



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

**UN MODELO DE REGRESIÓN LOGÍSTICA DE
EFECTOS MIXTOS PARA IDENTIFICAR LOS
FACTORES DE RIESGO DE LA SEROPOSITIVIDAD
ANTI-LEPTOSPIRA EN OVEJAS**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

P R E S E N T A:

AXEL MORENO ALFARO



DIRECTOR DE TESIS: DR.GABRIEL ARTEAGA TRONCOSO

2011



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno:
Moreno Alfaro Axel
5512973298
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
2. Tutor:
Dr.
Gabriel Arteaga Troncoso
3. Sinodal 1:
M. en C.
María del Pilar Alonso Reyes
4. Sinodal 2:
M. en E. I. O
Leticia Eugenia Gracia Medrano Valdelamar
5. Sinodal 3:
M. en C.
José Antonio Flores Díaz
6. Sinodal 4:
Dra.
Ruth Selene Fuentes García
7. Tesis
"Un modelo de regresión logística de efectos mixtos para identificar los factores de riesgo de la seropositividad anti-Leptospira en ovejas"
101 p.
2011

AGRADECIMIENTOS Y DEDICATORIAS

Agradezco profundamente y dedico este trabajo a mi mamá y a mi papá por todo el cariño y apoyo que me han dado a través mi vida; por ser un ejemplo de fortaleza y perseverancia. Siempre han creído en mí, me han motivado y apoyado, ustedes son la razón de todo lo que he logrado hasta hoy.

A mi hermana Yojana y mi tío Pedro por haber estado siempre presentes en los momentos buenos y malos que la vida nos ha puesto en el camino.

A Isaura por haber crecido conmigo, por su amor, comprensión y apoyo incondicional. ¡Gracias por ser parte de mí!

A todos los profesores que día a día, a lo largo de tantos años colaboraron con una pieza para que hoy, el rompecabezas de mi formación estuviera completo.

A mis amigos. ¡Gracias por todos los momentos que hemos y seguiremos viviendo juntos!

A mi tutor el Dr. Gabriel Arteaga por su ayuda y comprensión que me proporcionó mientras desarrollábamos este trabajo.

A mis sinodales por su tiempo, dedicación y aportes que me facilitaron la elaboración y corrección de este documento.

Finalmente le agradezco de todo corazón a la UNAM por haberme formado y brindado tantos momentos tan bellos.

ÍNDICE

INTRODUCCIÓN	I
CAPÍTULO 1. MODELOS LINEALES GENERALIZADOS: REGRESIÓN LOGÍSTICA.....	1
1.1.- MODELOS LINEALES CLÁSICOS	1
1.2.-MODELOS LINEALES GENERALIZADOS.....	1
1.2.1.-La familia exponencial.....	2
1.2.2.-Componentes del modelo lineal generalizado	3
1.2.3.-Estimación de β	4
1.2.5.-La devianza	9
Comparación de modelos usando la devianza.....	9
1.2.6.-Residuos.....	10
1.2.7.-Modelos lineales generalizados para datos binarios.....	11
1.3.-REGRESIÓN LOGÍSTICA	14
1.3.1.-El modelo	14
1.3.2.-Estimación de los parámetros.....	16
Algunos problemas en la estimación	17
1.3.3.-Interpretación de los coeficientes.....	19
Variable explicativa dicotómica.....	19
Variable explicativa nominal	20
Variable explicativa continua	22
1.3.4.-Inferencia	23
Intervalo de confianza para β_j	23
Intervalo de confianza para la razón de momios	24
Intervalo de confianza para $\pi(x)$	24
Pruebas de significancia	25
Devianza	26
Estadística χ^2 de Pearson.....	27
La estadística Hosmer-Lemeshow.....	28
CAPÍTULO 2. MODELO LINEAL GENERALIZADO MIXTO.....	29
2.1.-EFECTOS FIJOS	29
2.2.-EFECTOS ALEATORIOS	29
2.3.- ¿FIJO O ALEATORIO?	30
2.4.-MODELO LINEAL GENERALIZADO MIXTO	31
2.4.1.-El modelo	32
2.4.2.-Consecuencias de tener efectos aleatorios	33
2.4.3.-Estimación por máxima verosimilitud	34
2.4.5.-Predicción usando efectos aleatorios.....	36
2.5.-MODELOS MIXTOS PARA DATOS BINARIOS	36
2.5.2.-Ajustando un modelo de regresión logística con un efecto aleatorio.....	38
2.5.3.-Modelos con más de un efecto aleatorio.....	40
2.5.4.- Interpretación de los parámetros.....	41

CAPÍTULO 3. LEPTOSPIROSIS: MARCO TEÓRICO.....	43
3.1.-GENERALIDADES.....	43
3.2.-TAXONOMÍA.....	44
3.3.-RESERVORIOS Y VÍAS DE TRANSMISIÓN DE LEPTOSPIROSIS.....	44
3.4.-PATOGÉNESIS.....	46
3.5.-SIGNOS CLÍNICOS EN OVEJAS Y CABRAS.....	48
3.6.-LAS OVEJAS Y SU RELACIÓN CON OTROS RESERVORIOS.....	49
3.7.-MANIFESTACIONES CLÍNICAS EN HUMANOS.....	50
3.8.-DIAGNÓSTICO DIFERENCIAL.....	51
CAPÍTULO 4. LEPTOSPIROSIS EN OVEJAS: METODOLOGÍA Y DESCRIPCIÓN DEL ESTUDIO.....	53
4.1.-PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN DEL ESTUDIO.....	53
4.2.-OBJETIVOS DEL ESTUDIO.....	55
4.3.-HIPÓTESIS DE INVESTIGACIÓN.....	55
4.4.-POBLACIÓN OBJETIVO.....	55
4.5.-TIPO DE ESTUDIO.....	55
4.6.-LUGAR DE ESTUDIO.....	57
4.8.-REGISTRO DE LOS PARTICIPANTES.....	59
4.9.-SELECCIÓN DE LA MUESTRA.....	59
4.10.-PRUEBAS SEROLÓGICAS.....	60
CAPÍTULO 5. LEPTOSPIROSIS EN OVEJAS: APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA DE EFECTOS MIXTOS PARA LA DETECCIÓN DE FACTORES DE RIESGO.....	61
5.1.-DESCRIPCIÓN DE LA BASE.....	61
5.2.-VARIABLE RESPUESTA O DE INTERÉS.....	66
5.3.-ESTIMACIÓN DE LA SEROPREVALENCIA.....	67
5.3.1.- <i>Determinación de las serovariedades y títulos anti leptospira presentes en los animales.....</i>	<i>68</i>
5.4.-FACTORES DE RIESGO. MODELO DE REGRESIÓN LOGÍSTICA CON EFECTOS MIXTOS.....	69
5.4.1.- <i>Construcción del modelo.....</i>	<i>69</i>
Modelo nulo.....	69
Modelos con una variable explicativa.....	71
Modelos con múltiples variables explicativas.....	72
5.5.-MODELO LOGÍSTICO SIN EFECTO ALEATORIO.....	76
5.5.1.- <i>Bondad de ajuste y verificación del modelo.....</i>	<i>77</i>
5.6.-INTERPRETACIÓN DE LOS COEFICIENTES.....	81
CONCLUSIONES.....	83
DISCUSIÓN.....	84
ANEXO A: ESTIMADOR DE MUESTREO BIETÁPICO ESTRATIFICADO.....	88
ANEXO B: AIC Y BIC.....	90
GLOSARIO DE TÉRMINOS.....	91
BIBLIOGRAFÍA.....	94

INTRODUCCIÓN

La regresión logística ha sido en los últimos años una fuerte herramienta para el estudio de diversos fenómenos en distintas áreas de investigación, debido a su capacidad de incorporar variables con cualquier escala y la gran ventaja de modelar datos de tipo binario o binomial, los cuales se presentan frecuentemente en las diferentes áreas de estudio. El análisis de regresión logística permite cuantificar la relación que existe entre la probabilidad de ocurrencia de un determinado evento de interés y la presencia de uno o varios factores que intervienen en el proceso.

En el campo de la epidemiología la regresión logística ha sido ampliamente usada debido a su relativa facilidad de interpretación. Es en este contexto que se presenta el tema principal de nuestro documento. La leptospirosis es una enfermedad causada por la bacteria *Leptospira interrogans*. La infección se presenta en una gran variedad de especies mamíferas, muchas de ellas domésticas o silvestres causando pérdidas económicas, además de que es una enfermedad transmisible a los humanos. En el Estado de México se cuenta con un considerable inventario ganadero, particularmente ovino explotado bajo las condiciones ambientales que facilitan la transmisión animal-animal y animal-humano de leptospirosis. El impacto de la leptospirosis en la producción animal y su papel como zoonosis es mejor apreciado cuando esta enfermedad es definida por la relación individual en el huésped (especie animal), las serovariedades presentes en su ecosistema o nicho ecológico y los factores predisponentes para reducir el riesgo de la infección a los humanos en esta región (Prescott, 1993; Radostitis y cols., 1999; Levett, 2001).

En el capítulo 1 se describen los modelos lineales generalizados (extensión de los modelos lineales clásicos), familia de modelos a la que pertenece la regresión logística. Se determina la manera en la cual se estiman los parámetros del modelo, así como la interpretación de los mismos. También se definen los métodos de inferencia y de bondad de ajuste. El capítulo 2 define los efectos aleatorios y las diferencias que estos tienen con respecto a los efectos fijos. Se presenta el modelo lineal generalizado mixto, el cual tiene las ventajas de los modelos lineales generales clásicos además de la flexibilidad para tratar datos correlacionados entre sí y se exponen los métodos de ajuste y de inferencia. En el capítulo 3 en un sentido epidemiológico se describen los componentes de la historia natural de la leptospirosis con temas como: taxonomía, reservorios y vías de transmisión, patogénesis, signos clínicos en pequeños rumiantes, ovejas y su relación con otros reservorios, manifestaciones clínicas en humanos y su diagnóstico diferencial. En el capítulo 4 se lleva a cabo la descripción metodológica del trabajo, incluyendo los aspectos de justificación, planteamiento del problema, objetivos, hipótesis y diseño de la muestra. En el capítulo 5 se describe la base de datos utilizada y se estima la prevalencia de anticuerpos anti-*Leptospira interrogans*, así como de la frecuencia y el título serológico de las serovariedades identificadas. También se describe el procedimiento para construir el modelo logístico mixto, el cual tiene la finalidad de encontrar factores relacionados con la seropositividad y así clasificar rebaños

seropositivos a leptospirosis de acuerdo a las variables seleccionadas. La respuesta estudiada (seropositividad) fue determinada a partir de los títulos de anticuerpos a las diferentes serovariedades de *Leptospira interrogans* en el suero extraído de las ovejas y los predictores del modelo fueron obtenidos mediante una encuesta-cuestionario no validada. Un ajuste logístico simple también fue construido y comparado con el modelo final mixto eligiendo al “mejor”, sobre el cual se interpretan los parámetros y se verifica que sea adecuado.

