



APLICACIÓN DEL MODELO DE REGRESIÓN
LOGÍSTICA Y ANÁLISIS DE DIAGNÓSTICO EN
PACIENTES QUE CONSUMEN BENZODIAZEPINA
EN CENTRO DE SALUD FAMILIAR DE LA
COMUNA DE LLAYLLAY

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

Presentado por:
Bárbara Castillo Olivares

Profesor guía:
Marco Riquelme, PhD

Valparaíso, Chile, Abril 2017

Dedicado a mis padres...

AGRADECIMIENTOS

El primer agradecimiento quiero hacerlo llegar a los principales impulsores de mis sueños, pues sin ustedes queridos padres no hubiese podido llegar hasta donde estoy.

La cúspide de mi primera meta como estudiante la he cristalizado, y es gracias a la formación de la Universidad y a cada uno de los profesores, en especial a mi profesor guía Marco Riquelme quien tuvo la disposición y amabilidad de avanzar conmigo en este camino.

Afortunadme en la mayoría de los casos no estamos solos y mucho menos en este tipo de situaciones donde las cosas no están saliendo como esperamos, en ese momento acudí a Camilo Lillo quien estuvo ahí para responder mis dudas y Juan Ignacio que se dio el tiempo de leer conmigo una y otra vez.

Muchas gracias por el apoyo.

RESUMEN

Este trabajo ha considerado principalmente los riesgos a la dependencia que padecen las personas conforme al consumo irregular de drogas actualmente en Chile.

Para ello han sido ocupados datos proporcionados por el CESFAM (Centro de Salud Familiar) de la comuna de Llayllay, región de Valparaíso, tomando en cuenta a pacientes prescritos con Benzodiazepina desde los años 2013 al 2015 ¹.

El objetivo principal es estudiar la probabilidad de riesgo de dependencia de dichos pacientes a través del Modelo de Regresión Logística. Para ello se hará uso de métodos estadísticos que consisten principalmente en la selección de variables para ajustar un modelo que sea capaz de predecir de forma adecuada el riesgo de padecer dependencia de Benzodiazepina.

Existen por tanto covariables significativas que predicen ciertamente el riesgo de dependencia. También se lleva a cabo Análisis de Diagnóstico como; Distancia de Cook y Leverage, para detectar posibles casos atípicos que puedan afectar en las estimaciones de los parámetros del modelo ajustado.

Los resultados de las pruebas serán sujetos a interpretación a modo de lograr una visión clara acerca de la problemática de dependencia de drogas atenuante actualmente en la comuna de LlayLlay.

¹ver anexo

Índice general

AGRADECIMIENTOS	III
RESUMEN	IV
Lista de figuras	VII
Lista de tablas	1
1. INTRODUCCIÓN	3
1.1. Objetivos	5
1.1.1. Objetivo general	5
1.1.2. Objetivos específicos	5
1.2. Revisión bibliográfica	6
1.3. Hipótesis	8
1.4. Presentación de las variables	8
2. MODELO DE REGRESIÓN LOGÍSTICA	9
2.1. Modelo Lineal	10
2.1.1. Estructura del Modelo Lineal	10
2.1.2. Estructura del Error	10
2.1.3. Restricciones de los Modelos Lineales	11
2.2. Modelos Lineales Generalizados (GLM)	11
2.3. Familia Exponencial	11
2.3.1. Las distribuciones más representativas de la Familia Exponencial	12
2.4. Modelo de Regresión Logística Múltiple	13
2.4.1. Estimación de los parámetros	14

2.4.2. Razón de Chance (odds ratio)	14
2.5. Chi-squared Automatic Interaction Detection	15
2.5.1. Técnicas Potenciales	15
2.5.2. Algoritmo	16
2.6. Stepwise	17
2.6.1. Método Backward:	17
2.6.2. Método Forward:	18
2.6.3. Método Stepwise:	18
2.6.4. Criterio de Información Akaike (AIC)	18
3. DIAGNÓSTICO EN EL MODELO DE REGRESIÓN LOGÍSTICA	19
3.1. Análisis de Residuos	19
3.1.1. Residuos de Pearson	20
3.1.2. Residuos de Pearson Estandarizado	20
3.1.3. Residuos de la Devianza	21
3.2. Medidas de Influencia	21
3.2.1. Leverage	21
3.2.2. Distancia de Cook	22
3.2.3. Gráficos de Diagnósticos	22
4. APLICACIÓN	23
4.1. Descriptiva de los Datos	24
4.2. Resultados Método CHAID	33
4.3. Resultados Método Stepwise	36
4.4. Resultados Modelo de Regresión Logística	36
4.5. Resultados Análisis de Residuos	38
4.6. Resultados Análisis de Diagnóstico	43
4.7. Probabilidad de Ocurrencia	46
Conclusión	51
Referencias Bibliográficas	53
Anexo	55

Índice de figuras

4.1. Gráfico sectorial de la variable de respuesta.	24
4.2. Histograma de la edad de los pacientes.	25
4.3. Box plot de la edad de los pacientes.	26
4.4. Gráfico sectorial de la variable de sexo.	27
4.5. Gráfico sectorial de la variable tipo de enfermedad.	28
4.6. Gráfico sectorial de la variable tipo de fármaco.	29
4.7. Gráfico sectorial de la variable tipo de receta.	30
4.8. Histograma de la variable frecuencia.	31
4.9. Box plot de la variable de frecuencia.	32
4.10. Árbol de clasificación del método CHAID.	34
4.11. Residuos de la Devianza	42
4.12. Distancia de Cook	43
4.13. Leverage v/s Distancia de Cook	45
4.14. Probabilidad de ocurrencia según la variable frecuencia en los pacientes que padecen el tipo de enfermedad 4.	48
4.15. Probabilidad de ocurrencia según la variable frecuencia en los pacientes que padecen otro tipo de enfermedad.	50
4.16. Contentimiento para trabajar con los datos solicitados.	55

Índice de tablas

2.1. Algunas distribuciones de la Familia Exponencial	13
2.2. Análisis según tipo de variables	15
4.1. Distribución de frecuencia de la variable de respuesta.	24
4.2. Estadísticos de resumen de la variable edad.	25
4.3. Distribución de frecuencia del sexo del paciente.	26
4.4. Distribución de frecuencia del tipo de enfermedad del paciente.	28
4.5. Distribución de frecuencia de la variable tipo de fármaco.	29
4.6. Distribución de frecuencia de la variable tipo de receta.	30
4.7. Estadísticos de resumen de la variable frecuencia.	31
4.8. Características del árbol de clasificación.	33
4.9. Selección de las covariables.	36
4.10. Estimación de parámetros del modelo de regresión logística ajustado.	37
4.11. Cantidad de residuos significativos.	39
4.12. Pacientes con los residuos de Pearson más altos.	39
4.13. Características de los pacientes con residuos altos.	39
4.14. Cantidad de residuos de Pearson estandarizados significativos.	40
4.15. Pacientes con los residuos más altos.	40
4.16. Pacientes con los residuos más altos.	40
4.17. Características de los pacientes con residuos de Pearson estandarizados altos.	41
4.18. Residuos significativos de la Devianza.	41
4.19. Cantidad de pacientes con residuos de Devianza estandarizada significativos.	41
4.20. Cantidad de pacientes con distancia significativa.	43
4.21. Características de los pacientes con distancia de Cook más altos	44
4.22. Estimación de parámetros del modelo de regresión logística ajustado final.	45
4.23. Probabilidad de riesgo de dependencia en los pacientes que padecen el tipo de enfermedad 4.	47

4.24. Probabilidad de riesgo de dependencia en los pacientes que padecen otro tipo de enfermedad 49

INTRODUCCIÓN

A lo largo del tiempo la dependencia de drogas se ha suscitado como problema de gran envergadura para la sociedad, afectando tanto a consumidores como a los gobiernos encargados de generar políticas públicas de atención para dichos problemas. Este trabajo pretende estudiar la probabilidad de riesgo de dependencia de Benzodiazepina en los pacientes del CESFAM entre los años 2013 y 2015 pertenecientes a la comuna de Llaylay perteneciente a la provincia de San Felipe de Aconcagua, en la Región de Valparaíso, Chile. Específicamente en este estudio se trabajará con la base de datos facilitada por el CESFAM.

La base de datos consta de muchas covariables en donde a través de distintos métodos se seleccionarán las más significativas o las que mejor describan a la variable de respuesta y así ajustar un Modelo de Regresión Logística adecuado que sea capaz de predecir la probabilidad de ocurrencia al riesgo de dependencia.

Junto con ello la revisión bibliográfica ha de sugerir conceptualizaciones referentes a la comprensión de dependencia de drogas que existe por parte de los centros de estudio de temáticas afines, como también acerca de la creación, difusión y usos de lo que es la Benzodiazepina.

Luego del análisis de datos con las covariables más significativas seleccionadas se ajustará el Modelo de Regresión Logística para aplicar Análisis de Diagnóstico y determinar la productividad del modelo. Con el fin de obtener la probabilidad de riesgo de dependencia al fármaco de los pacientes pertenecientes al CESFAM de Llaylay.

Planteamiento del problema

Durante los últimos años en el mundo han sido utilizados diversa cantidad de fármacos por parte de la población mundial. Dentro de ellas están las Benzodiazepinas (BZD) las cuales son uno de los psicofármacos que durante estos últimos años han producido gran atención debido al aumento en su indicación. Junto con ello se han elevado también los efectos adversos asociados al uso prolongado del mismo, especialmente la presencia de dependencia y la asociación con el deterioro cognitivo.

En un estudio poblacional realizado en Santiago de Chile se halló una prevalencia anual en el uso de psicofármacos de 31,4 % siendo mayor en mujeres que en hombres (38,9 % vs 22,8 %). Un 42,9 % de la población había consumido alguna vez en su vida BZD y un 4,1 % podía considerarse como dependiente a las BZD, un porcentaje tres veces más alto que el promedio mundial. Además, se encontró que un 28,9 % de los psicofármacos consumidos eran automedicados (Busto, 1991). En otro estudio en consultorios del área norte de Santiago, la prevalencia anual del uso de psicofármacos fue de 29,4 %, 19,4 % correspondiendo a BZD. De éstas, un 38,8 % eran consumidas sin prescripción médica (Gonzalez, 1993).

La comuna de Llayllay, escenario en el cual se desarrolla este trabajo cuenta a su haber con un centro de salud familiar (CESFAM) en el cual existe un servicio de atención para la dependencia de drogas. Según datos principales de la oficina de estadística del CESFAM en el último año la atención de pacientes que presenten algún tipo de dependencia a las drogas lícitas o ilícitas presenta mayor frecuencia en el segmento etario de entre 35 a 39 años de edad. Por otro lado el segmento etario que presenta menor frecuencia de dependencia de drogas lícitas o ilícitas se concentra en los segmentos etarios de edades menores.

En este trabajo de titulación, se hará uso de métodos estadísticos para determinar la probabilidad de riesgo de dependencia de BDZ en los pacientes del CESFAM de la comuna de Llayllay entre los años 2013 a 2015.

Resulta imprescindible hacer énfasis en que se hace cada vez más latente el problema que enfrenta la evolución del consumo dependiente de drogas. Dicha evolución no es un fenómeno propio de Llayllay, ya que se repite a lo largo de Chile y el mundo. Esbozar un esfuerzo por intentar objetivar la dependencia de fármacos representa para la contingencia actual un desafío dada la precariedad de los datos disponibles en Chile y Llayllay actualmente para el segmento de drogas lícitas.

1.1 **Objetivos**

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

1.1.1 **Objetivo general**

Determinar a través del Modelo de Regresión Logística la probabilidad de riesgo de dependencia en los pacientes que consumen Benzodiazepinas del CESFAM de L Layllay desde el año 2013 al 2015, considerando las características de los pacientes con respecto a las covariables asociadas.

1.1.2 **Objetivos específicos**

- (i) Identificar los factores que influyen en la dependencia del fármaco.
- (ii) Ajustar un Modelo de Regresión Logística que permite determinar de forma adecuada la probabilidad de riesgo de dependencia.
- (iii) Realizar Análisis de Diagnóstico de Influencia para detectar la presencia de datos influyentes.
- (iv) Estudiar el riesgo de dependencia del paciente con respecto a su tipo de enfermedad.

1.2 Revisión bibliográfica

Diferentes consideraciones existen respecto al uso que tienen los fármacos actualmente en Chile. Existe al mismo tiempo un debate constante dentro de la formación de políticas públicas conforme a la flexibilidad con la cual se genera dependencia de fármacos en las personas en Chile y el mundo. La farmacodependencia está derivada del consumo reiterado de fármacos, en síntesis de cualquier sustancia que sirva para prevenir, curar, aliviar la enfermedad o para reparar sus consecuencias. La farmacodependencia es al mismo tiempo un trastorno conductual derivado de los efectos biológicos de una determinada sustancia, a raíz de ello una persona tiene disminuido el control sobre el consumo de esta. Los psicofármacos, puntualmente los usados para tratar problemas mentales, son los más idóneos a provocar farmacodependencia.

