuesta//
spearman TON_ALIM VELOCIDAD_S POTENCIA_S PESO_S TORQUE_S SOLIDO_S GRANULOMETRIA_S
PEBBLE_S AGUA_ALIM_S CARGUIO_BOLAS_S PRESIONBAT_12 CARGUIO_BOLAS_12 AGUA_CUBA_13
SOLALIMBAT_13 CARGUIO_BOLAS_13

//Matriz de correlación//
graph matrix TON_ALIM VELOCIDAD_S POTENCIA_S CARGUIO_BOLAS_S CARGUIO_BOLAS_12
CARGUIO_BOLAS_13, title(Matriz de Correlación)

//MLG GAMMA//

//glm gamma con variables correlacionadas con variable dependiente//
glm TON_ALIM VELOCIDAD_S POTENCIA_S PESO_S TORQUE_S SOLIDO_S GRANULOMETRIA_S
PEBBLE_S AGUA_ALIM_S CARGUIO_BOLAS_S PRESIONBAT_12 CARGUIO_BOLAS_12 AGUA_CUBA_13
SOLALIMBAT_13 CARGUIO_BOLAS_13, family(gamma) link(reciprocal)

//glm gamma eliminando variables no significativas//
glm TON_ALIM VELOCIDAD_S SOLIDO_S GRANULOMETRIA_S PEBBLE_S AGUA_ALIM_S
CARGUIO_BOLAS_S PRESIONBAT_12 AGUA_CUBA_13, family(gamma) link(reciprocal)

//MLG GAUSSIANO//

//glm gaussiano con variables correlacionadas con la variable dependiente//
glm TON_ALIM VELOCIDAD_S POTENCIA_S PESO_S TORQUE_S SOLIDO_S GRANULOMETRIA_S
PEBBLE_S AGUA_ALIM_S AGUA_TRATM_S CARGUIO_BOLAS_S PRESIONBAT_12 CARGUIO_BOLAS_12
AGUA_CUBA_13 SOLALIMBAT_13 CARGUIO_BOLAS_13, family(gaussian) link(identity)

//glm gaussiano eliminando variables no significativas//
glm TON_ALIM VELOCIDAD_S TORQUE_S SOLIDO_S GRANULOMETRIA_S PEBBLE_S AGUA_ALIM_S
CARGUIO_BOLAS_S PRESIONBAT2_12 AGUA_CUBA_13, family(gaussian) link(identity)
```