Finalmente se presentan las conclusiones generales del estudio y se da un breve glosario con algunos términos de los capítulos 3 y 4. Todos los análisis estadísticos se realizaron con el paquete STATA (versión 11, College Station, Texas, USA).

CAPÍTULO 1. MODELOS LINEALES GENERALIZADOS: REGRESIÓN LOGÍSTICA

En este capítulo se definen los modelos lineales generalizados (GLM), métodos de estimación, inferencia y medidas de bondad de ajuste. Posteriormente se presenta la regresión logística como miembro de esta familia y se detallan los mismos puntos que en los GLM, además de la interpretación de los parámetros y medidas especiales de bondad de ajuste.

1.1.- Modelos lineales clásicos

Los elementos de un modelo lineal clásico son: El conjunto de observaciones denotado por un vector columna $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$, el cual representa una variable aleatoria; Las covariables o variables explicativas son una matriz $\mathbf{X}_{n \times p}$ donde cada renglón de \mathbf{X} se asocia a una observación y cada columna a una covariable diferente; a cada variable se le asigna un coeficiente o parámetro desconocido; El conjunto de parámetros es un vector de dimensión p , generalmente denotado por $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$. Entonces la relación del vector de observaciones con las covariables es:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \text{ donde } E(\boldsymbol{\epsilon}) = 0, \text{ y por lo tanto } E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

Donde $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$ es el vector aleatorio de errores definido para cada valor de \mathbf{y} , por lo tanto $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$.

Los β_i se estiman minimizando $\boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \sum \epsilon_i^2 = \sum (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^2$ sobre los valores de $\boldsymbol{\beta}$, a este método se le conoce como mínimos cuadrados.

1.2.- Modelos lineales generalizados

Los modelos estadísticos clásicos, como los lineales, son útiles en situaciones donde los datos siguen una distribución normal y que puedan ser explicados linealmente, por lo tanto son relativamente sencillos para interpretar. Sin embargo los supuestos pueden ser bastante restrictivos, además de que los datos pueden ser de naturaleza "no normal". El ajuste de regresiones lineales a variables de respuesta de tipo categórico también puede arrojar resultados erróneos.

Avances en la estadística teórica y la computación permiten usar métodos análogos a los desarrollados por los modelos lineales clásicos en situaciones más generales:

1. La variable de respuesta puede tener una distribución diferente de la normal, ésta puede ser incluso categórica en vez de continua.

2. La relación entre la respuesta y los factores no necesita ser de forma lineal simple. Los modelos lineales generalizados son una extensión de los modelos lineales clásicos, introducidos por Nelder y Wedderburn (Dobson, 2005).

1.2.1.-La familia exponencial

Las respuestas y_i no necesariamente son gaussianas, sino que su distribución puede ser cualquiera de la familia exponencial y se puede escribir de la siguiente manera:

$$f(y_i; \theta_i) = s(y_i)t(\theta_i)e^{a(y_i)b(\theta_i)} \dots (1.1)$$

Donde a , b , s y t son funciones conocidas. La ecuación anterior puede ser escrita como:

$$f(y; \theta_i) = \exp[a(y_i)b(\theta_i) + c(\theta_i) + d(y_i)] \dots (1.2)$$

Donde $s(y_i) = \exp d(y_i)$ y $t(\theta_i) = \exp c(\theta_i)$.

Si $a(y_i) = y_i$, se dice que la distribución está en la forma canónica y $b(\theta)$ es llamado el parámetro natural de la distribución.

Si hay otros parámetros, además del de interés θ , se consideran como estorbo o ruido, formando parte de las funciones a , b , c y d y son tratados como si fueran conocidos.

Muchas de las funciones de distribución conocidas pertenecen a la familia exponencial. Por ejemplo, las distribuciones Poisson, normal y binomial pueden ser escritas en la forma canónica.

Tabla 1.1.- Algunas distribuciones pertenecientes a la familia exponencial

Distribución	Parámetro natural	c	d
Poisson	$\log \theta$	$-\theta$	$-\log y!$
Normal	$\frac{\mu}{\sigma^2}$	$-\frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2)$	$-\frac{y^2}{2\sigma^2}$
Binomial	$\log\left(\frac{\pi}{1-\pi}\right)$	$n \log(1-\pi)$	$\log \binom{n}{y}$

1.2.2.-Componentes del modelo lineal generalizado

Un modelo lineal generalizado consta de tres componentes:

1. Aleatorio: Es un vector de observaciones \mathbf{y} con n entradas y se supone que es realización de una variable aleatoria \mathbf{Y} cuyos elementos son distribuidos independientemente y pertenecen a la familia exponencial con media $\boldsymbol{\mu}$, es decir:

$$E(y_i) = \mu_i$$

2. Sistemático: La parte sistemática del modelo relaciona un vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)$ con un conjunto de variables predictoras mediante una relación lineal:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

Donde:

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Se le denomina la matriz de diseño y contiene los valores de las p variables explicativas para las n observaciones y $\boldsymbol{\beta}$ es un vector de parámetros desconocidos.

3. Función liga: Es una función $g(\cdot)$ monótona diferenciable que relaciona los componentes aleatorio y sistemático:

$$\eta_i = g(\mu_i) = \sum_{j=1}^p x_{ij}\beta_j \dots (1.3)$$

Para algunas distribuciones existen funciones liga especiales llamadas canónicas, éstas se dan cuando el parámetro natural es igual al predictor lineal. En estos casos se puede encontrar una estadística suficiente de igual dimensión que $\boldsymbol{\beta}$ en $\eta_i = \sum x_{ij} \beta_j$. Entonces:

$$b(\theta) = \eta \Rightarrow \mathbf{X}^T \mathbf{y}$$

Es una estadística suficiente para β .

La razón para usar ligas canónicas es que conducen a que el modelo tenga propiedades estadísticas deseables, particularmente en muestras pequeñas, aunque no existe ninguna otra razón para usar este tipo de ligas.

1.2.3.-Estimación de β

El método usual para estimar los parámetros en un modelo lineal generalizado es estimación por máxima verosimilitud. Aunque expresiones matemáticas explícitas pueden ser encontradas para algunos casos especiales, regularmente se necesitan métodos numéricos, los cuales son iterativos y están basados en el algoritmo Newton-Raphson (Dobson, 2005).

Considérense las variables independientes Y_1, Y_2, \dots, Y_n que satisfacen las propiedades de un modelo lineal generalizado. Se desea estimar los parámetros β .

Para cada Y_i la función de log-verosimilitud es:

$$l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i) \dots (1.4)$$

En estos modelos, las siguientes expresiones pueden encontrarse (Dobson, 2005):

$$E(Y_i) = \mu_i = \frac{-c'(\theta_i)}{b'(\theta_i)} \dots (1.5)$$

$$\text{var}(Y_i) = \frac{[b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)]}{[b'(\theta_i)]^3} \dots (1.6)$$

$$g(\mu_i) = \sum_{j=1}^p x_{ij} \beta_j = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i \dots (1.7)$$

La función de log-verosimilitud para todas las Y_i es:

$$l = \sum_{i=1}^n l_i = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i) \dots (1.8)$$

Para obtener el estimador máximo verosímil de β_j se necesita encontrar:

$$U_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^n \left[\frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j} \right] \dots (1.9)$$

Donde U_j es llamada *estadística de puntuación* para el j -ésimo parámetro (Dobson, 2005).

Encontrando cada derivada parcial por separado se tiene:

Diferenciando (1.4) y sustituyendo (1.5)

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)(y_i - \mu_i) \dots (1.10)$$

Además se cumple que

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\left(\frac{\partial \mu_i}{\partial \theta_i} \right)}$$

Diferenciando (1.5), y por (1.6)

$$\frac{\partial \mu_i}{\partial \theta_i} = \frac{-c''(\theta_i)}{b'(\theta_i)} + \frac{c'(\theta_i)b''(\theta_i)}{[b'(\theta_i)]^2} = b'(\theta_i)\text{var}(Y_i)$$

$$\therefore \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b'(\theta_i)\text{var}(Y_i)} \dots (1.11)$$

Finalmente por (1.7)

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \dots (1.12)$$

Por consiguiente la estadística de puntuación es:

$$U_j = \sum_{i=1}^n \left[\frac{y_i - \mu_i}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \dots (1.13)$$

La matriz de varianzas covarianzas de las U_j , conocida como la *matriz de información* \mathfrak{S} y que tiene entradas:

$$\mathfrak{S}_{jk} = E[U_j U_k] \dots (1.14)$$

$$\begin{aligned} \mathfrak{S}_{jk} &= E \left\{ \sum_{i=1}^n \left[\frac{Y_i - \mu_i}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \sum_{m=1}^n \left[\frac{Y_m - \mu_m}{\text{var}(Y_m)} x_{mk} \left(\frac{\partial \mu_m}{\partial \eta_m} \right) \right] \right\} \\ &= \sum_{i=1}^n \left[\frac{E[(Y_i - \mu_i)^2] x_{ij} x_{ik} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2}{[\text{var}(Y_i)]^2} \right] \dots (1.15) \end{aligned}$$

Debido a que $E[(Y_i - \mu_i)(Y_m - \mu_m)] = 0$ para toda $i \neq m$ ya que las Y_i son independientes. Si se usa que $\text{var}(Y_i) = E[(Y_i - \mu_i)^2]$, 1.15 puede ser simplificado a:

$$\mathfrak{S}_{jk} = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \dots (1.16)$$

El método basado en el algoritmo de Newton-Raphson llamado *la puntuación de Fisher* plantea que para encontrar la solución buscada hay que resolver la ecuación iterativa:

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + [\mathfrak{S}^{(m-1)}]^{-1} \mathbf{U}^{(m-1)} \dots (1.17)$$

Donde $\mathbf{b}^{(m)}$ es el vector de estimaciones de los parámetros β_1, \dots, β_j en la m -ésima iteración, $[\mathfrak{S}^{(m-1)}]^{-1}$ es la inversa de la matriz de información y $\mathbf{U}^{(m-1)}$ es el vector compuesto por las estadísticas de puntuación en la $m - 1$ -ésima iteración. La ecuación 1.17 puede ser escrita como:

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{b}^{(m)} = \mathbf{X}^T \mathbf{W} \mathbf{z} \dots (1.18)$$

Donde \mathbf{W} es una matriz diagonal $n \times n$ con elementos

$$w_{ii} = \frac{1}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

y \mathbf{z} tiene elementos

$$z_i = \sum_{k=1}^p x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)$$

La mayoría de los programas estadísticos incluyen un algoritmo basado en (1.17). Empiezan usando una aproximación inicial $\mathbf{b}^{(0)}$ para evaluar \mathbf{z} y \mathbf{W} , entonces (1.17) es resuelta para dar $\mathbf{b}^{(1)}$, el cual es usado a su vez para calcular mejores aproximaciones de \mathbf{z} y \mathbf{W} , y así hasta que la diferencia entre $\mathbf{b}^{(m-1)}$ y $\mathbf{b}^{(m)}$ es suficientemente pequeña, entonces $\mathbf{b}^{(m)}$ es tomado como el estimador máximo verosímil de $\boldsymbol{\beta}$ (Dobson, 2005).