La Organización Mundial de la Salud (OMS) define a la dependencia de drogas como un estado psíquico y algunas veces físico resultante de la interacción entre un organismo vivo y un producto psicoactivo, dicho estado se caracteriza por producir modificaciones de la conducta (psicoactivo) y otras reacciones que incluyen siempre un deseo incontrolable de consumir droga, continua o periódicamente, a fin de experimentar nuevamente sus efectos psíquicos y evitar a veces el malestar de su privación. En la actualidad, el concepto ha sido remplazado por “consumo problemático de drogas”. Cuando un médico receta una sustancia, se puede hablar de uso de drogas, es decir, consumo de drogas sin las consecuencias negativas antes descritas. El Servicio Nacional de Drogas y Alcohol (SENDA) entiende la dependencia de drogas como la situación fisiológica en la que el consumidor ha desarrollado una adaptación o tolerancia al uso de sustancias, cuya suspensión brusca puede generar una enfermedad denominada abstinencia.

En otras palabras la dependencia es comprendida como el conjunto de comportamientos y de reacciones que se refieren al impulso y la necesidad imperiosa de ingerir sustancias en forma continua o regular, teniendo como fin el sacie del malestar producido por sus efectos respecto a la privación. En términos generales la dependencia puede afectar al cuerpo denominándose dependencia física a actividades específicas del sistema nervioso central. Para efectos de este trabajo y sus conclusiones, se hará uso y entendimiento de las conceptualizaciones anteriores referentes a la dependencia de drogas.

En un contexto histórico la primera BZD desarrollada corresponde al clordiazepóxido, el cual tuvo inicialmente el nombre de metamínodiazepóxido. Esta BZD fue descubierta en 1959 por el científico Leo Sternbach (1908-2005), luego sintetizada en el año 1955 y finalmente comercializada a partir de 1957 con el nombre de librium, derivado de las sílabas finales de equilibrium (Ariza, 2012).

El nombre de Benzodiazepinas se debe a la presencia en su estructura química del anillo benzodiazepínico el cual está compuesto por un anillo de benceno unido a uno de diazepina, de 7 miembros heterocíclicos.

Algunas pruebas realizadas con este medicamento, en animales, demostraron que el compuesto era un efectivo hipnótico, ansiolítico y relajante muscular. Años después de haberse desarrollado dicho compuesto comenzó a comercializarse el Diazepam con el nombre de valium, el cual correspondía a una versión simplificada del clordiazepóxido, seguido por otras Benzodiazepinas.

Al igual que a existencia de dependencia en drogas duras, la dependencia en fármacos existe y es un campo poco atendido por la sociedad actualmente. Al enmarcarse en un contexto legal, la desregulación del consumo se presenta como un fenómeno el cual persiste y afecta diferentes segmentos de la población. Una de dichas dependencias a este tipo de drogas lícitas es la circunscrita a la Benzodiazepina. Para lograr dilucidar dicho fenómeno es necesario entender que una condición propia de la BZD es que esta se ajusta generalmente a un marco terapéutico. Frecuentemente el primer contacto con una BZD ocurre en un contexto clínico relacionado con una prescripción médica. Así, los pacientes con trastornos psiquiátricos corren un mayor riesgo de presentar dependencia a BZD en comparación con otros grupos diagnósticos. La teoría más difundida en relación al mecanismo de acción hipnótico y/o sedante de la BZD se centra en el GABA como el más conocido depresor endógeno del sistema nervioso central. Se cree que el uso crónico de BZD resulta en una regulación negativa del receptor GABA-inhibidor y cambios configuracionales del complejo receptor-agonista, disminuyendo la sensibilidad al agonista. Estos cambios explicarían los fenómenos de tolerancia, dependencia y síndrome de abstinencia asociados al consumo prolongado de BZD (Olivera, 2009).

Por otro lado la tolerancia y la dependencia de las BZD es creada con rapidez entre quienes las consumen, pues demuestran síntomas de abstinencia en tan solo 3 semanas de uso continuo. Una vez que se haya establecido una dependencia medicamentosa, el clínico tiende a establecer el consumo diario promedio de BZD por el paciente y comienza con reducciones de aproximadamente 2 mg (Ariza, 2012).

1.3 Hipótesis

Los pacientes que utilizan Benzodiazepinas dentro de su terapia en un periodo de tiempo prolongado, tienen un mayor riesgo a ser dependientes de este fármaco.

1.4 Presentación de las variables

La base de datos proporcionada por el CESFAM de la comuna de Llayllay contiene una serie de covariables con la facultad de caracterizar a los pacientes.

En este caso la variable de respuesta es riesgo de dependencia, en el cual se asume igual a 1 cuando el paciente tiene riesgo y 0 cuando el paciente no tiene riesgo a ser dependiente del fármaco.

Variable de respuesta:

$$\text{Riesgo de dependencia} = \begin{cases} 1; & \text{Con riesgo} \\ 0; & \text{Sin riesgo} \end{cases}$$

Las características de cada uno de los pacientes que serán estudiados para determinar el riesgo de dependencia al fármaco, están dadas a través de las siguientes **Covariables:**

- **género:** hombre o mujer.
- **edad:** entre 15 y 98 años.
- **tipo de fármaco:** Clonazepan de 0,5Mg, Clonazepan de 2Mg y Diazepan de 10Mg.
- **tipo de receta:** Crónica, Salud mental y Morbilidad.
- **frecuencia:** número de veces que el paciente retira el fármaco durante los 3 años.
- **tipo de enfermedad:** Depresión, Trastornos, Alcohol - Drogas, Epilepsia - Esquizofrenia - Problemas familiares, Repetición receta, Demencia - Dislímia - Insomnio, Otros.

MODELO DE REGRESIÓN LOGÍSTICA

El análisis de Regresión Logística es una técnica multivariante con gran importancia debido a sus aplicaciones en diferentes campos del conocimiento humano. El análisis de Regresión Logística Múltiple fue aplicado por primera vez a principio de los años 60 en un estudio referente a enfermedades coronarias (Truett et al. 1967). En el estudio del análisis de Regresión Logística, frecuentemente el conjunto de los datos contiene algunas observaciones atípicas o extremas en relación a los datos. En un Modelo de Regresión Logística se han de examinar los datos que están siendo utilizados, con el objetivo de determinar la existencia de uno o varios puntos que están controlando propiedades importantes del modelo. Para una comprensión clara del funcionamiento del Modelo de Regresión Logística es necesario identificar en que consiste el Modelo Lineal, el Modelo Lineal Generalizado y así también la Familia Exponencial.

El Modelo de Regresión Logística tiene como objetivo determinar la probabilidad de riesgo de dependencia a través de las covariables seleccionadas mediante la técnica de CHAID (*Chi-squared Automatic Interaction Detection*) y de Stepwise. El método de CHAID responde al objetivo específico número uno en vista de identificar las covariables influyentes en la dependencia del fármaco, en tanto el método Stepwise deja las covariables que mejor describen a la variable de respuesta, para este caso resulta en el riesgo de dependencia. Estos métodos ajustan de mejor forma el Modelo de Regresión Logística.

2.1 Modelo Lineal

El término Modelo Lineal se aplica cuando la media, la varianza o las covarianzas pueden ser expresadas mediante funciones lineales de parámetros desconocidos. (Galton, 1886) propuso un Modelo de Regresión Lineal en donde la respuesta Y_i , es variable aleatoria independiente distribuida normalmente con media μ_i y varianza constante σ^2 , es decir.

$$Y_i \sim N(\mu_i, \sigma^2), \quad \text{para } i=1, \dots, n \quad . \quad (2.1)$$

Por otro lado las variables explicativas (x_{ij}) proporcionan un conjunto de predictores lineales dados por la expresión.

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}, \quad \text{para } i=1, \dots, n \quad \text{y} \quad j=1, \dots, p \quad (2.2)$$

donde x_{ij} son las covariables o variables explicativas y los β_i son los coeficientes asociados a cada una de las covariables. El enlace de la ecuación (2.1) y (2.2) da paso a lo siguiente:

$$E(Y_i) = \mu_i = \eta_i. \quad (2.3)$$

2.1.1 Estructura del Modelo Lineal

El Modelo Lineal general se escribe de la siguiente forma: (ver; McCullagh, 1989)

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad (2.4)$$

donde Y_i es la variable de respuesta, $i=1, \dots, n$ que se modela por una función lineal de variables explicativas (x_{ij}) con $j=1, \dots, p$, más un término de error.

2.1.2 Estructura del Error

Los errores ε_i son independientes e idénticamente distribuidos con esperanza cero y varianza σ^2 :

$$E(\varepsilon_i) = 0 \quad \text{y} \quad Var(\varepsilon_i) = \sigma^2. \quad (2.5)$$

Son variables aleatorias independientes y normalmente distribuidas con varianza σ^2 y media constante igual a cero:

$$\varepsilon_i \sim N(0, \sigma^2), \quad \text{para } i=1, \dots, n \quad .$$

2.1.3 Restricciones de los Modelos Lineales

A pesar de la diversa utilidad, existen situaciones donde los Modelos Lineales resultan no ser apropiados, por ejemplo:

- Cuando el rango de Y está restringido.
- Cuando la varianza de Y depende de la media.

El Modelo Lineal Generalizado (GLM) es una generalización del Modelo Lineal, donde la variable de respuesta puede tener una distribución distinta de la distribución normal.

2.2 Modelos Lineales Generalizados (GLM)

El Modelo Lineal Generalizado (ver; Paula, 2008) se compone de un predictor lineal dado por:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}, \quad i=1, \dots, n \quad (2.6)$$

Existen dos funciones, una denominada la de enlace (2.7) y la otra es la función de la varianza (2.8):

La primera corresponde a la función de enlace la cual describe cómo la media, $E(Y_i) = \mu_i$, depende del predictor lineal:

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

La segunda función corresponde a la función de la varianza la cual describe cómo la varianza, $Var(Y_i)$ depende de la media:

$$Var(Y_i) = \phi V(\mu), \quad (2.8)$$

donde el parámetro de dispersión ϕ es conocido.

2.3 Familia Exponencial

La función de la varianza (2.8) juega un papel importante en la Familia Exponencial, ya que caracteriza la distribución. La distribución para la variable respuesta debe pertenecer a la Familia Exponencial (FE) bajo las condiciones de que Y sea una variable aleatoria cuya función de probabilidad depende de un parámetro θ . O cuando se diga que la distribución pertenece a la Familia Exponencial. Puede escribirse de la siguiente manera (Dobson, 2002):

$$f(y, \theta) = \exp(a(y)b(\theta) + c(\theta) + d(y)), \quad (2.9)$$

donde $s(y) = \exp(d(y))$ y $t(\theta) = \exp(c(\theta))$.

Si $\alpha(y) = y$, entonces se dice que la distribución tiene forma canónica o estándar. A su vez $b(\theta)$ recibe el nombre de parámetro natural de la distribución. Si hay otros parámetros

además del parámetro de interés θ estos son considerados parámetros molestos (*nuisance parameters*) que forman parte de las funciones a, b, c, d y son tratados como si fueran conocidos.

2.3.1 Las distribuciones más representativas de la Familia Exponencial

A continuación se presentan algunas de las distribuciones exponenciales más conocidas pertenecientes a la familia.

Distribución Normal

Y es una variable aleatoria $Y \sim N(\mu, \sigma^2)$ y su función de densidad:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\} = \exp\left[\left\{\frac{1}{\sigma^2}\left(\mu y - \frac{\mu^2}{2}\right) - \frac{1}{2}\log(2\pi\sigma^2) + \frac{y^2}{\sigma^2}\right\}\right], \quad (2.10)$$

donde $-\infty < \mu$, $y < -\infty$ y $\sigma^2 > 0$.

En consecuencia: $V(\mu) = 1$

Distribución Poisson

Cuando $Y \sim P(\mu)$ su función de probabilidad viene dada por:

$$p(x; \mu) = e^{-\mu} \mu^y / y! = \exp(y \log \mu - \mu - \log y!), \quad (2.11)$$

donde $\mu > 0$ y $y = 0, 1, 2, \dots$

En consecuencia: $V(\mu) = \mu$.

Distribución Binomial

Y^* es el porcentaje de éxitos en los experimentos n independientes, cada uno con probabilidad de μ . Suponemos que $nY^* \sim B(n, \mu)$. La función de probabilidad de Y^* se expresa como:

$$p(x; \mu) = \binom{n}{nY^*} \mu^{nY^*} (1 - \mu)^{n - nY^*} = \exp\left\{\log\binom{n}{nY^*} + nY^* \log \frac{\mu}{1 - \mu} + n \log(1 - \mu)\right\}, \quad (2.12)$$

donde $0 < \mu$, $Y^* < 1$.

En consecuencia: $V(\mu) = \mu(1 - \mu)$.

Cuadro representativo de las principales distribuciones pertenecientes a la familia exponencial:

A continuación se detalla la estimación de los parámetros de las distribuciones más representativas de la Familia Exponencial presentadas anteriormente. Cabe destacar que el Modelo de Regresión Logística es un caso particular del Modelo Lineal Generalizado también perteneciente a la Familia Exponencial.

Distribución	Parámetro Natural $b(\theta)$	$c(\theta)$	$d(y)$
Binomial	$\log(\pi/(1 - \pi))$	$n\log(1 - \pi)$	$\log\binom{n}{y}$
Normal	u/σ^2	$-\frac{u}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$	$-\frac{y^2}{2\sigma^2}$
Poisson	$\log(\theta)$	$-\theta$	$-\log(y!)$

Tabla 2.1: Algunas distribuciones de la Familia Exponencial

2.4 Modelo de Regresión Logística Múltiple

La Regresión Logística ha constituido uno de los principales métodos para modelar datos estadísticos. Aun cuando la respuesta de interés no es originalmente del tipo binario, algunos investigadores han dicotomizadas la respuesta de manera que la probabilidad de éxito se puede ajustar mediante Regresión Logística (Paula, 2008).