1.2.4.-Inferencia estadística y verificación del modelo

Inferencia acerca de los parámetros

Las dos principales herramientas de la inferencia estadística son intervalos de confianza y pruebas de hipótesis.

Para calcular un intervalo de confianza para los parámetros, se puede utilizar el hecho de que la distribución asintótica de los estimadores máximos verosímiles es aproximadamente normal. Si se denota $\hat{\beta}_j$ como el estimador máximo verosímil de β_j entonces:

$$\hat{\beta}_j \sim N(\beta_j, \mathfrak{S}_{jj}^{-1}) \dots (1.19)$$

Donde \mathfrak{S}_{jj}^{-1} es el j -ésimo elemento de la diagonal de \mathfrak{S}^{-1} . La inversa de la matriz de información (\mathfrak{S}) para muestras grandes, contiene las varianzas y covarianzas de los elementos de $\hat{\beta}$. Entonces un intervalo al $(1 - \alpha)100\%$ de confianza para β_j es:

$$\hat{\beta}_j \pm z_{1-\alpha/2} \sqrt{\mathfrak{S}_{jj}^{-1}} \dots (1.20)$$

El intervalo 1.20 también es conocido como intervalo de Wald al $(1 - \alpha)100\%$ de confianza.

De la misma manera se puede probar $H_0: \beta_j = 0$ con la estadística de prueba Wald:

$$z = \frac{\hat{\beta}_j}{SE}$$

Donde $SE = \sqrt{\mathfrak{S}_{jj}^{-1}}$ es el error estándar de $\hat{\beta}_j$. Esta estadística tiene aproximadamente una distribución normal estándar cuando $\beta_j = 0$.

También se puede comprobar lo anterior mediante la prueba de cociente de verosimilitudes. Si se denota el valor maximizado de la función de verosimilitud por ℓ_0 bajo $H_0: \beta_j = 0$ y por ℓ_1 cuando β_j no es necesariamente 0. La estadística de prueba del cociente de verosimilitudes es:

$$-2 \log \left(\frac{\ell_0}{\ell_1} \right) = -2[\log(\ell_0) - \log(\ell_1)] = -2[L_0 - L_1] \dots (1.21)$$

Donde L_0 y L_1 denotan la función log-verosimilitud maximizada. Bajo $H_0: \beta_j = 0$ esta estadística de prueba tiene una distribución ji-cuadrada con 1 grado de libertad.

El procedimiento de cociente de verosimilitudes también puede determinar un intervalo del $(1 - \alpha)100\%$ de confianza para β_j . Éste consiste en todos los valores β_0 para los cuales el p -value (valor p) es mayor que α en la prueba $H_0: \beta = \beta_0$. Para muestras pequeñas, este intervalo es preferible que el propuesto en 1.20.

1.2.5.-La devianza

Sean L_M y L_S los valores maximizados de las funciones de log-verosimilitud de los modelos M (el de interés) y el saturado (S). Éste último tiene un parámetro para cada observación y por lo tanto provee un ajuste perfecto para los datos.

Debido a que S tiene más parámetros, L_S es al menos tan grande como la log-verosimilitud maximizada L_M , es decir $L_S \geq L_M$. La *devianza* de un modelo lineal generalizado se define como:

$$D = \text{Devianza} = -2[L_M - L_S] \dots (1.22)$$

La devianza es la estadística de cociente de verosimilitudes para comparar M con S . Ésta es una prueba estadística para la hipótesis de que todos los parámetros que están en el modelo saturado, pero no en el modelo M , son cero.

Si la variable de respuesta Y_i tiene distribución normal, entonces D se distribuye exactamente como una Ji-cuadrada, sin embargo D depende de $\text{var}(Y_i) = \sigma^2$, la cual en la práctica es usualmente desconocida. Para Y_i no normal, la distribución de muestreo de D solamente puede ser Ji-cuadrada aproximada.

La devianza provee una medida de bondad de ajuste, debido a que ésta, prueba la hipótesis de que todos los parámetros no incluidos en el modelo son cero. Los grados de libertad son iguales al número de observaciones menos la cantidad de parámetros estimados. El *valor-p* es la cola derecha de probabilidad por encima de la estadística de prueba observada (1.22), de una distribución Ji-cuadrada. Valores grandes de D y/o *valores-p* pequeños brindan una fuerte evidencia de falta de ajuste del modelo.

Comparación de modelos usando la devianza

Ahora considérense dos modelos, denotados por M_0 y M_1 , tal que M_0 es un caso especial de M_1 , es decir, tienen la misma distribución de probabilidad y la misma función liga pero el componente lineal de M_0 es un caso especial del de M_1 .

Considérese la hipótesis nula

$$H_0: \boldsymbol{\beta} = \boldsymbol{\beta}_0 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_q \end{bmatrix}$$

correspondiente al modelo M_0 y una hipótesis más general

$$H_1: \boldsymbol{\beta} = \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

correspondiente a M_1 , con $q < p < N$.

Se puede probar H_0 vs. H_1 usando la diferencia entre las devianzas

$$\Delta D = D_0 - D_1 = -2[L_0 - L_S] - \{-2[L_1 - L_S]\} = -2[L_0 - L_1] \dots \dots (1.23)$$

Si ambos modelos describen bien los datos entonces $D_0 \sim \chi^2(n - q)$ y $D_1 \sim \chi^2(n - p)$ de modo que $\Delta D \sim \chi^2(p - q)$ siempre que la condición de independencia se mantenga. Si el valor de ΔD fuera consistente con la distribución $\chi^2(p - q)$, generalmente se escogería M_0 correspondiente a H_0 porque es más simple.

Si el valor de ΔD estuviera en la región crítica (es decir más grande que el cuantil $(1-\alpha)$ de una distribución $\chi^2(p - q)$) entonces se rechazaría H_0 en favor de H_1 con base en que el modelo M_1 provee una mejor descripción de los datos.

1.2.6.-Residuos

Para cualquier GLM, la estadística de bondad de ajuste sólo resume en general qué tan bien el modelo ajusta a los datos. Se obtendría una mejor comprensión si se compararan individualmente los valores observados y los ajustados. A continuación se presentan tres tipos de residuos.

El *residual de Pearson* definido como:

$$\text{Residual de Pearson} = e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{var}(y_i)}} \dots \dots (1.24)$$

El *residual estandarizado* toma $(y_i - \hat{\mu}_i)$ y lo divide entre el error estándar estimado de e_i , esto es:

$$\text{Residual estandarizado} = \frac{y_i - \hat{\mu}_i}{SE(e_i)} \dots (1.25)$$

Si la devianza es usada como una medida de discrepancia de un modelo lineal generalizado, entonces cada unidad contribuye una cantidad d_i a la medida, de modo que $\sum d_i = D$. Entonces se define:

$$\text{Residual de la devianza} = r_D = \text{sign}(y - \mu) * \sqrt{d_i} \dots (1.26) \quad ^1$$

Para los cuales se cumple que $\sum r_D^2 = D$ (Agresti 2002, Hosmer 1989, McCullagh 1983, Collet 2003).

1.2.7.-Modelos lineales generalizados para datos binarios

Muchas variables tienen sólo dos posibles categorías, por ejemplo, si hoy se usó o no el transporte público, si se tuvo o no un examen físico el año pasado, en general si se tiene éxito o fracaso.

Primero se define la variable aleatoria binaria:

$$Z = \begin{cases} 1 & \text{se tiene un éxito} \\ 0 & \text{en otro caso} \end{cases}$$

Con probabilidades $P(Z = 1) = \pi$ y $P(Z = 0) = 1 - \pi$. Si se tienen n de tales variables Z_1, \dots, Z_n independientes con $P(Z_j = 1) = \pi_j$, entonces su función de densidad conjunta de probabilidad es:

$$\prod_{j=1}^n \pi_j^{z_j} (1 - \pi_j)^{1-z_j} = \exp \left[\sum_{j=1}^n z_j \log \left(\frac{\pi_j}{1 - \pi_j} \right) + \sum_{j=1}^n \log (1 - \pi_j) \right] \dots (1.27)$$

La cual es miembro de la familia exponencial. Ahora para el caso en que las π_j sean las mismas, se define:

¹ $\text{sign}(y - \mu)$ es el signo de la diferencia entre paréntesis.

$$Y = \sum_{j=1}^n Z_j$$

Por consiguiente Y es el número de éxitos en n ensayos, y de aquí que Y siga una distribución binomial

$$P(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} \quad , \quad y = 0, 1 \dots n \dots \dots (1.28)$$

Finalmente, considérese el caso general de N variables independientes Y_1, \dots, Y_N correspondientes al número de éxitos en N diferentes subgrupos o estratos, por lo tanto $Y_i \sim \text{binomial}(n_i, \pi_i)$.

Se quiere describir la proporción de éxitos $P_i = Y_i/n_i$, en cada grupo en términos de variables explicativas que caracterizan al subgrupo. Debido a que $E(Y_i) = n_i \pi_i$, se tiene que $E(P_i) = \pi_i$, entonces se puede modelar las probabilidades π_i en términos de un modelo lineal general como:

$$g(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

El caso más simple se da cuando la función g es la función identidad, es decir

$$\pi = \mathbf{x}\boldsymbol{\beta}$$

Éste es llamado *modelo de probabilidad lineal* porque la posibilidad de un éxito cambia linealmente en \mathbf{x} . El ajuste es simple, pero desafortunadamente tiene un defecto estructural. Las probabilidades toman valores entre 0 y 1, mientras que el dominio de las funciones lineales es toda la recta real, de esta puede predecir $\pi < 0$ o $\pi > 1$, lo cual contradice la teoría de la probabilidad.

Para asegurarse que π esté en el intervalo $[0,1]$ se puede modelarla usando una distribución de probabilidad acumulada:

$$\pi = \int_{-\infty}^t f(s) ds$$

Donde $f(s)$ es una función de densidad (es decir $f(s) \geq 0$ y $\int_{-\infty}^{\infty} f(s) ds = 1$) y es llamada distribución de tolerancia.

Si se usa la distribución normal como distribución de tolerancia, se le conoce como *modelo probit* (Agresti, 2002).

$$\pi = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2\right] ds = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Donde Φ la función de probabilidad acumulada para la distribución normal estándar.

Éste es para el caso de una sola variable explicativa x , sin embargo se puede generalizar. La función $\text{ligag}(\cdot)$ es la inversa de la función de probabilidad acumulada normal, esto es:

$$\Phi^{-1} = \beta_1 + \beta_2 x$$

Otro modelo que da resultados numéricos semejantes al probit, aunque computacionalmente más sencillo, es el logístico o logit, donde la distribución de tolerancia es:

$$f(s) = \frac{\beta_2 \exp(\beta_1 + \beta_2 s)}{[1 + \exp(\beta_1 + \beta_2 s)]^2}$$

Por lo tanto:

$$\pi = \int_{-\infty}^x f(s) ds = \frac{\exp(\beta_1 + \beta_2 x)}{1 + \exp(\beta_1 + \beta_2 x)}$$

El modelo logístico es el más popular para trabajar con datos binarios y es fundamental para el desarrollo de este trabajo, motivo por el cual se pondrá mayor énfasis en su desarrollo.