El Modelo de Regresión Logística se compone de una variable de respuesta la cual toma valores 0 o 1. Por consiguiente esta variable aleatoria sigue una distribución Bernoulli con parámetro p , o Binomial con $(1, p)$, con función de probabilidad dada por:

$$f(y; p) = p^y(1 - p)^{1-y} = \exp\left(\frac{y\log\left(\frac{p}{1-p}\right) - \log\left(1 + \exp\left(\log\left(\frac{p}{1-p}\right)\right)\right)}{1} + 0\right). \quad (2.13)$$

Por lo tanto, la distribución de la variable de respuesta pertenece a la Familia Exponencial con:

- Parámetro natural: $\theta = \log\frac{p}{1-p}$
- Parámetro de escala: $\phi = 1$
- $b(\theta) = \log(1 + \exp(\theta))$ y $c(y; \phi) = 0$

Considerando la disposición de p variables independientes expresadas por el vector $X' = (x_1, x_2, \dots, x_p)$ y relacionando la probabilidad de que ocurra un determinado suceso independiente denotado por el vector X' con probabilidad condicional $P(Y = 1/X) = \pi(x)$ en función de p variables independientes que pueden ser cuantitativas o cualitativas según el tipo de diseño de estudio (ver; Solano, 2007).

El logit del Modelo de Regresión Logística Múltiple se presenta por la siguiente ecuación:

$$g(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p,$$

donde $\beta_0, \beta_1, \dots, \beta_p$ son parámetros del Modelo de Regresión Logística Múltiple. En este caso el Modelo de Regresión Logística es:

$$\pi(x) = p_j = \frac{e^{g(x)}}{1 + e^{g(x)}}, \quad j=1, \dots, p.$$

2.4.1 Estimación de los parámetros

Continuando con la investigación del artículo de Solano, para la estimación de los parámetros del modelo se utiliza el método de máxima verosimilitud.

Suponiendo que existe una muestra de n observaciones independientes (x_i, y_i) , $i = 1, 2, \dots, n$. Para ajustar el modelo se requiere tener estimadores del vector $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$. Las ecuaciones de verosimilitud que resultan pueden ser expresadas de la siguiente forma:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0,$$

y

$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0,$$

para $i = 1, 2, \dots, n$.

Para encontrar la solución de este conjunto de ecuaciones se utiliza métodos iterativos y corresponden a los estimadores máximo verosimil de los parámetros como aquellos que maximizan la función de verosimilitud.

2.4.2 Razón de Chance (odds ratio)

Se define el odds ratio como la razón que se establece entre la ocurrencia de un suceso respecto a su no ocurrencia. Esta razón se denomina con el término Odd representado de la siguiente forma: (ver; Almazán, 2000)

$$Odd = \frac{p}{q} = \frac{p}{(1-p)}. \quad (2.14)$$

Considerando dicha definición (2.14) la relación entre Odd y p podemos expresarla de la siguiente forma:

$$p = \frac{Odd}{1 + Odd}. \quad (2.15)$$

Como se sabe, los Odds varían desde 0 a $+\infty$. Téngase en cuenta que p varía desde 0 a 1. Por lo tanto, cuando p está muy cerca de 1:

$$Odd = \frac{p}{1-p} = \frac{1}{0} = +\infty. \quad (2.16)$$

Y en caso inverso:

$$Odd = \frac{p}{1-p} = \frac{0}{1} = 0. \quad (2.17)$$

2.5 Chi-squared Automatic Interaction Detection

El método que se detalla a continuación es de gran utilidad para seleccionar las covariables más significativas de la base de datos.

Al Modelo de Regresión Logística ingresan todas las covariables existentes, entonces se procede a realizar el método de CHAID para seleccionar las covariables más significativas, el cuál responderá al objetivo específico número uno. Este método se explica a continuación.

La técnica de CHAID es usada en mayor medida para la investigación de grandes cantidades de datos categóricos. Está conformada por la rama de la AID (*Automatic Interaction Detection*), dicha rama ha sido diseñada para una variable dependiente categorizada. El método de CHAID es una técnica propuesta inicialmente por (Kass, 1980) y posteriormente una extensión fue propuesta por (Biggs, 1991).

El modelo de CHAID procede mediante la creación de un árbol de decisión. Al mismo tiempo este clasifica casos en grupos o pronostica valores de una variable (criterio) dependiente basada en valores de covariables (predictores). El procedimiento proporciona herramientas de validación para análisis de clasificación exploratorios y confirmatorios. En el análisis, depende del tipo de variable para realizar el test, por ejemplo:

Variable de respuesta	Covariable	Técnica
cuantitativa	cuantitativa	Chi-cuadrado
cuantitativa	cuantitativa	ANOVA
cuantitativa	cuantitativa	Regresión
cuantitativa	cuantitativa	Logística

Tabla 2.2: Análisis según tipo de variables

2.5.1 Técnicas Potenciales

La técnica de CHAID puede ser utilizada en diversos procedimientos que facilitan la visión de resultados, a raíz de esto es mucho más práctico obtener interpretaciones orientadas a la técnica. Según el IBM Knowledge Center el procedimiento puede ser utilizado para distintas técnicas, por ejemplo en primer lugar la técnica de **segmentación** identifica

las personas que pueden ser miembros de un grupo específico. En segundo lugar la técnica **estratificación** asigna los casos a una categoría de entre varias, por ejemplo, grupos de alto riesgo, riesgo intermedio y bajo riesgo. En tercer lugar la técnica de **predicción** es la encargada de crear reglas y de utilizarlas para predecir eventos futuros, por ejemplo la verosimilitud de que una persona cause mora en un crédito o el valor de reventa potencial de un vehículo o una casa. En cuarto lugar La técnica de **reducción de datos y clasificación de variables** selecciona un subconjunto útil de predictores a partir de un gran conjunto de variables para utilizarlo en la creación de un modelo paramétrico formal. En quinto lugar la técnica de **identificación de interacción** se ocupa de identificar las relaciones que pertenecen sólo a subgrupos determinados y especifica dichas relaciones en un modelo paramétrico formal. En sexto lugar la técnica de **fusión de categorías y discreción de variables continuas** vuelve a codificar las variables continuas y las categorías de los predictores del grupo, con una pérdida mínima de información.

2.5.2 Algoritmo

Diversos son los procedimientos que se conjugan a lo largo de la técnica de CHAID. Se definen en ella una serie de elementos que dan forma al árbol de clasificación en donde son seleccionadas las covariables más significativas para el modelo (Kass, 1980). Para conocer en profundidad dichos procesos de la técnica es necesario ver en detalle.

1. Para cada predictor, sus categorías son tabulación cruzada con las categorías de la variable de respuesta.
2. Encuentra la pareja de categorías del predictor (considerando sólo pares permitidos según lo determine el tipo de predictor) cuya sub-tabla es menos significativamente diferente. Si esta significación no alcanza un valor crítico, entonces fusiona las dos categorías. Por último ha de considerarse esta fusión como una nueva categoría compuesta y repetir el paso.
3. Para cada categoría compuesta que consta de tres o más de las categorías originales, el algoritmo de CHAID encuentra la división binaria más significativa (limitado por el tipo de predictor). Si este significado sobrepasa cierto límite pone en práctica la división para formar dos nuevas categorías. Posteriormente ha de repetirse el paso dos.
4. Calcula la importancia de cada predictor fusionado de manera óptima y aísla el más significativo. Si la importancia es mayor que un valor de criterio, subdivide los datos de acuerdo a las categorías (fusionadas) del predictor elegido.
5. Para cada partición de los datos que aún no ha sido analizada es necesario volver al paso número uno. Este paso puede modificarse mediante la exclusión de otras particiones de análisis con un número de observaciones.

Paso condicional:

6. algoritmo CHAID requiere una prueba de significación de la tabla de contingencia reducida. Se necesita el multiplicador de Bonferroni para tener en cuenta el número de formas. El cálculo será distinto según la reducción de una de las siguientes tres categorías que se utilice, las cuales pueden ser **variables ordinales**: como en la AID, un predictor monótonico es uno cuyas categorías se encuentran en una escala ordinal, ésta implica que solo pueden agruparse categorías continuas, **variables nominales**: una vez más como en la AID convencional, un predictor libre es uno cuyas categorías son puramente nominales, esto implica que cualquier agrupación de categorías es permisible y **variables ordinales con valores perdidos**: al predictor, se le llama “predictor flotante”, esto surge cuando se permite una categoría desconocida, la información se recodifica a través de alguna técnica que asigna un valor a la información faltante y estos no alteran los análisis del AID.

Una vez seleccionadas las covariables que mejor describen a la variable de respuesta a través del método CHAID, se procede a la realización de Stepwise para definir las covariables más significativas, el método de Stepwise se explica a continuación.

2.6 Stepwise

Un método ampliamente aplicado en el método de Regresión Logística es Stepwise. La selección de variables se basa en la comparación de la devianza entre modelos, el procedimiento habitual es el llamado Stepwise (paso a paso) en el que, mediante contrastes condicionales de razón de verosimilitudes se comparan modelos con diferentes variables. Este método sirve para encontrar el modelo de ajuste más adecuado al momento de seleccionar las covariables que pretendan explicar la variable de respuesta, para lo cual selecciona un subconjunto que resulte suficientemente explicativo. No obstante en la práctica, la selección del subconjunto de covariables de los Modelos de Regresión se deja en manos de procedimientos relativamente automáticos. Este método consta de 3 procedimientos: (Cañadas, 2013)

2.6.1 Método Backward:

En este método se comienza por incluir en el modelo a todas las covariables disponibles. Según su capacidad explicativa se van eliminando del modelo de una en una. En concreto, la primera variable que se elimina es aquella que presenta un menor coeficiente de correlación parcial con la variable de respuesta.

2.6.2 Método Forward:

En este método se comienza por un modelo que no contiene ninguna covariable y se añade como primera de ellas a la que presente un mayor coeficiente de correlación (en valor absoluto) con la variable de respuesta. En los pasos sucesivos se va incorporando al modelo aquella variable que presenta un mayor coeficiente de correlación parcial con la variable dependiente dadas las covariables ya incluidas en el modelo. El procedimiento se detiene cuando el incremento en el coeficiente de determinación debido a la inclusión de una nueva covariable en el modelo ya no es importante.

2.6.3 Método Stepwise:

Este método es uno de los más empleados y consiste en una combinación de los dos anteriores. En el primer paso se procede como en el método Forward pero a diferencia de éste en el que cuando una variable entra en el modelo posteriormente ya no vuelve a salir. En el procedimiento Stepwise es posible que la inclusión de una nueva variable afecte a que otra incluida previamente en el modelo resulte redundante y sea “expulsada” de él.

La selección de modelos no está exenta de polémica, ya que según sea el método utilizado “backward”, “forward”, o “stepwise” puede dar lugar a modelos totalmente diferentes. El empleo de estos métodos nunca debe sustituir al criterio del investigador (Silva, 2004), ni considerar sin más a las variables que se quedan en el modelo, como principales responsables del efecto en la variable dependiente.

2.6.4 Criterio de Información Akaike (AIC)

Un procedimiento más sencillo para seleccionar variables explicativas en un Modelo Logístico es a través del método de Akaike. Este criterio se define en base a la log-verosimilitud y al número de parámetros del modelo, el valor del AIC se calcula a través de (ver Cañadas, 2013):

$$AIC = -2 * L_{modelo} + k * \text{número parámetros estimados.}$$

Con $k = 2$ y $L = \text{log-verosimilitud}$.

Este criterio penaliza los modelos con muchos parámetros frente a los modelos más parsimoniosos. Se elige el modelo con menor valor de AIC. El proceso continúa hasta que la inclusión o eliminación de alguna variable incrementa el valor del AIC en vez de reducirlo.

DIAGNÓSTICO EN EL MODELO DE REGRESIÓN LOGÍSTICA

Una vez comprobado el ajuste global del Modelo de Regresión Logística, se estudiará la falta de ajuste a nivel de cada observación y cómo afectan las observaciones al modelo general. Para eso, se realiza Análisis de los Residuos y de los valores influyentes.

El Análisis de Diagnóstico es de suma importancia para el Modelo de Regresión Logística, porque se puede ver de forma clara a través de sus gráficos los datos fuera de tendencia. A continuación, en este capítulo se explica el Análisis de Residuos, el cual tiene como objetivo identificar la existencia de datos atípicos y luego estudiar Análisis de Diagnóstico para detectar datos que puedan ser influyentes, este Análisis de Diagnóstico responderá al objetivo específico número tres.

3.1 Análisis de Residuos

A continuación se presenta el Análisis de Residuos, el cual detecta la presencia de puntos aberrantes y evalúa la adecuación de la distribución propuesta para la variable de respuesta.

Existen varios tipos de residuos que sugieren si una observación es atípica o no.

Los residuos son las diferencias entre los valores estimados por el modelo y los valores observados. Sin embargo, muchas veces se utilizan los residuos estandarizados (Cayuela, 2015). Para ver de mejor forma los residuos, se pueden analizar una serie de gráficos, como

por ejemplo:

- Histograma de los residuos
- Gráfico de residuos frente a valores estimados
Estos dos gráficos pueden indicar falta de linealidad, heterocedasticidad (varianza no constante) y valores atípicos.

3.1.1 Residuos de Pearson

Los Residuos de Pearson se definen como (ver; Solano, 2007):

$$r_j = \frac{y_j - m_j \hat{p}_j}{\sqrt{m_j \hat{p}_j (1 - \hat{p}_j)}}, \quad (3.1)$$

donde, y_j representa el número de respuestas, $y = 1$, entre los m_j individuos (en este caso pacientes) con $X_j = x$ (algunos individuos que tienen el mismo valor x), $j = 1, 2, \dots, p$.

$$\hat{p} = \hat{\pi}(x) = \frac{e^{\hat{g}(x)}}{1 + e^{\hat{g}(x)}}, \quad (3.2)$$

$$y \quad \hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p.$$

El Residuo de Pearson es similar al Residuo estudentizado usado en la Regresión Lineal. Así, un Residuo de Pearson en valor absoluto mayor que 2 indica un dato atípico. La estadística X^2 de Pearson es la suma de cuadrados de los Residuos de Pearson.

$$X_P^2 = \sum_{i=1}^J r_j^2. \quad (3.3)$$

3.1.2 Residuos de Pearson Estandarizado

Según (Solano, 2007) los Residuos de Pearson Estandarizado están definidos por:

$$r_{sj} = \frac{r_j}{\sqrt{1 - h_j}}, \quad j=1, \dots, p \quad (3.4)$$

donde r_j son los Residuos de Pearson y h_j es el Leverage, esto es, el elemento de la diagonal principal de la matriz H (se explica a continuación en Análisis de Diagnóstico).

3.1.3 Residuos de la Devianza

Los Residuos de la Devianza están definidos de la siguiente forma:

$$d_j = \pm \left[2 \left(y_j \ln \left(\frac{y_j}{m_j \hat{p}_j} \right) + (m_j - y_j) \ln \left(\frac{m_j - y_j}{m_j (1 - \hat{p}_j)} \right) \right) \right]^{1/2}, \quad (3.5)$$

La Devianza es la suma de cuadrados de los Residuos de la Devianza:

$$X_D^2 = \sum_{i=1}^p d_j^2. \quad (3.6)$$

Si la Devianza es mayor que 4 en valor absoluto entonces la observación correspondiente es atípica.

Las Medidas de Influencia en el Modelo de Regresión Logística son presentadas en el siguiente párrafo.

3.2 Medidas de Influencia

En las técnicas de diagnóstico pueden existir observaciones que situándose lejos del resto pueden influir en las estimaciones del modelo. Las Medidas de Influencia detectan los valores influyentes analizando el efecto que esos valores tienen en los parámetros del modelo (Solano, 2007) con el fin de discriminar las variables con mayor influencia para las estimaciones del modelo.

Son de gran utilidad para evaluar la presencia de datos influyentes sobre el Modelo de Regresión Logística ajustado, para efectos de este trabajo se tomarán en cuenta las Medidas de Influencia de Leverage y la Distancia de Cook.

3.2.1 Leverage

Leverage es utilizado para detectar datos atípicos o que estén fuera de tendencia con respecto a las demás observaciones, es por eso que aquellos datos que tengan valores fuera de rango en el modelo de regresión tendrán valores altos en el estadístico de Leverage. En tal caso, serán designados como observaciones con alto valor de Leverage. El estadístico de Leverage tiene un recorrido cerrado entre 0 (que es su valor mínimo) y 1 (que es el valor máximo) (Fernández, 2004).

Son los elementos de la diagonal de la matriz de predicción H . El leverage para la observación i -ésima es el elemento i -ésimo de la diagonal principal de la matriz H , h_i , toma valores entre 0 y 1 con un valor medio de p/n .

$$H = X_*(X_*'X_*)^{-1}X_*, \quad i=1, \dots, n \quad (3.7)$$

donde $X_* = V^{1/2}X$ y $V = \text{diag}(\hat{\pi}_i(1 - \hat{\pi}_i))$.

El estadístico de influencia derivado de h_i es la Distancia de Cook presentada a continuación.

3.2.2 Distancia de Cook

La Distancia de Cook mide la influencia de la estimación de los parámetros del modelo y se define como (Solano, 2007):

$$\Delta B_j = \frac{r_{sj}^2 h_j}{(1 - h_j)}, \quad (3.8)$$

donde r_{sj} : Residuos de Pearson estandarizado. Si $\Delta B_j > 1$, la j -ésima observación es influyente en los valores de los parámetros estimados.

Tanto Leverage como la Distancia de Cook muestran los datos o en este caso pacientes de la base de datos que puedan ser influyentes, esto se ve de mejor forma a través de gráficos, como por ejemplo los que se mencionan a continuación.

3.2.3 Gráficos de Diagnósticos

Los Gráficos de Diagnóstico son de gran utilidad para detectar la presencia de datos influyentes en el Modelo de Regresión Logística. Para obtener una descripción rápida de la información proporcionada por las estadísticas, se pueden utilizar los siguientes gráficos (Solano, 2007):

- Delta Chi-cuadrado vs la probabilidad estimada.
- Delta Devianza vs la probabilidad estimada.
- Distancia de Cook vs la probabilidad estimada.
- Delta Chi-cuadrado vs leverage.
- Delta Devianza vs leverage.
- Distancia de Cook vs leverage.

APLICACIÓN

En el presente estudio se trabajó con la totalidad de pacientes prescritos con Benzodiazepinas durante los años 2013, 2014 y 2015 pertenecientes al CESFAM de Llaylay, los cuales corresponden a 824 pacientes.

En este capítulo se presentan los resultados correspondientes a las metodologías expuestas en el capítulo 2 y 3. En primer lugar comenzando por la construcción del árbol de clasificación y luego el método Stepwise, la conjunción de estos dos elementos da cuerpo a la selección de las covariables a ser incluidas en el Modelo de Regresión Logística. Posteriormente se realiza un Análisis de Residuos en el cual se muestran los pacientes que presentan datos atípicos y finalmente la ejecución de un Análisis de Diagnóstico para ver los pacientes que son influyentes y puedan afectar en las estimaciones del riesgo de dependencia.

Con la obtención de los resultados se espera guiar una interpretación idónea a la naturaleza del fenómeno de dependencia de fármacos.

4.1 Descriptiva de los Datos

A continuación se presenta un análisis marginal de la variable de respuesta y las covariables consideradas en la base de datos. La descripción de los datos permite en este punto obtener una visión marginal de las características de cada una de las variables pertenecientes a la base de datos lo cual ayuda y provee de una mejor comprensión de como las variables mencionadas se relacionan con el fenómeno de dependencia de drogas.

La primera variable presentada es la variable de respuesta la cual toma valor 0 o 1 dependiendo si el paciente tiene riesgo o no a ser dependiente del fármaco, respectivamente.

Variable de respuesta: El riesgo de dependencia se presenta a continuación como la variable explicativa. El detalle de la variable de respuesta se presenta en la tabla 4.1 y en la figura 4.1.

Dependencia	Nº de pacientes	% de pacientes
Sin riesgo	349	42,35 %
Con riesgo	475	57,65 %
Total	824	100 %

Tabla 4.1: Distribución de frecuencia de la variable de respuesta.

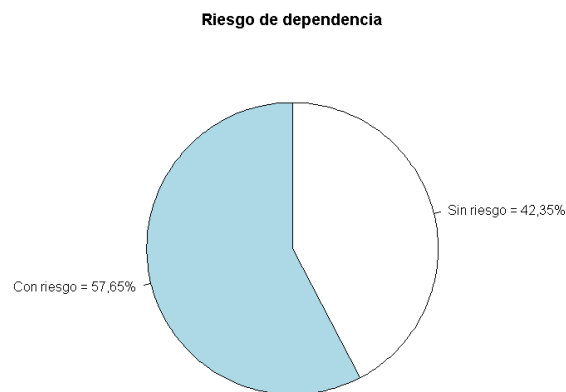


Figura 4.1: Gráfico sectorial de la variable de respuesta.

La variable de respuesta representa el riesgo de dependencia en los pacientes, de los 824 pacientes 475 se encuentra con riesgo a ser dependiente del fármaco que corresponde al 57,65 % de los pacientes. Por otro lado 349 de 824 pacientes los cuales representan el 42,35 % de la población no tiene riesgo a ser dependiente del fármaco.

Variable edad: Esta variable está representada en años y es la segunda variable tratada en la base de datos, que corresponde a una variable cuantitativa de carácter continuo. Que se presenta en la tabla 4.2 y en la figura 4.2 y 4.3.

Variable	Observaciones	Promedio	Desviación estándar	Mínimo	Máximo
edad	824	51,10	17,37	15	98

Tabla 4.2: Estadísticos de resumen de la variable edad.

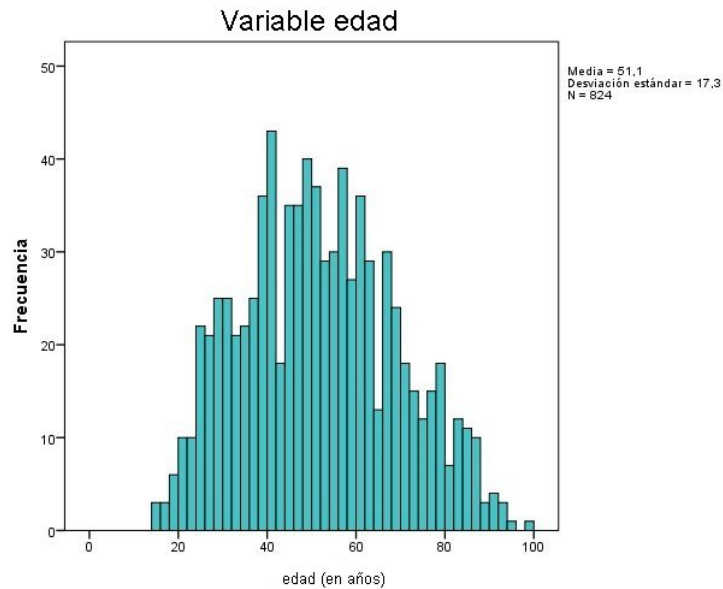


Figura 4.2: Histograma de la edad de los pacientes.

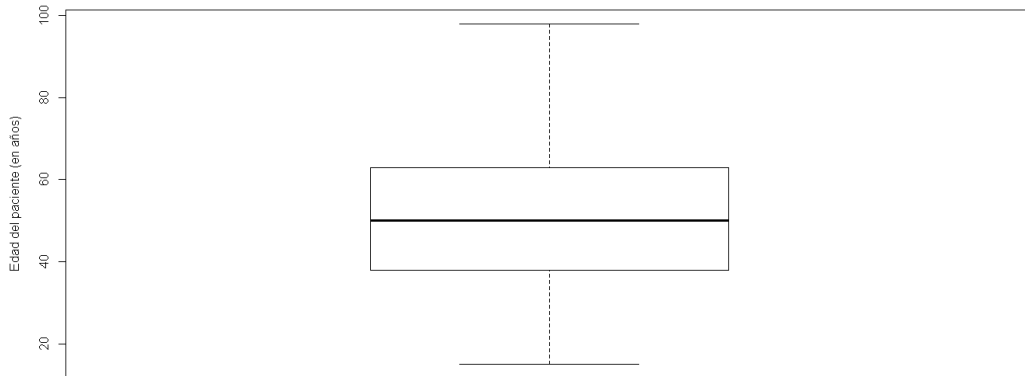


Figura 4.3: Box plot de la edad de los pacientes.

En la figura 4.2 se puede notar que la mayor frecuencia de edad se centra en los 41 años. De los pacientes que asistieron al CESFAM durante los tres años, el menor de ellos tiene 15 años mientras que el paciente con mayor edad tiene 98 años. Se puede visualizar una razonable simetría en las edades, mientras que la mayor parte de los pacientes se centran en la edad de 50 años aproximadamente. En la figura 4.3 el box plot representa en la parte superior el máximo valor no atípico donde se encuentra el paciente que tiene mayor edad, luego al comienzo de la caja está el cuartil tres que se encuentra entre los 65 años aproximadamente, en el centro está la mediana y luego el cuartil uno que se encuentra entre los 38 años aproximadamente, en la parte inferior que es el mínimo valor no atípico y ahí está el paciente que tiene menor edad. En este caso no hay valores atípicos porque no existen puntos o valores fuera del gráfico.

Variable sexo: Esta variable caracteriza el género del paciente en la base de datos que se presenta en la tabla 4.3 y en la figura 4.4.

sexo	Nº de pacientes	% de pacientes
Mujer	580	70,39 %
Hombre	244	29,61 %
Total	824	100 %

Tabla 4.3: Distribución de frecuencia del sexo del paciente.