1.3.-Regresión logística

1.3.1.-El modelo

Los datos binarios son la forma más común de datos categóricos. En la sección anterior se dio una breve introducción a los modelos lineales generalizados que aceptan datos de este tipo. La regresión logística es la más usada de los GLM.

Los modelos de regresión se han convertido en una herramienta fundamental para el análisis estadístico donde el principal objetivo es describir la relación entre una variable de respuesta y un conjunto de variables explicativas.

Los modelos de regresión clásicos son aquellos donde se trata de explicar una variable de respuesta, que generalmente es de tipo continuo, sin embargo, existen casos donde tal variable es de tipo discreto, tomando dos o más posibles valores. Ejemplos de este tipo de datos pueden ser sí-no, verdadero-falso, 0-1, presencia-ausencia de cierta característica, etc. Para este caso la regresión logística permite discriminar entre dos poblaciones, en términos de un conjunto de variables, en el papel de predictores. También para representar probabilidades de ocurrencia de un evento, como función de una serie de variables explicativas. Estos modelos permiten analizar por ejemplo la relación entre edad y la probabilidad de presencia de cierta enfermedad del corazón.

Existen varias razones por las que es más conveniente usar el modelo logístico, es decir usar la función de probabilidad acumulada de la distribución logística para modelar π_i , las más importantes son:

- 1.- Satisface que $0 < \pi < 1$, evitando así las inconsistencias que produce un modelo lineal clásico.
- 2.- La interpretación se puede dar de manera intuitiva como el logaritmo de los momios, los cuales se definirán más adelante.
- 3.- Es computacionalmente más sencilla que, por ejemplo, el modelo probit.
- 4.- Los datos binarios pueden ser resumidos mediante estadísticas suficientes cuando la transformación logística es usada.
- 5.- Las variables explicativas pueden ser de carácter dicotómico, ordinal o nominal.

Supóngase que se tienen n observaciones binomiales de la forma $P_i = y_i/n_i$, para $i = 1, 2, \dots, n$, donde el valor esperado de la variable aleatoria asociada con la i -ésima observación, y_i , está dada por $E(Y_i) = n_i\pi_i$, donde π_i es la correspondiente probabilidad de respuesta.

Entonces de acuerdo a la sección anterior se puede modelar la relación entre π_i y p variables explicativas por medio de la distribución de probabilidad acumulada de la función logística, esto es:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}} \dots (1.29)$$

Y por lo tanto

$$1 - \pi_i = \frac{1}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}$$

La transformación *logit* está definida como:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \dots (1.30)$$

Donde el término:

$$\frac{\pi_i}{1 - \pi_i}$$

es llamado *odds* de la ocurrencia del evento y es la razón entre la probabilidad de ocurrencia del evento entre la probabilidad de no ocurrencia; indica cuanto más probable es que ocurra el evento a que no ocurra. Entonces:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \log\left(\frac{e^{\eta_i}/(1 + e^{\eta_i})}{1/(1 + e^{\eta_i})}\right) = \log(e^{\eta_i}) = \eta_i \dots (1.31)$$

Se tiene que la función $\text{logit}(\pi_i)$ es la liga del modelo, además es igual al parámetro natural, por lo tanto se tiene que es una liga canónica, entonces se pueden encontrar estadísticas suficientes para los parámetros como se señalaba en las ventajas de usar este modelo.

1.3.2.-Estimación de los parámetros

Supóngase que los datos binomiales de la forma y_i éxitos en n_i , $i = 1, 2, \dots, n$, ensayos están disponibles, donde la transformación logística de la correspondiente probabilidad π_i , es modelada como una combinación lineal de los valores de las variables explicativas X_1, \dots, X_p , entonces:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

Se necesitan estimar los $p+1$ parámetros. Esto se hace por el método de máxima verosimilitud, la cual es

$$L(\beta) = \prod_{i=1}^n \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \dots \dots (1.32)$$

Esta verosimilitud depende de la probabilidad de éxito π_i , que a la vez está relacionada con los parámetros a través de (1.29), entonces puede ser considerada como función de β . El problema es entonces encontrar los valores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ que maximicen $L(\beta)$, o equivalentemente $\log L(\beta)$, por lo tanto

$$\begin{aligned} \log L(\beta) &= \sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} + y_i \log \pi_i + (n_i - y_i) \log(1 - \pi_i) \right\} \\ &= \sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} + y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + n_i \log(1 - \pi_i) \right\} \\ &= \sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} + y_i \eta_i - n_i \log(1 + e^{\eta_i}) \right\} \dots \dots (1.33) \end{aligned}$$

Donde $\eta_i = \sum_{j=0}^p \beta_j x_{ij}$ y $x_{i0} = 1$ para todos los valores de i . La derivada de esta función de log-verosimilitud respecto a los $p+1$ parámetros desconocidos β_j es:

$$U_j = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n n_i x_{ij} e^{\eta_i} (1 + e^{\eta_i})^{-1}, \quad j = 0, 1, \dots, p \dots (1.34)$$

Igualando U_j a cero queda un sistema de $p+1$ ecuaciones no lineales de los parámetros desconocidos y que sólo pueden ser resueltas numéricamente. La mayoría de los paquetes utilizan la alternativa mencionada en (1.2.3) y que resuelve de manera iterativa la ecuación:

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + [\boldsymbol{\mathfrak{J}}^{(m-1)}]^{-1} \mathbf{U}^{(m-1)}$$

Donde $\mathbf{U}^{(m-1)}$ es el vector formado por U_j $j = 0, 1, \dots, p$ en la $(m-1)$ -ésima iteración, y siguiendo la forma de (1.16) para el caso particular del modelo logístico se tiene que la matriz de información $\boldsymbol{\mathfrak{J}}^{(m-1)}$ tiene elementos

$$\sum_{i=1}^n x_{ij} x_{ik} n_i \pi_i (1 - \pi_i) \dots (1.35)$$

Evaluados en $\hat{\beta}_j$ $j = 0, 1 \dots, p$ en la $(m-1)$ -ésima iteración.

Algunos problemas en la estimación

Un problema que puede presentarse en la estimación del modelo de regresión logística, es la no convergencia del algoritmo de maximización de la verosimilitud. En la mayoría de los casos esta falla es consecuencia de los patrones de datos llamados separación completa (o predicción perfecta) y casi separación completa (o predicción casi perfecta).

La separación completa ocurre cuando la variable de respuesta separa a una variable explicativa o a una combinación de ellas. Como ejemplo considérense los siguientes datos:

Y	X_1	X_2
0	1	3
0	2	2
0	3	-1
0	3	-1
1	5	2
1	6	4
1	10	1
1	11	0

En este ejemplo Y es la variable respuesta, X_1 y X_2 variables explicativas. Se nota que todas las observaciones con $Y = 0$ tienen valores de X_1 menores o iguales a tres y todos los datos con $Y = 1$ tienen valores de X_1 mayores a tres. Otra manera de ver esto es que X_1 predice perfectamente a Y , ya que $X_1 \leq 3$ corresponde a $Y = 0$ y $X_1 > 3$ corresponde a $Y = 1$. En términos de las probabilidades que se predicen, $P(Y = 1 | X_1 \leq 3) = 0$ y $P(Y = 1 | X_1 > 3) = 1$, sin la necesidad de estimar un modelo.

La predicción perfecta usualmente ocurre cuando hay un error de parte de la persona que codifica los datos. Probablemente se ha incluido otra versión de la variable respuesta como variable explicativa. Otro posible escenario para que ocurra este problema, es cuando el tamaño de muestra es muy pequeño. En el ejemplo anterior no hay razón para que Y tenga que ser cero cuando $X_1 \leq 3$. Si el tamaño de muestra fuera lo suficientemente grande probablemente se hubieran encontrado algunas observaciones con $Y = 1$ y $X_1 \leq 3$. La predicción casi perfecta ocurre cuando Y se puede predecir perfectamente para casi todos los valores de X_1 .

Para estos patrones, las estimaciones máximo verosímiles no existen. Debido a que existe predicción perfecta, mientras mayor sea el valor del coeficiente estimado mayor será el valor de la verosimilitud (Bruin, J. 2006, Paul D., Allison).

Algunas formas de tratar estos problemas son:

- En el caso de separación completa, asegurarse que no se está usando otra versión de la variable respuesta como variable explicativa.
- Si existe predicción casi completa, la mejor estrategia es “no hacer nada” o eliminar las categorías de la variable explicativa con las que se puede predecir Y sin error. Esto se debe a que la verosimilitud maximizada sí existe para las otras variables. Esto no ocurre con la predicción perfecta.
- No incluir a X en el modelo.
- Se pueden colapsar algunas categorías de X , si es que esto tiene sentido.

- Aumentar el tamaño de muestra o utilizar métodos exactos (Bruin, J. 2006, Paul D., Allison).

1.3.3.-Interpretación de los coeficientes

Para efectos de esta sección supóngase el modelo más simple donde sólo se tiene una variable explicativa y que Y es una variable binaria (1,0). Se dará una interpretación para cada una de las posibilidades que se tengan para la variable predictora, ya sea dicotómica, nominal o continua.

Variable explicativa dicotómica

El primer caso es la interpretación del parámetro del modelo cuando la variable explicativa es dicotómica. Asíumase que la variable x está codificada como cero o uno. Bajo estas condiciones sólo se tienen las posibilidades mostradas en la siguiente tabla.

Tabla 1.2.- Valores del modelo de regresión logística cuando x es dicotómica

	$x = 1$	$x = 0$
$y = 1$	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
$y = 0$	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$

Los momios del resultado donde la respuesta es un éxito ($Y=1$) entre los individuos con $x=1$ es $\pi(1)/1 - \pi(1)$. Equivalentemente para $x=0$ el momio está definido como $\pi(0)/1 - \pi(0)$. El logaritmo de los momios como se definió en (1.30) es llamado logit, por ejemplo

$$\text{logit}(1) = \log\left(\frac{\pi(1)}{1 - \pi(1)}\right)$$

La razón de momios, denotada por ψ , se define como la razón de los momios cuando $x = 1$ y $x = 0$

$$\psi = \frac{\pi(1)/1 - \pi(1)}{\pi(0)/1 - \pi(0)} \dots (1.36)$$

Por lo tanto el logaritmo de ψ es la diferencias de los logit

$$\log(\psi) = \log \left[\frac{\pi(1)/1 - \pi(1)}{\pi(0)/1 - \pi(0)} \right] = \text{logit}(1) - \text{logit}(0) \dots (1.37)$$

Si se sustituyen los valores de la tabla (1.2) se obtiene:

$$\psi = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1} \dots (1.38)$$

Y la diferencia de los logit es:

$$\log(\psi) = \log(e^{\beta_1}) = \beta_1 \dots (1.39)$$

Entonces se puede interpretar a β_1 como la diferencia de logit(1) y logit(0), y a e^{β_1} como la razón de momios.

Variable explicativa nominal

Supóngase que en vez de dos categorías la variable independiente toma $k > 2$ distintos valores. Por ejemplo variables que denoten país, estado de residencia, religión etc. Cada una de estas variables tiene un número fijo discreto de resultados con escala de medida nominal. Sería inadecuado tratar estos valores como escala de intervalo. Dado este caso se deben formar variables de diseño (o *variables dummy*), las cuales son usadas para representar las categorías de la variable regularmente llamada *factor*.

En general, si un factor tiene k posibles valores, entonces será necesario crear $k - 1$ variables de diseño.

Si se supone que la j -ésima variable independiente, X_j , tiene k_j niveles. Las $k_j - 1$ variables dummy serán denotadas como D_{ju} $u = 1, 2, \dots, k - 1$ y los coeficientes para estas variables serán denotados como B_{ju} $u = 1, 2, \dots, k - 1$. Por lo tanto el modelo logit con la j -ésima variable en escala nominal será:

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \dots + \sum_{u=1}^{k-1} \beta_{ju} D_{ju} + \dots + \beta_p x_p \dots \dots (1.40)$$

Existen diferentes métodos de codificación para las variables de diseño, los cuales llevan a diferentes interpretaciones de los coeficientes estimados para éstas (Hosmer y Lemeshow, 1989). Con relación a un método con fácil interpretación de la razón de momios es el siguiente:

Supóngase que sólo se tiene una variable explicativa y ésta es categórica con k niveles, entonces se crean las variables de diseño como se muestra en la tabla 1.3.

Tabla 1.3.- Ejemplo de codificación de variables de diseño

Variable original	Variables de diseño			
	D_1	D_2	D_3	D_{k-1}
Nivel 1	0	0	0	0
Nivel 2	1	0	0	0
Nivel 3	0	1	0	0
....				
Nivel k	0	0	0	1

Ahora se fija un nivel con el cual los demás serán comparados, por ejemplo el Nivel 1, éste será la categoría de referencia. Con esto se tiene que

$$\begin{aligned} \log[\hat{\psi}(\text{nivel } i, \text{nivel } 1)] &= \widehat{\text{logit}}(\text{nivel } i) - \widehat{\text{logit}}(\text{nivel } 1) \\ &= [\hat{\beta}_0 + \hat{\beta}_{11}(D_1 = 0) + \hat{\beta}_{12}(D_2 = 0) + \dots + \hat{\beta}_{1i}(D_i = 1) + \dots + \hat{\beta}_{1(k-1)}(D_{k-1} = 0)] \\ &\quad - [\hat{\beta}_0 + \hat{\beta}_{11}(D_1 = 0) + \dots + \hat{\beta}_{k-1}(D_{k-1} = 0)] \\ &= \hat{\beta}_{1i} \dots \dots (1.41) \end{aligned}$$

Por lo tanto:

$$\log[\hat{\psi}(\text{nivel } i, \text{nivel } 1)] = \hat{\beta}_{1i} \dots \dots (1.42)$$

Entonces se puede interpretar a $\hat{\beta}_{1i}$ como el logaritmo de la razón de momios entre el *nivel* i y la categoría de referencia (en este caso el nivel 1). Tomando la exponencial de ambos lados se tiene que la relación entre la razón de momios y el coeficiente estimado es de nuevo la función exponencial.

$$\hat{\psi}(\text{nivel } i, \text{nivel } 1) = e^{\hat{\beta}_{1i}} \dots (1.43)$$

Variable explicativa continua

Supóngase que se tiene sólo una variable explicativa y ésta es continua. Bajo la consideración de que el logit es lineal en la variable continua x , su es: $\text{logit}(x) = \beta_0 + \beta_1 x$. De aquí se sigue que el coeficiente de la pendiente, β_1 , da el cambio en el logaritmo de la razón de momios para un incremento de una unidad en x , esto es $\beta_1 = \text{logit}(x + 1) - \text{logit}(x)$ para cualquier valor de x . En muchas ocasiones el cambio en una unidad no será de mucho interés, por lo tanto, para dar una útil interpretación para variables de escala continua se necesita desarrollar un método para un cambio arbitrario de c unidades en la variable.

El logaritmo de la razón de momios para un cambio de c unidades en x es obtenido de la diferencia de logits

$$\text{logit}(x + c) - \text{logit}(x) = c\beta_1 \dots (1.44)$$

Y la razón de momios asociada es obtenida sacando la exponencial a esta diferencia

$$\psi(c) = \psi(x + c, x) = \exp(c\beta_1) \dots (1.45)$$

Una estimación puede ser calculada reemplazando β_1 por su estimador máximo verosímil $\hat{\beta}_1$.

Interacciones

Las interacciones pueden ser explicadas por términos adicionales que son incluidos en el componente sistemático del modelo, el cual muestra el producto de variables explicativas que interactúan. Por ejemplo, si hay una interacción entre x_1 y x_2 , el componente lineal puede ser representado como

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

Lo anterior se puede hacer cuando se piense que la influencia de x_1 sobre la variable de respuesta se modifique en función de cuál sea el valor de x_2 o viceversa.

Esta idea puede extenderse a más variables, pudiéndose así incorporar expresiones que involucren tres o más de ellas. Si se incluye un término de cierto orden, se deben incluir entonces todos los de grado menor.

Se puede observar que el coeficiente de x_1 en la ecuación no es constante sino que depende de x_2 o viceversa, ya que:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 = \beta_0 + (\beta_1 + \beta_3 x_2) x_1 + \beta_2 x_2$$

El grado en que influye el aumento de una unidad en el cociente de momios es igual a:

$$\exp(\beta_1 + \beta_3 x_2) = \exp(\beta_1) \exp(\beta_3 x_2)$$

1.3.4.-Inferencia

Intervalo de confianza para β_j

La matriz \mathfrak{S} con elementos (1.35) puede ser escrita como:

$$\mathfrak{S} = \mathbf{X}' \mathbf{Diag}[n_i \pi_i (1 - \pi_i)] \mathbf{X} \dots \dots (1.46)$$

Donde \mathbf{X} y $\mathbf{Diag}[n_i \pi_i (1 - \pi_i)]$ son matrices. La primera es la de diseño y la segunda es diagonal $n \times n$ con elementos $n_i \pi_i (1 - \pi_i)$. Por lo tanto la matriz de varianza-covarianza estimada de $\boldsymbol{\beta}$ es

$$\widehat{cov}(\boldsymbol{\beta}) = \{\mathbf{X}' \mathbf{Diag}[n_i \pi_i (1 - \pi_i)] \mathbf{X}\}^{-1} \dots \dots (1.47)$$

Entonces

$$\hat{\beta}_j \pm z_{1-\alpha/2} \sqrt{Var(\hat{\beta}_j)} \dots \dots (1.48)$$

Es un intervalo Wald al $(1 - \alpha)100\%$ de confianza para β_j . La varianza de β_j es el j -ésimo elemento de la diagonal de $\widehat{cov}(\boldsymbol{\beta})$.

Intervalo de confianza para la razón de momios

La razón de momios, ψ , es usualmente el parámetro de interés en la regresión logística debido a su fácil interpretación. En teoría, para muestras suficientemente grandes, la distribución de $\hat{\psi}$ será normal. Desafortunadamente, el tamaño de muestra regularmente excede al del estudio. Por lo tanto, la inferencia está regularmente basada en $\log(\hat{\psi})$, el cual tiende a seguir una distribución normal para tamaños de muestra mucho menores. Por ejemplo en el caso de un sólo predictor dicotómico: $\log(\hat{\psi}) = \hat{\beta}_1$.

Para encontrar un intervalo de confianza para ψ se tiene que calcular primero uno para $e^{\hat{\beta}_1}$. Una manera de hacerlo es construir un intervalo para $\hat{\beta}_1$ y sacando la exponencial a ambos extremos de éste, es decir

$$\exp [\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} * \widehat{SE}(\hat{\beta}_1)] \dots (1.49)$$

En el caso de una variable continua se tiene que el intervalo de confianza para la razón de momios en un cambio de x en c unidades es

$$\exp[c\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}}(c * \widehat{SE}(\hat{\beta}_1))]$$

Intervalo de confianza para $\pi(x)$

Una vez que $\hat{\beta}$ ha sido obtenido, la relación entre la probabilidad de respuesta estimada y los valores x_1, x_2, \dots, x_p de las variables explicativas puede ser expresada como:

$$\text{logit}[\hat{\pi}(\mathbf{x})] = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p$$

o equivalentemente:

$$\hat{\pi}(\mathbf{x}) = \frac{\exp[\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p]}{1 + \exp[\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p]} \dots (1.50)$$

De aquí que un intervalo al $(1 - \alpha)100\%$ de confianza para el logit con \mathbf{x} fijo y muestras grandes es:

$$\hat{\eta} \pm z_{1-\alpha/2} * SE(\hat{\eta}) \dots (1.51)$$

Donde $\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1x_1 + \dots + \hat{\beta}_px_p$ y $SE(\hat{\eta})$ es la raíz cuadrada de:

$$\text{var}(\hat{\beta}_0 + \hat{\beta}_1x_1 + \dots + \hat{\beta}_px_p) = \text{var}(\hat{\beta}_0) + \sum_{i=1}^n x_i^2 \text{var}(\hat{\beta}_i) + 2 \sum x_i x_j \text{cov}(\hat{\beta}_i, \hat{\beta}_j)$$

Las varianzas y covarianzas anteriores son tomadas de la matriz de varianza- covarianza de $\hat{\beta}$. Por lo anterior y por (1.50) un intervalo al $(1 - \alpha)100\%$ de confianza para $\pi(\mathbf{x})$ con \mathbf{x} fijo es

$$\left(\frac{\exp [\hat{\eta} - z_{1-\frac{\alpha}{2}} * SE(\hat{\eta})]}{1 + \exp [\hat{\eta} - z_{1-\frac{\alpha}{2}} * SE(\hat{\eta})]} , \frac{\exp [\hat{\eta} + z_{1-\frac{\alpha}{2}} * SE(\hat{\eta})]}{1 + \exp [\hat{\eta} + z_{1-\frac{\alpha}{2}} * SE(\hat{\eta})]} \right) \dots (1.52)$$

Pruebas de significancia

En el modelo logístico, después de la estimación de los parámetros, la primera tarea comúnmente concierne al cálculo de la significancia de las variables.

Para el modelo logit, $H_0: \beta = 0$ expresa que la probabilidad de éxito es independiente de la variable asociada a β . La estadística Wald (sección 1.2.4) es simple. Para muestras grandes la fórmula 1.53 tiene una distribución normal estándar cuando $\beta = 0$. Equivalentemente $z^2 = (\hat{\beta}/SE(\hat{\beta}))^2$ tiene una distribución nula Ji-cuadrada con un grado de libertad.

$$z = \frac{\hat{\beta}}{SE(\hat{\beta})} \dots (1.53)$$

Aunque la prueba Wald es adecuada para muestras grandes, el contraste de cociente de verosimilitudes (sección 1.5) es más poderoso y más fiable para los tamaños de muestra casi

siempre usados en la práctica. La estadística de prueba es $-2[L_0 - L_1]$, que también tiene una distribución Ji-cuadrada con un grado de libertad.