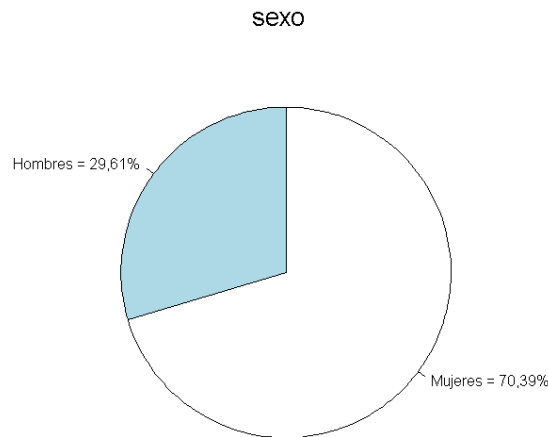


Figura 4.4: Gráfico sectorial de la variable de sexo.

La caracterización de la variable sexo indica que de los 824 pacientes 580 son de sexo femenino correspondiente al 70,39% del total, mientras que 244 son de sexo masculino pertenecientes al 29,61% de la población, esto indica que los pacientes prescritos en mayor medida con benzodiazepinas corresponden a mujeres.

Variable tipo de enfermedad: El tipo de enfermedad caracteriza a los pacientes según la enfermedad que padezcan, alcanzando 7 categorías de enfermedades. La distribución de la variable tipo de enfermedad se presenta en la tabla 4.4 y en la figura 4.5. Las 7 categorías de tipo de enfermedad que padece el paciente están agrupadas de la siguiente manera:

- T1 = tipo de enfermedad 1: Depresión.
- T2 = tipo de enfermedad 2: Trastornos.
- T3 = tipo de enfermedad 3: Alcohol y Drogas.
- T4 = tipo de enfermedad 4: Epilepsia, Esquizofrenia, Hipertensión, Tartamudéz y Problemas familiares.
- T5 = tipo de enfermedad 5: Repetición de receta.

- T6 = tipo de enfermedad 6: Demencia, Dislímia, Insomnio y Efecto adverso.
- T7 = tipo de enfermedad 7: Otros.

tipo de enfermedad	Nº de pacientes	% de pacientes
T1	388	47,09 %
T2	284	34,47 %
T3	39	4,73 %
T4	71	8,62 %
T5	11	1,33 %
T6	12	1,46 %
T7	19	2,31 %
Total	824	100 %

Tabla 4.4: Distribución de frecuencia del tipo de enfermedad del paciente.

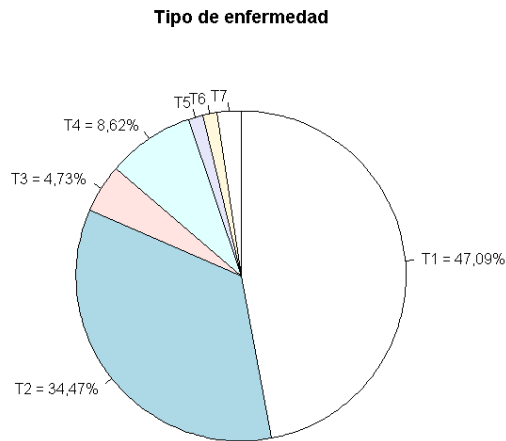


Figura 4.5: Gráfico sectorial de la variable tipo de enfermedad.

En la tabla 4.4 y en la figura 4.5 se puede percibir que la mayor cantidad de pacientes se centra en el diagnóstico 1 y 2, los cuales pertenecen a los casos padecientes con depresión o algún tipo de trastorno.

Variable tipo de fármaco: La Benzodiazepina es una familia de fármacos de las cuales solo tres tipos son distribuidas a los pacientes atendidos en el CESFAM. La distribución de la variable tipo de fármaco se presenta en la tabla 4.5 y en la figura 4.6.

tipo de fármaco	Nº de pacientes	% de pacientes
Clonazepan 2Mg.	126	15,29 %
Clonazepan 0,5Mg.	545	66,14 %
Diazepan 10 Mg.	153	18,57 %
Total	824	100 %

Tabla 4.5: Distribución de frecuencia de la variable tipo de fármaco.

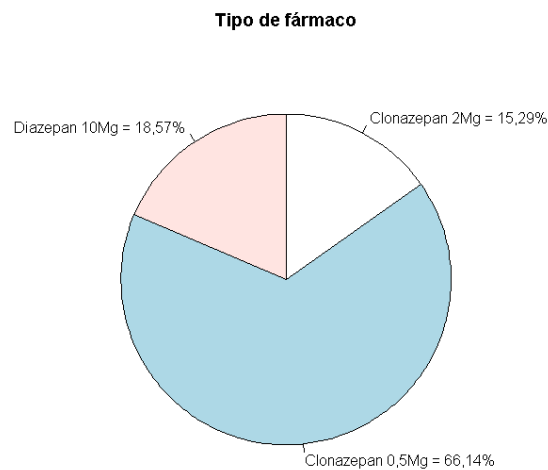


Figura 4.6: Gráfico sectorial de la variable tipo de fármaco.

Con respecto a la tabla 4.5 se puede visualizar que de los tres tipos de fármacos que entrega en CESFAM, el que tiene una mayor prescripción en los pacientes es el Clonazepan

de 0,5 Mg. Perteneciente al 63,14% de los pacientes.

Variable tipo de receta: El tipo de receta es la caracterización del paciente según su tipo de enfermedad, si tiene una enfermedad de tipo crónica es una enfermedad de larga duración y por lo general de progresión lenta, si es de tipo salud mental corresponde a una condición de trastornos y si es una enfermedad de morbilidad es por tiempo determinado. La distribución de frecuencia de la variable tipo de receta se presenta en la tabla 4.6 y en la figura 4.7.

tipo de receta	Nº de pacientes	% de pacientes
Crónica	540	65,53 %
Salud Mental	185	22,45 %
Morbilidad	99	12,01 %
Total	824	100 %

Tabla 4.6: Distribución de frecuencia de la variable tipo de receta.

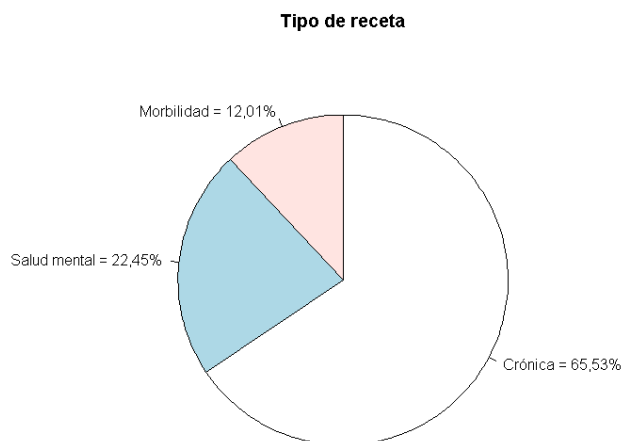


Figura 4.7: Gráfico sectorial de la variable tipo de receta.

En el cuadro 4.6 se puede ver que la mayoría de los pacientes padece una enfermedad de tipo crónica de larga duración y por lo general de progresión lenta. El porcentaje de los pacientes crónicos es de 65,53 %. Esto se ve de mejor forma en la figura 4.7.

frecuencia: Esta variable representa el número de veces que asiste el paciente a retirar el fármaco. Se presenta en la tabla 4.7 y en la figura 4.8 y 4.9.

Variable	Observaciones	Promedio	Desviación estándar	Mínimo	Máximo
frecuencia	824	7,41	8,17	1	30

Tabla 4.7: Estadísticos de resumen de la variable frecuencia.

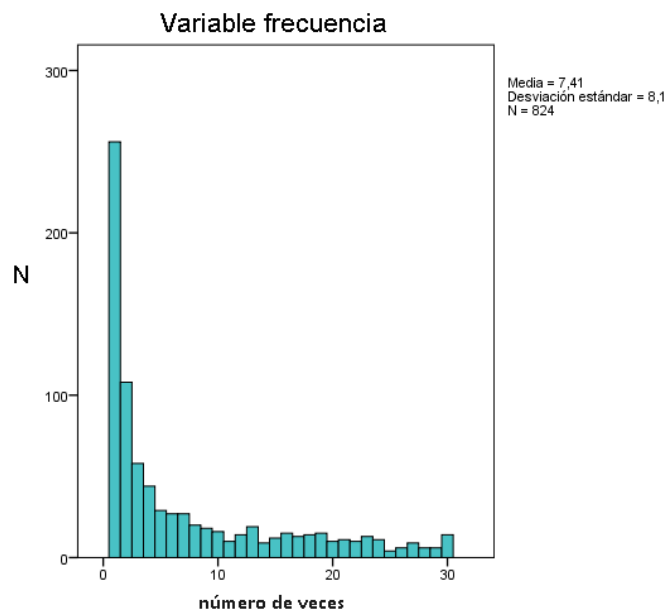


Figura 4.8: Histograma de la variable frecuencia.

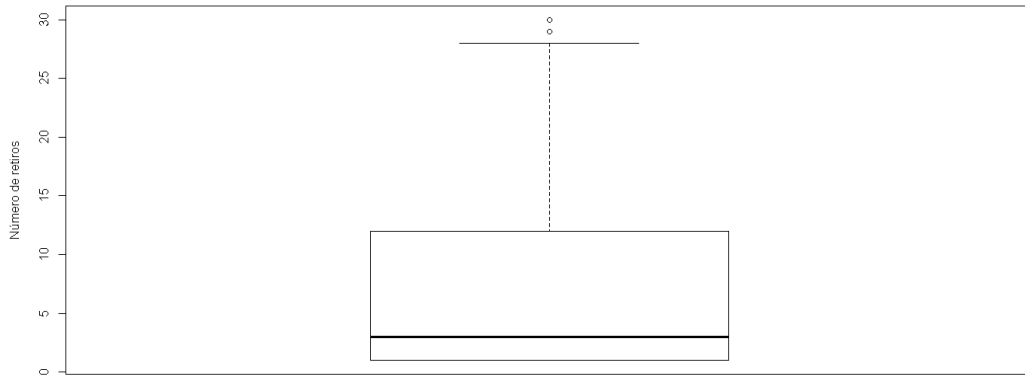


Figura 4.9: Box plot de la variable de frecuencia.

En la tabla 4.7 se distingue que los pacientes que más asistieron al CESFAM a retirar el fármaco, lo hicieron en un rango de retiro de 30 veces durante los 3 años, sin embargo como puede observarse en el histograma, la mayoría de ellos asistió solo una vez, esto se ve claramente en la figura 4.8.

En la figura 4.9 el box plot representa en la parte superior el máximo valor no atípico donde se encuentran los pacientes que más veces han asistido al CESFAM durante los tres años a retirar el fármaco, luego al comienzo de la caja está el cuartil tres que se encuentra entre las 12 veces aproximadamente, en el centro está la mediana y luego el cuartil uno que se encuentra en los pacientes que asistieron una vez. En este caso hay presencia de valores atípicos, se puede ver en el gráfico dos puntos que sobresalen de la parte superior y son los pacientes que han asistido más veces al CESFAM a retirar el fármaco durante los tres años.

4.2 Resultados Método CHAID

A continuación se presentan los resultados obtenidos mediante la aplicación del método de CHAID, que corresponde a una técnica para seleccionar las variables que mejor predigan la variable de respuesta la cual para este caso es riesgo de dependencia.

La figura 4.10 muestra el árbol de clasificación y selecciona las covariables más significativas.

Especificaciones	
Variable de respuesta	riesgo de dependencia
Covariables	edad, sexo, tipo de enfermedad, artículo, tipo de receta, frecuencia.
Resultados	
Covariables incluidas	frecuencia, edad y tipo de receta.
Número de nodos	9
Número de nodos terminales	6
Profundidad	2

Tabla 4.8: Características del árbol de clasificación.

La figura 4.10 es una segmentación que se realizó mediante la técnica CHAID correspondiente a un tipo de árbol y útil al momento de encontrar las variables condicionantes de la probabilidad de riesgo de dependencia. Esta segmentación produjo varios grupos que contienen a las variables más significativas.