Bondad de ajuste

Después de ajustar un modelo a un conjunto de observaciones, es natural investigar sobre la medida en que los valores ajustados de la variable de respuesta bajo el modelo se comparan con los datos. Existen varias estadísticas que miden la discrepancia entre una proporción binomial observada y la ajustada.

Devianza

Una medida de discrepancia ampliamente usada, es la devianza (1.22) la cual compara la log-verosimilitud maximizada bajo los modelos L_M (el actual) y L_S (el saturado) mediante la estadística $D = -2[L_M - L_S]$.

Para datos binomiales, L_M tiene la forma

$$L_M = \sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} + y_i \log \hat{\pi}_i + (n_i - y_i) \log(1 - \hat{\pi}_i) \right\} \dots (1.54)$$

Bajo el modelo saturado, las probabilidades ajustadas serán las mismas que las proporciones observadas $\tilde{\pi}_i = y_i/n_i$ $i = 1, 2, \dots, n$. Por lo tanto la log-verosimilitud maximizada para L_S es:

$$L_S = \sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} + y_i \log \tilde{\pi}_i + (n_i - y_i) \log(1 - \tilde{\pi}_i) \right\} \dots (1.55)$$

La devianza está entonces dada por:

$$D = -2[L_M - L_S]$$

$$= 2 \sum_{i=1}^n \left\{ y_i \log \left(\frac{\tilde{\pi}_i}{\hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{1 - \tilde{\pi}_i}{1 - \hat{\pi}_i} \right) \right\} \dots (1.56)$$

Si el número ajustado de éxitos bajo L_M es $\hat{y}_i = n_i \hat{\pi}_i$, la devianza puede ser escrita como

$$D = 2 \sum_{i=1}^n \left\{ y_i \log \left(\frac{y_i}{\hat{y}_i} \right) + (n_i - y_i) \log \left(\frac{1 - y_i}{1 - \hat{y}_i} \right) \right\} \dots (1.57)$$

en la cual se observa que esta estadística compara las observaciones y_i con su correspondiente valor ajustado \hat{y}_i bajo el actual modelo.

Para el importante caso donde las observaciones son binarias ($n_i = 1 \quad i = 1, 2, \dots, n$), la devianza es:

$$D = -2 \sum_{i=1}^n \{ \hat{\pi}_i \log(\hat{\pi}_i) + \log(1 - \hat{\pi}_i) \} \dots (1.58)$$

Ésta depende de las observaciones binarias y_i sólo a través de las probabilidades ajustadas $\hat{\pi}_i$, y entonces no dice nada acerca de la concordancia entre las observaciones y sus correspondientes valores ajustados. Consecuentemente, la devianza del modelo ajustado para datos binarios no puede ser usada como medida de bondad de ajuste.

Estadística χ^2 de Pearson

Una de las alternativas más populares a la devianza es la χ^2 de Pearson definida por

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)} \dots (1.59)$$

Ambas, la devianza y la χ^2 tienen la misma distribución asintótica χ^2_α . El valor numérico de las dos estadísticas generalmente diferirán, pero la diferencia rara vez será de importancia práctica. Grandes diferencias entre las dos estadísticas puede ser tomada como un indicador de que la aproximación Ji-cuadrada a la distribución de la devianza o de la estadística χ^2 no es adecuada.

La estadística Hosmer-Lemeshow.

En contraste con la devianza, la estadística Hosmer-Lemeshow es una medida de bondad de ajuste que puede ser usada en el modelado de datos binarios. De hecho, si los datos están de manera binomial, se debe tratar de desagruparlos antes de que la estadística sea evaluada.

Con el fin de calcular el valor de la estadística, los datos son primero ordenados ascendentemente con respecto a sus correspondientes probabilidades ajustadas bajo el modelo asumido. Los valores ordenados son entonces separados en $g \geq 3$ conjuntos de tamaño similar. Los datos observados de la variable de respuesta binaria en cada grupo son sumados. Lo mismo se hace para sus probabilidades ajustadas. La adición de las y_i y de las $\hat{\pi}_i$ en cada conglomerado de datos son el número de éxitos observados y estimados (en el mismo grupo) respectivamente. Estos valores son entonces comparados usando una estadística Ji-cuadrada.

Supóngase que hay m_i observaciones en el i -ésimo grupo $i = 1, 2, \dots, g$, donde la cantidad observado de éxitos es o_i y el correspondiente número esperado es e_i . También, sea $\bar{\pi}_i$ la probabilidad promedio de éxito en el i -ésimo grupo, de modo que $\bar{\pi}_i = e_i/m_i$. La estadística Hosmer-Lemeshow está entonces dada por:

$$X_{HL}^2 = \sum_{i=1}^g \frac{(o_i - m_i \bar{\pi}_i)^2}{m_i \hat{\pi}_i (1 - \bar{\pi}_i)} \dots \dots (1.60)$$

Usando estudios de simulación, esta estadística ha mostrado tener una distribución aproximada Ji-cuadrada con $(g-1)$ grados de libertad, cuando el modelo ajustado es apropiado, conduciendo a una prueba de bondad de ajuste adecuada para datos binarios no agrupados.

CAPÍTULO 2. MODELO LINEAL GENERALIZADO MIXTO

En este capítulo se definirán los conceptos de efecto fijo y aleatorio, así como las diferencias entre ellos. El modelo lineal generalizado mixto se presenta como una extensión de un GLM. Al igual que en el capítulo anterior la regresión logística se desarrolla como un caso particular y se detallan sus características. También se darán algunos puntos para tomar en cuenta a la hora de decidir qué tipo de efecto tiene una variable.

2.1.-Efectos fijos

Los efectos de un factor (variable explicativa) son siempre de dos tipos en términos de los parámetros. El primer caso es el de los parámetros que son considerados como constantes; el efecto asociado a estos es llamado fijo. Son atribuibles a un conjunto finito de niveles de un factor que ocurre en los datos y en los cuales se tiene interés.

Un ejemplo de efectos fijos podría ser el siguiente. Supóngase que se tiene un ensayo clínico en el cual una droga es administrada en cuatro diferentes niveles de dosis. Si y_{ij} es el dato de respuesta para la j -ésima persona recibiendo la dosis i , se puede empezar modelando la media como: $E[y_{ij}] = \mu_i = \mu + \alpha_i$, donde μ es un promedio general y α_i es el efecto en la respuesta debido a la dosis i . Las α_i son efectos fijos por que los cuatro niveles de dosis usados en el ensayo clínico son los únicos niveles que están siendo estudiados. Son las dosis en las que se fija la atención del investigador. Si todos los factores tomados en cuenta son de este tipo, el modelo es llamado de efectos fijos.

2.2.-Efectos aleatorios

El segundo tipo de efecto, llamado aleatorio, es el de los parámetros considerados como realizaciones de una variable aleatoria. Éstos son atribuibles generalmente a un conjunto infinito de niveles de un factor, del cual sólo una muestra aleatoria se considera para ocurrir en los datos.

Como ejemplo, considérese un estudio realizado en Guatemala mediante la encuesta de salud infantil y materna. La variable respuesta es si el niño cuenta con todas las vacunas que le corresponden. Los datos provienen de 2159 niños que fueron tomados de 1595 familias, que a su vez viven en 161 comunidades, por lo tanto, existen tres niveles de variación; en cada uno de ellos se tomaron ciertas variables explicativas. En el nivel niño, las variables consideradas fueron la edad del infante y el orden de nacimiento del mismo. En el nivel familia se midió la raza, edad y situación laboral de la mamá, así como el grado de educación de la madre y el padre. Las variables correspondientes al nivel comunidad son la proporción de población indígena y si la comunidad

es rural o urbana. Es razonable pensar que las familias y las comunidades son una muestra aleatoria que provienen de una distribución. Estos datos pueden ser modelados introduciendo efectos aleatorios para la familia y la comunidad, controlando así la posible correlación creada por el anidamiento de los niveles. El resto de las variables pueden ser usadas con efectos fijos. Un modelo adecuado para este estudio se puede escribir de la siguiente manera:

$$\text{logit}(\pi_{ijk}) = \beta_0 + \boldsymbol{\beta}'_1 \mathbf{x}_{ijk} + \boldsymbol{\beta}'_2 \mathbf{x}_{jk} + \boldsymbol{\beta}'_3 \mathbf{x}_k + f_{jk} + c_k$$

Donde \mathbf{x}_{ijk} , \mathbf{x}_{jk} y \mathbf{x}_k son vectores de variables explicativas para los niveles niño, familia y comunidad; $\boldsymbol{\beta}'_1$, $\boldsymbol{\beta}'_2$, $\boldsymbol{\beta}'_3$ son sus correspondientes vectores de parámetros fijos. Los efectos aleatorios para familia y comunidad son f_{jk} y c_k respectivamente. Las familias y las comunidades fueron elegidas al azar con el fin de tratarlas como una representación de toda la población de Guatemala, por lo tanto se puede hacer inferencia acerca de ella. Está es una característica de los efectos aleatorios: Estos pueden ser usados como base para hacer inferencia de la población de la cual provienen. Al ser una variable aleatoria, los datos podrían ser útiles para hacer inferencia acerca de la varianza de tal variable, es decir acerca de la magnitud de la variación entre familias y entre comunidades en el caso del ejemplo anterior.

2.3.- ¿Fijo o aleatorio?

Algunas veces la decisión sobre si cierto efecto es fijo o aleatorio no es inmediatamente obvio. Al tratar de decidir esto, el contexto de los datos, la manera en la cual fueron recolectados y el ambiente del cual provienen son principios determinantes. Al considerar estos puntos, la pregunta importante es: ¿Los niveles de las variables serán considerados como una muestra aleatoria de una población de valores y por tanto tienen una distribución? Otra forma de decir lo anterior es: ¿Los niveles de un factor provienen de una distribución de probabilidad? y ¿Hay suficiente información acerca de él para decidir si sus niveles en los datos son una muestra aleatoria? Respuestas negativas a las preguntas anteriores significa que debemos tratar al coeficiente como fijo y estimar los efectos de sus niveles. Por otro lado, respuestas afirmativas indican que debemos tratar al factor como aleatorio y estimar el componente de la varianza. En tal caso, cuando también existe interés en la realización de la variable aleatoria, se puede usar un procedimiento de predicción para los efectos (Agresti, 2002; Collet, 2003; McCulloch y Serly, 2008).

Con el fin de ayudar a decidir si un factor es considerado como fijo o aleatorio, el diagrama de decisión de la figura 2.1 puede ser usado.

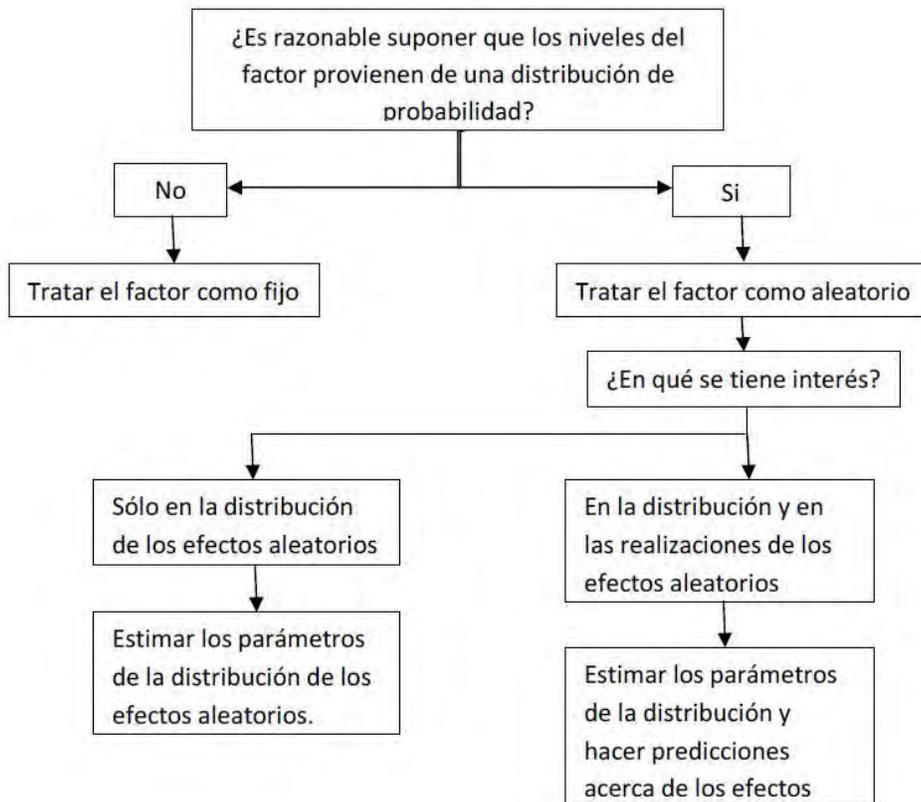
2.4.-Modelo lineal generalizado mixto

Muchos modelos estadísticos pueden contener tanto efectos fijos como aleatorios. A éstos se les llama modelos de efectos mixtos.

Los modelos de efectos mixtos son usados para describir la relación entre una variable respuesta y una covariable en datos agrupados de acuerdo a uno o más factores de clasificación.

El modelo lineal generalizado mixto (GLMM por sus siglas en inglés) tiene dos principales ventajas: la primera, los datos no necesariamente se suponen normalmente distribuidos o que la media no necesariamente es una combinación lineal de los parámetros, pero sí lo es alguna función de la media. La segunda es la aleatoriedad de los efectos mixtos, lo que permite particularizar a los individuos asociando una relación de varianza-covarianza. Así, el modelo mixto condiciona los efectos fijos con los aleatorios dando una asociación entre grupos o niveles de anidamiento, permitiendo que el modelo sea estudiado para dar un comportamiento distinto a los individuos. Otorgando distinta variabilidad a individuos de distintos grupos, por un lado o dando distinta variabilidad a elementos dentro del mismo grupo.

Figura 2.1 Árbol de decisión entre fijo o aleatorio.



2.4.1.-El modelo

Sea \mathbf{y} el vector de respuesta, al igual que en el modelo lineal generalizado éste contiene elementos independientes, sólo que en esta ocasión condicionados a los efectos aleatorios, cada uno con una densidad perteneciente a la familia exponencial, es decir:

$$y_i | \mathbf{u}_i \sim \text{indep. } f_{Y_i | \mathbf{u}_i}(y_i | \mathbf{u}_i)$$

$$f_{Y_i | \mathbf{u}_i}(y_i | \mathbf{u}_i) = \exp \left\{ \frac{[y_i \theta_i - d(\theta_i)]}{a(\phi)} - c(y_i, \phi) \right\} \dots \dots (2.1)$$

La fórmula (2.1) es una extensión de (1.1), llamada familia exponencial de dispersión. El parámetro θ_i es llamado natural o canónico y ϕ es llamado parámetro de dispersión. Cuando ϕ es conocido (2.1) se simplifica a la familia exponencial natural (1.1).

El siguiente paso es modelar una función de la media como un modelo lineal tanto en los efectos fijos como aleatorios.

$$E[y_i | \mathbf{u}_i] = \mu_i \dots \dots (2.2)$$

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u} \dots \dots (2.3)$$

$g(\cdot)$ es la función liga. \mathbf{x}_i^T y $\boldsymbol{\beta}$ son el i ésimo renglón de la matriz de observaciones y el vector de efectos fijos respectivamente. A diferencia de un GLM, el modelo mixto generalizado incluye \mathbf{z}_i^T que es el renglón i de la matriz de diseño para los efectos aleatorios y \mathbf{u} el vector de parámetros correspondiente. Notemos que μ_i es la media condicional de y_i dado \mathbf{u} . Para completar la especificación del modelo se asigna una distribución a \mathbf{u} :

$$\mathbf{u} \sim f_{\mathbf{u}}(\mathbf{u}) \dots \dots (2.4)$$

Usualmente (más no necesariamente) se suele suponer que el vector \mathbf{u} sigue una distribución normal multivariada $N(\mathbf{0}, \boldsymbol{\Sigma})$. La matriz $\boldsymbol{\Sigma}$ depende de las componentes de varianza desconocidas y posiblemente de parámetros de correlación.

2.4.2.-Consecuencias de tener efectos aleatorios

El promedio de \mathbf{y} puede ser calculado por la media condicional:

$$\begin{aligned} E[y_i] &= E[E[y_i|\mathbf{u}]] \\ &= E[\mu_i] \\ &= E[g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u})] \dots (2.5) \end{aligned}$$

En general, (2.5) no puede ser simplificada, debido a la función no lineal g^{-1} . La varianza marginal de \mathbf{y} se calcula mediante:

$$\begin{aligned} \text{var}(y_i) &= \text{var}(E[y_i|\mathbf{u}] + E[\text{var}(y_i|\mathbf{u})]) \\ &= \text{var}(\mu_i) + E[a(\phi)v(\mu_i)] \\ &= \text{var}\left(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u})\right) + E\left[a(\phi)v\left(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u})\right)\right] \dots (2.6) \end{aligned}$$

Se necesita conocer la forma de g o la distribución condicional de \mathbf{y} para poder simplificar esta última expresión. El uso de efectos aleatorios introduce una correlación entre las observaciones que tienen a cualquiera de ellos en común. Asumiendo independencia condicional de los elementos de \mathbf{y} , se tiene:

$$\begin{aligned} \text{cov}(y_i, y_j) &= \text{cov}(E[y_i|\mathbf{u}], E[y_j|\mathbf{u}]) + E[\text{cov}(y_i, y_j|\mathbf{u})] \\ &= \text{cov}(\mu_i, \mu_j) + E[0] \\ &= \text{cov}\left(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u}), g^{-1}(\mathbf{x}_j^T \boldsymbol{\beta} + \mathbf{z}_j^T \mathbf{u})\right) \dots (2.7) \end{aligned}$$

Las dos funciones dentro de la última covarianza de 2.7 son monótonas en \mathbf{u} , resultando no correlacionadas negativamente (Agresti, 2002; McCulloch y Serly, 2008).

2.4.3.-Estimación por máxima verosimilitud

De (2.1), (2.3) y (2.4) la función de verosimilitud puede escribirse como

$$L = \int \prod_i f_{Y_i|u}(y_i|\mathbf{u}) f_{\mathbf{u}}(\mathbf{u}) d\mathbf{u} \dots (2.8)$$

Donde, la integración es sobre la distribución q -dimensional de \mathbf{u} .

Las ecuaciones de verosimilitud para los efectos fijos son difíciles numéricamente, sin embargo se pueden escribir de manera muy general, de (2.8)

$$l = \log \int f_{Y|u}(\mathbf{y}|\mathbf{u}) f_{\mathbf{u}}(\mathbf{u}) d\mathbf{u} = \log f_Y(\mathbf{y}) \dots (2.9)$$

Por lo tanto

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \int f_{Y|u}(\mathbf{y}|\mathbf{u}) f_{\mathbf{u}}(\mathbf{u}) d\mathbf{u} / f_Y(\mathbf{y}) \\ &= \int \left[\frac{\partial}{\partial \boldsymbol{\beta}} f_{Y|u}(\mathbf{y}|\mathbf{u}) \right] f_{\mathbf{u}}(\mathbf{u}) d\mathbf{u} / f_Y(\mathbf{y}) \dots (2.10) \end{aligned}$$

Ya que $f_{\mathbf{u}}(\mathbf{u}) d\mathbf{u}$ no depende de $\boldsymbol{\beta}$. También

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} f_{Y|u}(\mathbf{y}|\mathbf{u}) &= \left(\frac{1}{f_{Y|u}(\mathbf{y}|\mathbf{u})} \frac{\partial f_{Y|u}(\mathbf{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} \right) f_{Y|u}(\mathbf{y}|\mathbf{u}) \\ &= \frac{\partial \log f_{Y|u}(\mathbf{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} f_{Y|u}(\mathbf{y}|\mathbf{u}) \dots (2.11) \end{aligned}$$

Entonces (2.10) puede ser reescrita como 2.12.

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \int \frac{\partial \log f_{Y|U}(\mathbf{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} f_{Y|U}(\mathbf{y}|\mathbf{u}) f_U(\mathbf{u}) d\mathbf{u} / f_Y(\mathbf{y}) \\ &= \int \frac{\partial \log f_{Y|U}(\mathbf{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} f_{U|Y}(\mathbf{u}|\mathbf{y}) d\mathbf{u} \dots \dots (2.12) \end{aligned}$$

Usando la derivada de la log-verosimilitud para un GLM (fórmula 2.12) resulta:

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \int \mathbf{X}^T \mathbf{W}^* (\mathbf{y} - \boldsymbol{\mu}) f_{U|Y}(\mathbf{u}|\mathbf{y}) d\mathbf{u} \\ &= \mathbf{X}^T E[\mathbf{W}^* | \mathbf{y}] \mathbf{y} - \mathbf{X}^T E[\mathbf{W}^* \boldsymbol{\mu} | \mathbf{y}] \dots \dots (2.13) \end{aligned}$$

Donde: $\mathbf{W}^* = \text{diag} \{ [a(\phi) v(\mu_i) g_\mu(\mu_i)]^{-1} \}$. Por lo tanto la ecuación de verosimilitud para los efectos fijos es:

$$\mathbf{X}^T E[\mathbf{W}^* | \mathbf{y}] \mathbf{y} = \mathbf{X}^T E[\mathbf{W}^* \boldsymbol{\mu} | \mathbf{y}] \dots \dots (2.14)$$

Estas ecuaciones deben ser resueltas numéricamente, usando métodos como la cuadratura de Gauss-Hermite (Collet 2003) y Newton-Raphson (Dobson 2002) para efectos aleatorios distribuidos normalmente.