La figura 4.10 está conformada por 9 nodos. En el nodo 0 se recoge el total de la población, en dicho total se dividen los pacientes que tienen riesgo a ser dependientes del fármaco y los que no. El 57,6% de los pacientes tiene riesgo a ser dependiente del fármaco, en tanto el 42,4% de los pacientes no tiene riesgo de ser dependiente del fármaco. Desde el nodo 0 se dividen cuatro subnodos los cuales se forman mediante la covariable denominada “frecuencia” (altamente significativa de acuerdo al *valor-p*). CHAID ha encontrado que algunos de estos grupos no eran significativamente distintos por lo cual los ha fusionado en un solo grupo. La caracterización de los cuatro nodos es la siguiente:

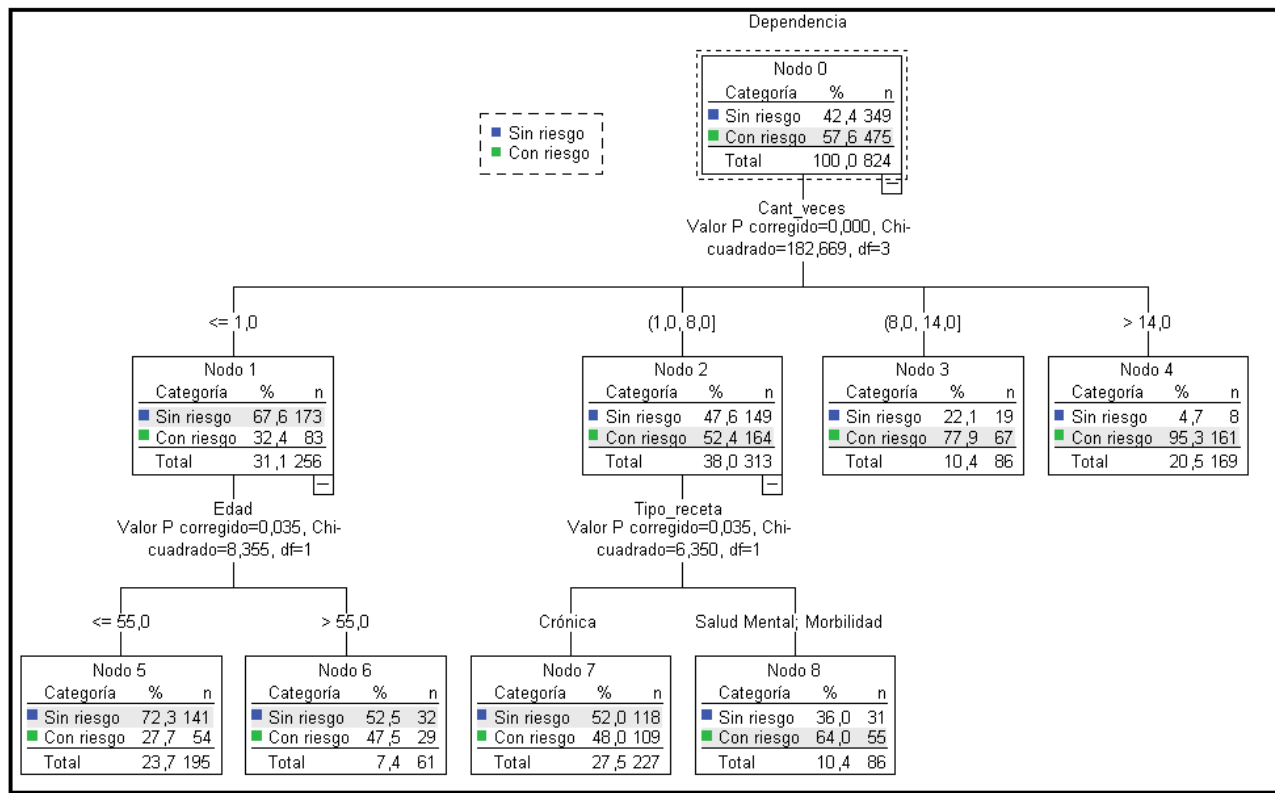


Figura 4.10: Árbol de clasificación del método CHAID.

- En el nodo 1 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 32,4% en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 67,6%. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido una vez al CESFAM a retirar el fármaco durante los años 2013 al 2015.
- En el nodo 2 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 52,4% en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 47,6%. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido entre 1 y 8 veces al CESFAM a retirar el fármaco durante los años 2013 al 2015.
- En el nodo 3 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 77,9% en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 22,1%. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido entre 8 y 14 veces al CESFAM a retirar el fármaco durante los años 2013 al 2015.

- En el nodo 4 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 95,3 % en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 4,7 %. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido más de 14 veces al CESFAM a retirar el fármaco durante los años 2013 al 2015.

Del nodo 1 se desprende la covariable “edad”, la cual se subdivide en dos nodos, el nodo 5 y el nodo 6. La caracterización de estos dos nodos es la siguiente:

- En el nodo 5 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 27,7 % en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 72,3 %. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido una vez al CESFAM a retirar el fármaco durante los años 2013 al 2015 y tienen 55 años o menos.
- En el nodo 6 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 47,5 % en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 52,5 %. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido una vez al CESFAM a retirar el fármaco durante los años 2013 al 2015 y tienen más de 55 años.

Del nodo 2 se desprende la covariable “tipo de receta”, la cual se subdivide en dos nodos, el nodo 7 y el nodo 8. La caracterización de estos dos nodos es la siguiente:

- En el nodo 7 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 48 % en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 52 %. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido entre 1 y 8 veces al CESFAM a retirar el fármaco durante los años 2013 al 2015 y padecen de una enfermedad de tipo crónica.
- En el nodo 8 el porcentaje de pacientes que tiene riesgo de ser dependiente del fármaco es de 64 % en tanto el porcentaje de los pacientes que no tiene riesgo de dependencia es de 36 %. Los pacientes pertenecientes a este nodo corresponden al rango etario de los que han asistido entre 1 y 8 veces al CESFAM a retirar el fármaco durante los años 2013 al 2015 y padecen de una enfermedad de tipo salud mental o morbilidad.

A través de este árbol puede obtenerse información relevante sobre la caracterización de los pacientes con mayor y menor porcentaje de riesgo de dependencia, así como también los rangos etarios que describen dichos porcentajes:

- Los pacientes con mayor riesgo a ser dependientes, corresponden a los que han asistido durante los años 2013 al 2015 más de 14 veces al CESFAM a retirar el fármaco.
- Los pacientes con un menor riesgo a ser dependientes, corresponden a los que han asistido al CESFAM una vez a retirar el fármaco y tienen 55 o menos años de edad.

4.3 Resultados Método Stepwise

A continuación el Método Stepwise (paso a paso) ha seleccionado dos de seis covariables de la base de datos exponiendo los valores del AIC general y en detalle. Para efectos de esta interpretación de datos la siguiente tabla muestra cuales son las variables más significativas que deben ser incluidas en el modelo:

AIC=913,38			
dependencia \sim tipo de enfermedad + frecuencia			
Covariable	gl	Devianza	AIC
+ edad	1	895,38	913,38
+ tipo de receta	2	894,28	914,28
+ sexo	1	896,72	914,72
+ articulo	2	896,42	916,42
- tipo de enfermedad	6	916,98	920,98
- frecuencia	1	1101,99	1115,99

Tabla 4.9: Selección de las covariables.

El proceso de Stepwise comienza incluyendo todas las covariables al modelo y testeando todas sus posibles combinaciones y termina cuando el modelo tiene un menor AIC. Según el valor del AIC, este modelo es el que mejor se ajusta, ya que si se elimina la covariable tipo de enfermedad el valor del AIC aumenta en 920,98, sin embargo si se eliminase la covariable cantidad de veces el valor del AIC aumentaría a 1115,99. El mejor modelo será el que incluya las covariables tipo de enfermedad y frecuencia.

4.4 Resultados Modelo de Regresión Logística

La Regresión Logística ha constituido uno de los principales métodos de modelado de datos estadísticos. En el Modelo de Regresión Logística ingresan todas las covariables, desde el cual son seleccionadas las más significativas según el *valor-p*. El Modelo de Regresión Logística propuesto en capítulo 2 aplicado a las observaciones analizadas puede ser definido como sigue:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

A continuación se presentan los resultados obtenidos mediante el Modelo de Regresión Logística. Dichos resultados corresponden a la estimación de los parámetros del modelo.

Coficiente	Estimación	Error estándar	Valor Z	$Pr(> z)$	Odds
constante	-0,984	0,394	-2,498	0,012	0,374
frecuencia	0,186	0,019	9,841	0,000	1,204
tipo de enfermedad 4	-1,125	0,366	-3,072	0,002	0,325
tipo de enfermedad 5	-2,532	0,983	-2,576	0,009	0,079
AIC = 917,75					

Tabla 4.10: Estimación de parámetros del modelo de regresión logística ajustado.

- tipo de enfermedad 4: Epilepsia, Esquizofrenia, Hipertensión, Tartamudez o Problemas familiares.
- tipo de enfermedad 5: Repetición de la receta.

Estas covariables son de mayor influencia para predecir el riesgo de dependencia en los pacientes, frecuencia ($Pr(> |z|) = 0,000$), luego dos de las categorías de la covariable tipo de enfermedad, las cuales son tipo de enfermedad 4 ($Pr(> |z|) = 0,002$) y tipo de enfermedad 5 ($Pr(> |z|) = 0,009$), las tres covariables seleccionadas tienen un *valor-p* muy cercano cero y son significativas a un nivel de significancia del 0,05.

Interpretación de los parámetros:

- Estimación: Las estimaciones son los betas, como en el caso de la regresión simple, los betas no son interpretados, el número en sí no tiene una interpretación pero la magnitud indica la importancia de la variable (mayor valor absoluto mayor importancia de la variable, es decir más significativa es la variable).
- Error estándar: Es la raíz de la varianza del estimador, es análogo a una desviación estándar del estimador β_i por ejemplo para β_i se tiene una estimación puntual (el número de la salida) y esa estimación tiene asociada una varianza (es decir no es fija). Este valor de error estándar indica que esa variable es la estimación de cada parámetro. Se espera que este valor sea pequeño y se pueda construir el intervalo de confianza para los β_i de la siguiente forma:

$$[\beta_i - 1,96 * EE \quad ; \quad \beta_i + 1,96 * EE]$$

Donde EE: error estándar. 1,96 es el percentil 0,975 de la distribución normal, ya que todo estimador de verosimilitud máxima se distribuye normal.

- Valor Z: Es el valor de β estandarizado. Para tamaños de muestras grandes se puede obtener la distribución del estimador del parámetro, que es $Z = \beta/EE$, está demostrado que ésta es normal y por lo tanto el valor es grande. También se sabe que el parámetro es significativo (distinto de cero) cuando su puntuación Z es grande o el valor-p es pequeño.
- Valor-p: según la dócima $H_0 : \beta_i = 0$ v/s $H_1 : \beta_i \neq 0$. El valor-p es la probabilidad de rechazar H_0 , cuando H_0 es verdadero (es la mínima probabilidad de cometer el error tipo I). En resumen cuando el valor-p es menor que 0,05 se rechaza H_0 , eso quiere decir que las covariables no influyen en el modelo, en otras palabras $H_0 : \beta_i = 0$ y la covariable es significativa. En este caso la variable más significativa es frecuencia porque el valor p está más cercano a cero.
- Odds: Indica el chance, es decir cuanto más probable es el éxito que el fracaso en el momento que la variable de respuesta toma el valor cero. La variable frecuencia tiene un odds=1,204 es mayor a 1, entonces la asociación es positiva, es decir que la cantidad de veces se asocia al riesgo de dependencia del paciente, por tanto se considera como un factor de riesgo. Las variables tipo de enfermedad 4 y tipo de enfermedad 5 tienen un odds menos a 1 entonces su asociación es negativa, es decir los pacientes que padecen este tipo de enfermedades no se asocian con la variable de respuesta que es riesgo de dependencia, por tanto se considera un factor de protección.

4.5 Resultados Análisis de Residuos

En la subsección 3.1 del capítulo anterior se describieron los distintos tipos de residuos que se pueden calcular en un modelo de regresión logística. La distribución de los residuos, bajo la hipótesis nula de que el modelo se ajusta bien a los datos, se puede asumir como una normal de media 0 y varianza estimada menor que 1. A efectos de diagnóstico asumiremos que un residuo es significativamente distinto de 0 si su valor absoluto es mayor que 2.

Residuos de Pearson

- La tabla 4.11 muestra los residuos que son significativos:

No significativos	Significativos
805	19

Tabla 4.11: Cantidad de residuos significativos.

19 de los 824 pacientes tienen residuos de Pearson significativos.

- La tabla 4.12 ordena el conjunto de datos de mayor a menor valor de los residuos, se obtiene lo siguiente (mostrando solo los residuos más altos):

ID paciente	785	12	652	763	507	638
residuos	7,142	5,888	5,888	4,230	3,487	3,298

Tabla 4.12: Pacientes con los residuos de Pearson más altos.

- Caracterización de los pacientes con los residuos más altos:

ID paciente	edad	género	tipo de enfermedad	tipo de fármaco	tipo de receta	frecuencia	dependencia
785	63	mujer	depresión	Clonazepan2Mg	crónica	24	no
12	62	mujer	depresión	Clonazepan0,5Mg	salud mental	22	no
653	62	mujer	depresión	Clonazepan0,5Mg	crónica	22	no
763	67	hombre	depresión	Diazepan10Mg	crónica	16	no
507	48	hombre	Epilepsia-Esquizofrenia	Diazepan10Mg	crónica	22	no
638	15	mujer	Epilepsia-Esquizofrenia	Clonazepan2Mg	crónica	24	no

Tabla 4.13: Características de los pacientes con residuos altos.

Puede verse en la tabla 4.13 que los residuos más altos corresponden a pacientes que no tienen riesgo a ser dependientes. El residuo de Pearson más alto corresponde a una mujer de 63 años la cual padece depresión, consume Clonazepan de 2 Mg, su tipo de receta es crónica

y durante los tres años asistió 24 veces al CESFAM a retirar el fármaco. La característica en común de estos 6 pacientes con residuos altos es que ninguno de ellos tiene riesgo a ser dependiente del fármaco. Por otro lado la covariable frecuencia se mueve en un rango de entre 16 y 24 asistencias a retirar el fármaco entre los años 2013 al 2015.

Residuos de Pearson Estandarizados

- La tabla 4.14 muestra los residuos de Pearson estandarizados, de los cuales son significativos aquellos cuyo valor absoluto sea mayor de 2:

No significativos	Significativos
805	19

Tabla 4.14: Cantidad de residuos de Pearson estandarizados significativos.

19 de los 824 pacientes tienen residuos de Pearson estandarizados significativos.

- La tabla 4.15 muestra el valor de los 6 residuos de Pearson estandarizados más significativos:

ID paciente	12	69	111	191	299	353
residuos	-5,897	-2,411	2,199	-2,246	2,199	-2,894

Tabla 4.15: Pacientes con los residuos más altos.

- Al igual que antes, la tabla 4.16 ordena los valores residuales de mayor a menor y puede verse qué valor tienen en las covariables (mostrando solo los valores más altos):

ID paciente	785	12	652	763	507	638
residuos	7,142	5,888	5,888	4,230	3,487	3,298

Tabla 4.16: Pacientes con los residuos más altos.

- Característica de los pacientes con los residuos de Pearson estandarizados más altos:

ID Paciente	edad	género	tipo de enfermedad	tipo de fármaco	tipo de receta	frecuencia	dependencia
785	63	mujer	depresión	Clonazepan2Mg	crónica	24	no
12	62	mujer	depresión	Clonazepan0,5Mg	salud mental	22	no
653	62	mujer	depresión	Clonazepan0,5Mg	crónica	22	no
763	67	hombre	depresión	Diazepan10Mg	crónica	16	no
507	48	hombre	Epilepsia-Esquizofrenia	Diazepan10Mg	crónica	22	no
638	15	mujer	Epilepsia-Esquizofrenia	Clonazepan2Mg	crónica	24	no

Tabla 4.17: Características de los pacientes con residuos de Pearson estandarizados altos.

Residuos de la Devianza

- La tabla 4.18 muestra los residuos significativos de la Devianza:

No significativos	Significativos
815	9

Tabla 4.18: Residuos significativos de la Devianza.

9 de los 824 pacientes tiene residuos de la Devianza significativos.

Residuos de la devianza estandarizados

- El cuadro 4.19 muestra los Residuos de la devianza estandarizados significativos:

No significativos	Significativos
814	10

Tabla 4.19: Cantidad de pacientes con residuos de Devianza estandarizada significativos.

10 de los 824 pacientes tiene residuos de la devianza estandarizado significativos.

Gráficos de los residuos:

Cuando existe un elevado número de observaciones, como en este caso, es útil la representación gráfica de los residuos. Para verlos de forma clara y así poder analizarlos.

El siguiente gráfico muestra los valores de los residuos de la devianza para cada paciente, las rectas $y = 2$ e $y = -2$ representan los cuantiles 0.025 y 0.975, con lo que, entre esas dos líneas estarían el 95 % de los residuos.

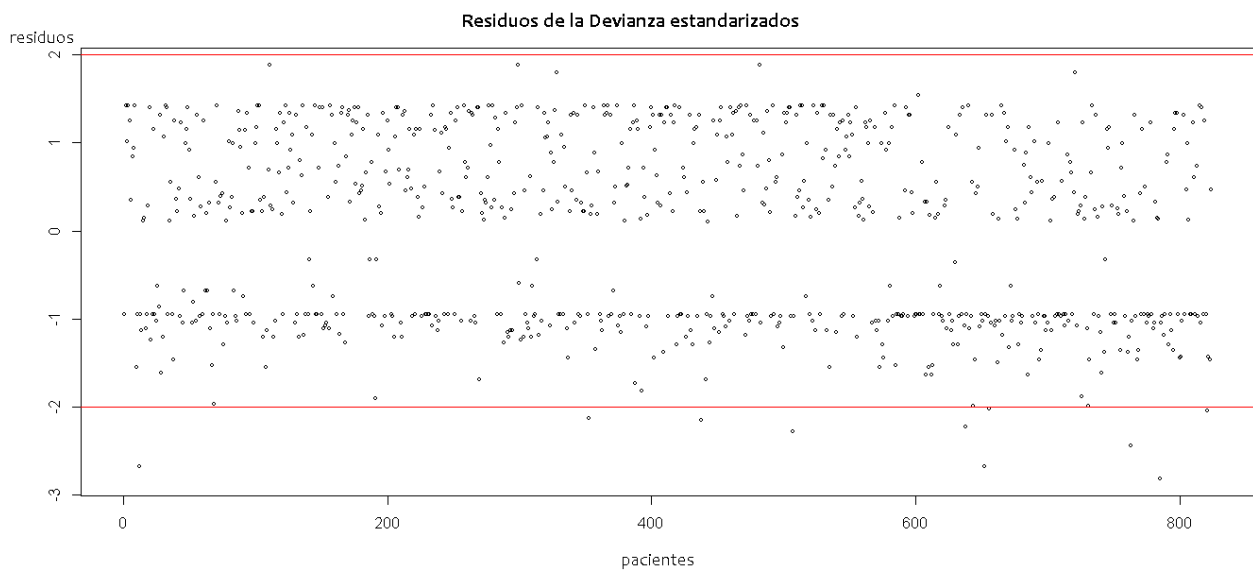


Figura 4.11: Residuos de la Devianza

La figura 4.11 muestra los residuos de los 824 pacientes, donde sólo los que se encuentran fuera de las rectas son mayores que 2 en valor absoluto y eso quiere decir que son significativos. Tal como se muestra en la tabla 4.20 10 de los 824 pacientes presentan alto residuo de la Devianza estandarizado y en la figura 4.11 se pueden ver claramente los 10 que están fuera de la línea roja horizontal.

4.6 Resultados Análisis de Diagnóstico

Medidas de influencia - Distancia de Cook:

- La tabla 4.20 muestra la cantidad de pacientes con Distancia de Cook significativa:

No significativos	Significativos
824	0

Tabla 4.20: Cantidad de pacientes con distancia significativa.

De los 824 pacientes, ninguno tiene una distancia significativa (mayor que uno).

- Representación gráfica:

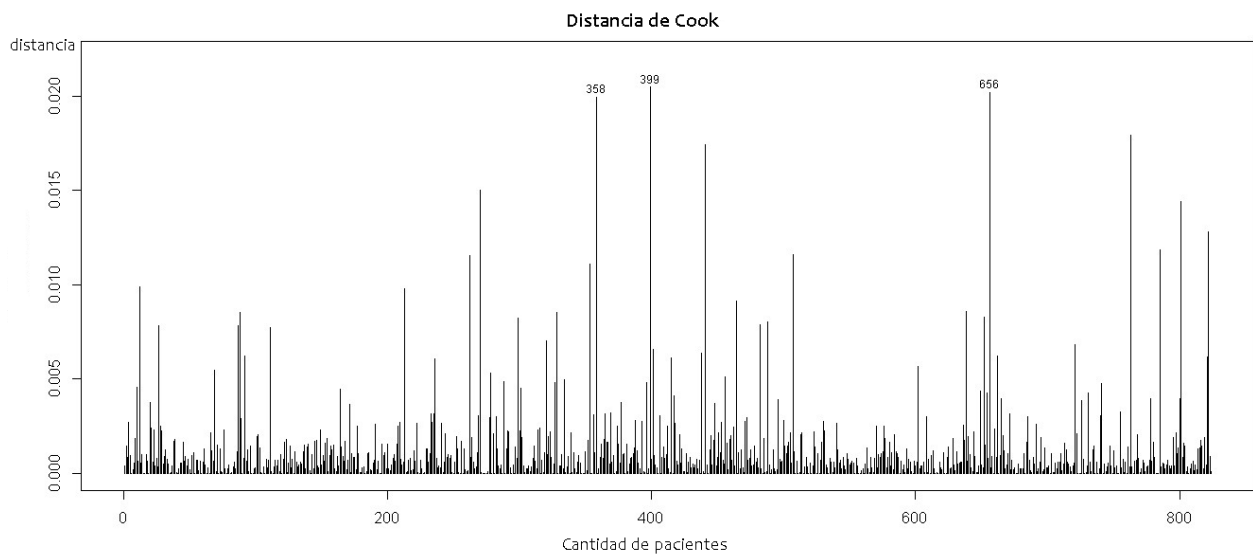


Figura 4.12: Distancia de Cook

Se puede observar que existen tres Distancias de Cook relativamente altas, sin embargo éstas no son significativas, por lo estudiando anteriormente ninguna de ellas es mayor

que 1. Pero sobre salen del resto de los pacientes y se puede ver claramente que los pacientes que pueden ser influyentes son el paciente 358, 399 y 656

- Características de los tres pacientes con alta distancia de Cook:

ID Paciente	edad	género	diagnóstico	artículo	tipo de receta	frecuencia	dependencia
358	31	mujer	depresión	Clonazepan0,5Mg	crónica	1	no
399	66	mujer	depresión	Clonazepan2Mg	salud mental	1	no
656	66	mujer	depresión	Clonazepan0,5Mg	morbilidad	1	no

Tabla 4.21: Características de los pacientes con distancia de Cook más altos

Algunas características similares que tienen los tres pacientes con la mayor distancia de Cook es que son en primer lugar mujeres que poseen como diagnóstico la depresión, en segundo lugar han asistido una vez al CESFAM a retirar el fármaco durante los 3 años y no tienen riesgo a ser dependientes.

Medidas de influencia - Leverage:

Leverage al igual que la Distancia de Cook mencionada anteriormente, muestra los pacientes que pueden ser influyentes para el modelo de regresión logística. La importancia de la gráfica de Leverage es que el dato que esté más alejado de los otros en la línea horizontal, puede ser influenciado en las estimaciones de los parámetros del modelo ajustado.

El gráfico que se muestra a continuación en la figura 4.13 muestra Leverage versus Distancia de Cook, de esta forma se ve claramente los puntos influyentes.

La figura 4.13 muestra las mismas observaciones influyentes, donde se destacan los pacientes 656, 441 y 399. Aquí se encuentran los 824 pacientes como se puede ver en la gráfica la mayoría se centra en el punto cero que significa que estos no son influenciados pero a medida que asciende en torno al 1 se ven algunos puntos dispersos de los cuales el punto con mayor influencia sería el paciente 399, porque se encuentra más cercano a 1 o más alejado en la dirección horizontal.

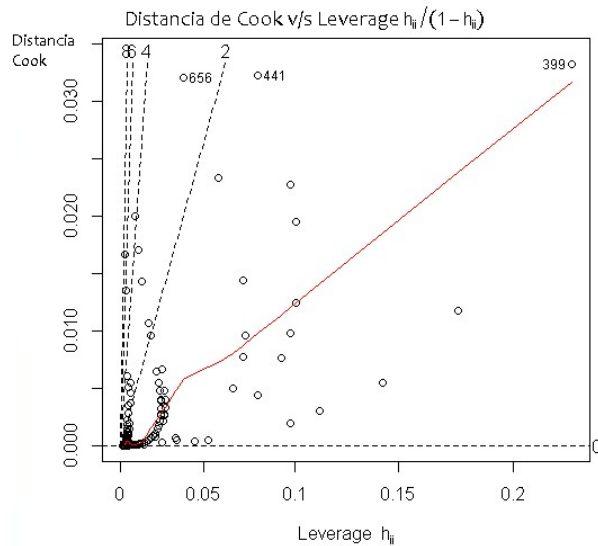


Figura 4.13: Leverage v/s Distancia de Cook

Estos tres pacientes con mayor distancia de Cook y mayor valor en Leverage, fueron eliminados de la base de datos para realizar nuevamente el modelo de regresión logística para ver si hay variación en los datos. El resultado del modelo final fue el siguiente:

Variable	Coefficiente	Estimación	Error estándar	Valor Z	$Pr(> z)$	Odds
	Constante	-0,716	0,151	0,273	0,000	0,489
x_1	frecuencia	0,019	10,539	9,841	0,000	1,216
x_2	Tipo de enfermedad 4	-1,059	0,348	-3,040	0,001	0,347
x_3	Tipo de enfermedad 5	-2,991	1,185	-2,525	0,009	0,050
AIC = 905,03						

Tabla 4.22: Estimación de parámetros del modelo de regresión logística ajustado final.

Tipo de enfermedad 4: Epilepsia, Esquizofrenia, Hipertensión, Tartamudez o Problemas familiares.

Tipo de enfermedad 5: Repetición de la receta.

Como se puede apreciar en la tabla 4.22, las mismas covariables siguen siendo significativas y los coeficientes cambian su valor, pero no de forma significativa. La única gran diferencia que se puede ver, es que en este modelo final el valor del AIC es menor en comparación al modelo que incluye los 824 pacientes. Eso quiere decir que el modelo es mejor.

Por consiguiente el modelo de regresión logística ajustado quedaría expresado de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i$$

4.7 Probabilidad de Ocurrencia

Como se puede percibir, las covariables con mayor significancia y que mejor describen la variable de respuesta son frecuencia y tipo de enfermedad, sin embargo tipo de enfermedad es una covariable categórica dividida en 7 grupos, por ello el conjunto más significativo son las personas que padecen: Epilepsia, Esquizofrenia, Hipertensión, Tartamudez o Problemas familiares (tipo de enfermedad 4).

A continuación se presenta una tabla que muestra la probabilidad exacta que tiene el paciente de sufrir riesgo de dependencia del fármaco según la cantidad de veces que el paciente haya asistido durante los años 2013, 2014 y 2015 y según el diagnóstico que padezca, esto responde al objetivo específico número 4.

También se muestra el gráfico correspondiente a esta probabilidad donde se ve de mejor forma como afecta el riesgo de dependencia en los pacientes que padecen cierto tipo de enfermedad con respecto al número de veces que ellos asisten a la farmacia del CESFAM a retirar el fármaco.

Probabilidad de riesgo de dependencia de los pacientes que padecen: Epilepsia, Esquizofrenia, Hipertensión, Tartamudez o Problemas familiares (tipo de enfermedad 4):

frecuencia	probabilidad
1	0,56
2	0,6
3	0,65
4	0,69
5	0,73
6	0,77
7	0,8
8	0,83
9	0,85
10	0,88
11	0,9
12	0,91
13	0,93
14	0,94
15	0,95
16	0,96
17	0,97
18	0,97
19	0,98
20	0,98
21	0,98
22	0,99
23	0,99
24	0,99
25	0,99
26	0,99
27	0,99
28	0,99
29	0,99
30	0,99

Tabla 4.23: Probabilidad de riesgo de dependencia en los pacientes que padecen el tipo de enfermedad 4.

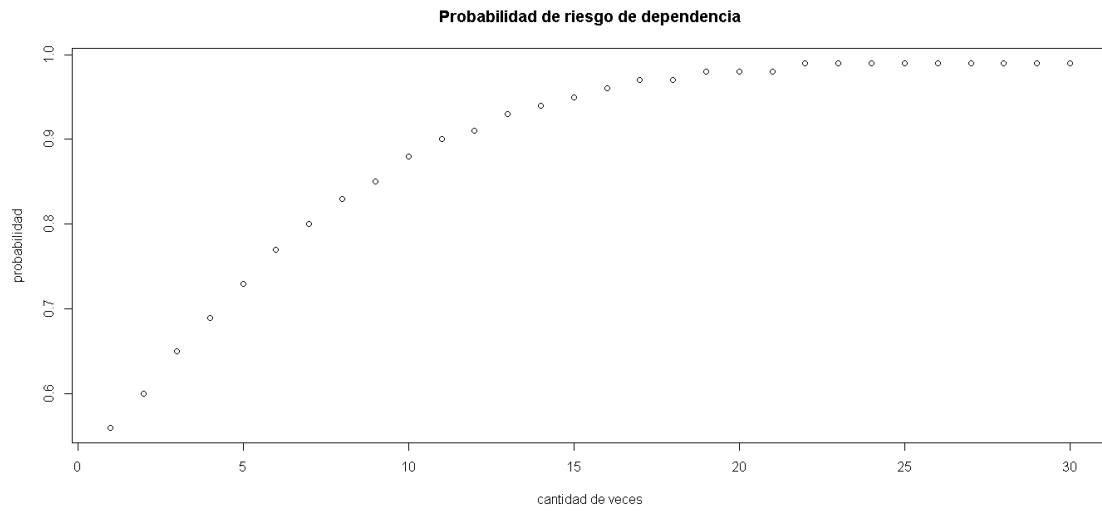


Figura 4.14: Probabilidad de ocurrencia según la variable frecuencia en los pacientes que padecen el tipo de enfermedad 4.

Probabilidad de riesgo de dependencia de los pacientes que no padecen: Epilepsia, Esquizofrenia, Hipertensión, Tartamudez o Problemas familiares, pero sí tienen otro tipo de enfermedad:

frecuencia	Probabilidad
1	0,38
2	0,42
3	0,47
4	0,52
5	0,57
6	0,61
7	0,66
8	0,7
9	0,74
10	0,77
11	0,81
12	0,84
13	0,86
14	0,88
15	0,9
16	0,92
17	0,93
18	0,94
19	0,95
20	0,96
21	0,97
22	0,97
23	0,98
24	0,98
25	0,98
26	0,99
27	0,99
28	0,99
29	0,99
30	0,99

Tabla 4.24: Probabilidad de riesgo de dependencia en los pacientes que padecen otro tipo de enfermedad

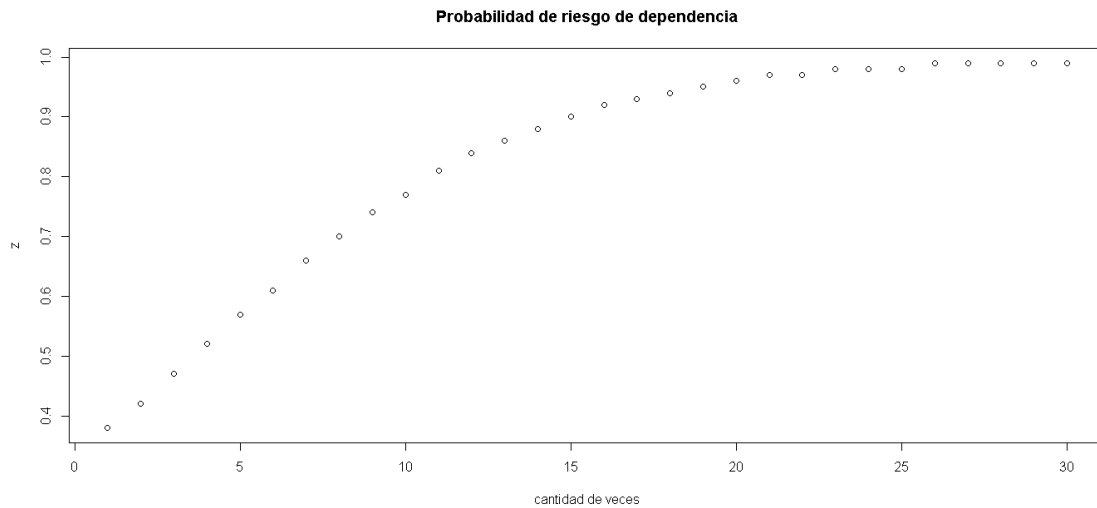


Figura 4.15: Probabilidad de ocurrencia según la variable frecuencia en los pacientes que padecen otro tipo de enfermedad.

Como puede apreciarse en las tablas y en los gráficos, la probabilidad de riesgo de dependencia es más alta si el paciente pertenece al grupo de tipo de enfermedad seleccionado como significativo. También se puede apreciar que mientras más veces el paciente asiste al CESFAM a retirar el fármaco, la probabilidad es muy cercana a 1. Lo cual quiere decir que el riesgo es mayor.

Conclusión

El objetivo de este trabajo fue el de estudiar el análisis de diagnóstico en el modelo de regresión logística, para detectar individuos influyentes se efectuó una evaluación de diagnóstico a través de Leverage y Distancia de Cook, lo cual contribuye a tener una idea sobre que pacientes influyentes distorsionan la estimación de los parámetros del modelo de regresión logística ajustado. Los gráficos considerados para hacer este análisis muestran que las observaciones con datos atípicos que fueron eliminadas del modelo, luego fue ajustado sin considerar las observaciones influyentes, son adecuado para predecir la probabilidad de riesgo de dependencia. Según el Modelo ajustado, las variables más relacionadas con el riesgo de dependencia son: frecuencia y tipo de enfermedad con un alto nivel de significancia.

El énfasis de este trabajo de titulación estuvo reflejado en el marco del proyecto que hubo en la edición de la base de datos y posteriormente de las pruebas que fueron ejecutadas dentro de la misma. A través del uso metodológico se pudo dar cuenta de factores relacionados al consumo de la benzodiazepina, por ejemplo mientras más tiempo se consuma el fármaco, la probabilidad de que el paciente tenga riesgo de dependencia, es mayor. Al igual si lo consume en mayor cantidad, la probabilidad de riesgo aumenta. Los pacientes que consumen benzodiazepinas son aquellos que sufren algún tipo de trastorno o problemas relacionados con el alcohol y las drogas. Está comprobado metodológicamente para efectos de este trabajo que los pacientes que tienen riesgo de ser consumidores de benzodiazepina son aquellos que su tipo de enfermedad está relacionado con Epilepsia, Esquizofrenia, Hipertensión, Tartamudez o Problemas familiares. Puede verse a priori que existen diferencias de sexo en la cantidad de consumo de benzodiazepina por parte de los pacientes de la comuna de Llayllay, siendo las mujeres las mayores consumidoras. Puede pensarse ergo que en la comuna de Llayllay la medicalización no es un asunto ajeno a la realidad cotidiana, ya que gran parte de la población se ve expuesta por intermedio de instituciones de salud pública al acceso a fármacos que poseen un grado de dependencia para ellos. Consecuentemente con ello existe un énfasis en los mecanismos de acceso que tienen los pacientes al fármaco, ya que pudo constatar en este trabajo que gran parte de los pacientes con alto riesgo de dependencia repetían la

prescripción médica. La falta de regulación en la entrega de fármacos siguiendo el conducto regular de la institución posee notorias falencias dentro de su funcionamiento al momento de retirar los fármacos de la farmacia, ya que se entregaban más fármacos de los suscritos. Diferentes elementos constituyen para el CESFMAN de Llayllay focos negativos en vistas del mejoramiento del servicio de salud pública y eventualmente la salud de los asistentes a ella.

Referencias Bibliográficas

Almazán, A., Camarero, L. y Mañas, B. (2000). *“Regresión Logística: Fundamentos y aplicación a la investigación sociológica”*. Departamento de Sociología I, UNED.

Ariza M, Tania, & Granados, Reina. (2012). *“Análisis Bibliométrico de las Revistas Iberoamericanas más Relevantes Afines a la Psicología Clínica y Salud delJournal Citation reports”*. Terapia psicológica, 30 (3), 89-102.

Biggs, D., B. De Ville, and E. Suen (1991). *“A method of choosing multiway partitions for classification and decision trees”*. Journal of Applied Statistics 18 (1), 49?62.

Busto, MA. (1991) *“Benzodiazepinas: La tranquila adicción de Santiago”*. Santiago: CORSAPS.

Cañadas, L. (2013). *“Regresión logística. Tratamiento computacional con R”*. Departamento de Estadística e Investigación Operativa. Facultad de Ciencias. Universidad de Granada.

Cayuela, L. (2015). *“Modelos lineales: Regresión, ANOVA y ANCOVA”*. Area de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Departamental 1 ? DI. 231, c/ Tulipán s/n. E-28933 Móstoles (Madrid), España.

Chatterjee, S., Timothy B., Gangseog Ryu, Michael S. McCarthy, David L. Mothersbaugh, Sandra Milberg, Gary J. Gaeth (2000). *“Asymmetric Competition in Choice and the Leveraging of Competitive Disadvantages”*. Journal of Consumer Research, 27 3 291-308.

D.W. Hosmer. & S. Lemeshow (1982) *“A review of goodness of fit statistics for use in the development of logistic regression models”*. American Journal of Epidemiology, 115:92?106.

Dobson, Annette J. (1945). *“An introduction to generalized linear models”*. University of British Columbia, Canada: Chapman and Hall/CRC, second edition.

Fernandez, T. (2004). *“Introducción al análisis cuantitativo en sociología: Diagnóstico de los residuos en el modelo de regresión lineal”*. COLMEX - CES.

Galton, F. (1886). *“Anthropological miscellanea: Regression towards mediocrity in hereditary stature”*. The Journal of the Anthropological Institute of Great Britain and Ireland, 15:246-263.

Gonzalez M.I., Rojas M. G., Díaz B. (1993) *“Uso de psicofármacos por consultantes a nivel primario (comunicación personal)”*.

Kass, G. V. (1980). *“An exploratory technique for investigating large quantities of categorical data”*. Applied Statistics 29 (2), 119-127.

McCullagh, P. & J.A. Nelder. (1989). *“Generalized Linear Models”*. London New York: Chapman and Hall.

Olivera V, Mauricio. (2009). *“Dependencia a benzodiazepinas en un centro de atención primaria de salud: Magnitud del problema y orientaciones para el manejo integral”*. Revista chilena de neuro-psiquiatría, 47(2), 132-137.

Paula, G. A., Carrasco, J. M., and Ortega, E. M. M. (2008). *“Log-modified Weibull regression models with censored data: sensitivity and residual analysis”*. Computational Statistics and Data Analysis 52, 4021-4039.

Silva, L. C., and I. M. Barroso (2004): *“Regresión Logística. La Muralla”*.

Solano, O, Agustina, R, Bartolo, F, Giraldo, O y Salinas, A. (Agosto 2007). *“Análisis de Diagnóstico en el Modelo de Regresión Logística: Una aplicación”*. PESQUIMAT, N°1, 16.

Truett, J.; Cornfield, J.;Kannel, W. *“A multivariate analysis of the risk of coronary heart disease in Framingham”*. J. Chronic Diseases, v.20, p.511-524, (1967).

Anexo

MINISTERIO DE SALUD
SERVICIO DE SALUD ACONCAGUA
CENTRO DE SALUD FAMILIAR LLAY LLAY
DIRECCION

CERTIFICADO

Quien suscribe certifica que este Centro de Salud ha entregado datos estadísticos de un periodo de 3 años (2013 al 2015) de usuarios que requieren medicamentos psicotrópicos, resguardando la privacidad de datos sensibles de cada paciente, con el fin de incluir este estudio en el trabajo de titulación de la Srta. Barbara Castillo, estudiante de Ingeniería en Estadística de la Universidad de Valparaíso.

Se extiende el presente certificado a petición del interesado, para los fines que estime conveniente.



REGA. PAOLA CARCAMO CASTILLO
DIRECTORA
CENTRO DE SALUD FAMILIAR LLAYLLAY

Llay Llay, 20 de marzo 2017.-

SERVICIO DE SALUD ACONCAGUA
CENTRO DE SALUD FAMILIAR
LLAY-LLAY

Figura 4.16: Consentimiento para trabajar con los datos solicitados.