Se puede usar un resultado similar a (2.12) para calcular los estimadores máximo verosímiles de los parámetros de la distribución de los efectos aleatorios. Sea $\boldsymbol{\varphi}$ el vector de tales parámetros, entonces

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\varphi}} &= \int \frac{\partial \log f_U(\mathbf{u})}{\partial \boldsymbol{\varphi}} f_{U|Y}(\mathbf{u}|\mathbf{y}) d\mathbf{u} \\ &= E \left[\frac{\partial \log f_U(\mathbf{u})}{\partial \boldsymbol{\varphi}} \middle| \mathbf{y} \right] \dots \dots (2.15) \end{aligned}$$

No es posible hacer más simplificaciones, a menos que se especifique la distribución de los efectos aleatorios (McCulloch y Serly, 2008).

2.4.4.-Inferencia para los parámetros en el modelo

Después del ajuste, se puede proceder a realizar inferencia acerca de los parámetros fijos de la manera usual. Por ejemplo, la prueba de razón de verosimilitud puede comparar modelos anidados. Las propiedades asintóticas en un GLMM aplican mientras el número de conglomerados aumenta, en lugar de cuando la cantidad de observaciones dentro del conglomerado se incrementa.

Hacer inferencia acerca de los efectos aleatorios es más complicado. Por ejemplo, algunas veces un modelo es un caso especial de otro en el cual un componente de varianza es igual a cero. El modelo más simple cae en el límite del espacio de parámetros relativo al modelo más complejo, por lo tanto la inferencia basada en la verosimilitud ordinaria no aplica. La distribución asintótica de la estadística de razón de verosimilitudes para la situación más común es conocida, al probar $H_0: \sigma^2 = 0$ contra $H_a: \sigma^2 > 0$ para un modelo que contiene sólo un componente de varianza. La distribución nula es igual a una mezcla de variables χ_0^2 (es decir, degenerada en cero) y χ_1^2 . El valor de cero ocurre cuando $\hat{\sigma} = 0$, en tal caso las verosimilitudes maximizadas son idénticas bajo H_0 y H_a . Cuando $\hat{\sigma} > 0$ y la estadística de prueba observada es t , el valor de probabilidad (Valor-p) para este contraste es $\frac{1}{2}P(\chi_1^2 > t)$, la mitad del Valor-p que aplica para una prueba asintótica χ_1^2 normal. Para probar más de un componente de varianza, la mezcla de la distribución se vuelve más compleja, y es más sencillo usar una prueba de puntajes (score statistic).

2.4.5.-Predicción usando efectos aleatorios

El uso de efectos aleatorios en un modelo implica heterogeneidad de ciertas cantidades de interés, tales como las razones de momios y las estimaciones de las mismas son casi siempre combinaciones lineales de efectos fijos y aleatorios.

Dados los datos, la distribución condicional de $(\mathbf{u}|\mathbf{y})$ contiene la información acerca de los efectos aleatorios \mathbf{u} . Una predicción para \mathbf{u} es $E(\mathbf{u}|\mathbf{y})$, su media posterior dado los datos. Calcular $E(\mathbf{u}|\mathbf{y})$ requiere integración numérica, esta media depende de $\boldsymbol{\beta}$ y $\boldsymbol{\Sigma}$ (los parámetros de la distribución de los efectos aleatorios, la matriz de varianzas-covarianzas en el caso normal), por lo tanto en la práctica se substituyen $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\Sigma}}$ en la aproximación (McCulloch y Serly, 2008).

2.5.-Modelos mixtos para datos binarios

En esta sección, es descrito un modelo mixto para el análisis de datos binarios. Aunque el modelo de regresión logística será usado, otras funciones liga como la probit pueden ser usadas de manera similar.

2.5.1.-Modelo de regresión logística con un efecto aleatorio

Supóngase que y_{ij} es la respuesta 0, 1 a cierto experimento para el individuo j en un grupo i , entonces un modelo razonable para la probabilidad de éxito y una variable explicativa es:

$$E[y_{ij}|u_i] = \pi(x_{ij}) = \frac{e^{\beta_0 + \beta_1 x_{ij} + u_i}}{1 + e^{\beta_0 + \beta_1 x_{ij} + u_i}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ij} + u_i)}} \dots (2.16)$$

O equivalentemente

$$\text{logit}[\pi(x_{ij})] = \beta_0 + \beta_1 x_{ij} + u_i \dots (2.17)$$

Y

$$y_{ij}|u_i \sim \text{indep. bernoulli}[\pi(x_{ij})]$$

$$u_i \sim i. i. d N(0, \sigma_u^2).$$

Condicionando en u_i , y_{ij} sigue un modelo de regresión logística con intersección que varía con el subíndice i (con el grupo). Así, condicionado en u_i , β_1 tiene la misma interpretación que en el modelo de regresión logística usual.

Debido a que u_i son efectos aleatorios, existen dos diferencias importantes entre el modelo logit usual y el mixto. Primero, marginalmente y_{ij} no sigue exactamente una forma logística, esto es debido a:

$$\begin{aligned} E[y_{ij}] &= E[E[y_{ij}|u_i]] = E[\pi(x_{ij})] \\ &= E\left[\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ij} + u_i)}}\right] \\ &= \int_{-\infty}^{\infty} \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ij} + u_i)}} \frac{1}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{1}{2\sigma_u^2}u_i^2} du_i \dots (2.18) \end{aligned}$$

Entonces 2.18 no puede ser evaluada en forma cerrada, y en particular, no es exactamente de la forma logística, esto es, $1/(1 + e^{\beta_0^* + \beta_1^* x_{ij}})$, para alguna elección de β_0^* y β_1^* . Sin embargo, puede ser bien aproximado por un modelo logit marginal:

$$\text{logit}(E[y_{ij}]) \approx \beta_0^* + \beta_1^* x_{ij}$$

Donde:

$$\beta_0^* = \frac{\beta_0}{\sqrt{1 + \lambda\sigma_u^2}} \quad y \quad \beta_1^* = \frac{\beta_1}{\sqrt{1 + \lambda\sigma_u^2}}$$

Para $\lambda = 256/75\pi$.

Segundo, las observaciones y_{ij} y y_{ik} están correlacionadas debido a que ambas comparten el mismo efecto aleatorio u_i . Usando (2.7) se obtiene:

$$\begin{aligned} \text{cov}(y_{ij}, y_{ik}) &= \text{cov}(E[y_{ij}|u_i], E[y_{ik}|u_i]) + E[\text{cov}(y_{ij}, y_{ik}|u_i)] \\ &= \text{cov}\left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ij} + u_i)}}, \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ik} + u_i)}}\right) + 0 \\ &= \int_{-\infty}^{\infty} \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ij} + u_i)}} \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ik} + u_i)}} \frac{1}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{1}{2\sigma_u^2}u_i^2} du_i \\ &\quad - \int_{-\infty}^{\infty} \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ij} + u_i)}} \frac{1}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{1}{2\sigma_u^2}u_i^2} du_i \\ &\quad \times \int_{-\infty}^{\infty} \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{ij} + u_i)}} \frac{1}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{1}{2\sigma_u^2}u_i^2} du_i \dots \dots (2.19) \end{aligned}$$

> 0, ya que $\sigma_u^2 > 0$

2.5.2.-Ajustando un modelo de regresión logística con un efecto aleatorio

El modelo de regresión logística para datos binarios se extiende fácilmente para incluir un efecto aleatorio. Supongamos que hay n grupos de observaciones de la forma y_i/n_i , donde y_i es el valor observado de una variable de respuesta de tipo binomial Y_i , asociada con la i – ésima proporción, $i = 1, 2, \dots, n$. Se supone que la variable de respuesta depende de k variables explicativas,

X_1, X_2, \dots, X_k , las cuales toman valores $x_{1i}, x_{2i}, \dots, x_{ki}$ para la i – ésima observación, y de un efecto aleatorio, u_i , debido a el i – ésimo nivel de un factor U .

Si la correspondiente probabilidad de respuesta es π_i , la variable aleatoria Y_i tiene una distribución binomial con parámetros n_i y π_i . La dependencia de π_i con las variables explicativas y el efecto aleatorio es modelado por

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i \dots \dots (2.20)$$

El efecto aleatorio se distribuye $N(0, \sigma_u^2)$. Entonces la ecuación anterior puede ser escrita:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \sigma_u z_i \dots \dots (2.21)$$

Donde z_i es una realización de una variable aleatoria normal estándar Z_i . La ecuación anterior es sólo para enfatizar que la desviación estándar del efecto aleatorio forma parte del componente lineal del modelo.

Si se denota $\eta_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$ para el componente lineal correspondiente a los efectos fijos, el modelo se escribe

$$\text{logit}(\pi_i) = \eta_i + \sigma_u z_i \dots \dots (2.22)$$

Y la verosimilitud de la i – ésima de n observaciones está dada por

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma_u, z_i) &= \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \\ &= \binom{n_i}{y_i} \frac{\{exp(\eta_i + \sigma_u z_i)\}^{y_i}}{\{1 + exp(\eta_i + \sigma_u z_i)\}^{n_i}} \dots \dots (2.23) \end{aligned}$$

La verosimilitud de las n observaciones es $\prod_{i=1}^n L(\boldsymbol{\beta}, \sigma_u, z_i)$. Esta función depende de $k + 2$ parámetros desconocidos $\beta_0, \beta_1, \dots, \beta_k, \sigma_u$ y también de la realización de variables aleatorias Z_1, Z_2, \dots, Z_n las cuales tienen una distribución $N(0,1)$.

Usando el procedimiento de la sección (2.4.3) se encuentra la función de verosimilitud marginal mediante:

$$L(\boldsymbol{\beta}, \sigma_u) = \prod_{i=1}^n \int_{-\infty}^{\infty} \binom{n_i}{y_i} \frac{\{exp(\eta_i + \sigma_u z_i)\}^{y_i}}{\{1 + exp(\eta_i + \sigma_u z_i)\}^{n_i}} \frac{exp(-z_i^2/2)}{\sqrt{2\pi}} dz_i \dots (2.24)$$

Una manera de llevar a cabo la integración, es el método numérico llamado fórmula de Gauss-Hermite o cuadratura, de acuerdo con el cual:

$$\int_{-\infty}^{\infty} f(u)e^{-u^2} du \approx \sum_{r=1}^m c_r f(s_r), \dots (2.25)$$

Donde los valores de c_r y s_r están dados en tablas estándar, y m es el número de puntos de cuadratura. La integral en la ecuación (2.24) puede ser expresada como una suma y entonces la verosimilitud se convierte en

$$\pi^{-n/2} \prod_{i=1}^n \binom{n_i}{y_i} \sum_{r=1}^m c_r \left[\frac{\{exp(\eta_i + \sigma_u s_{r_i} \sqrt{2})\}^{y_i}}{\{1 + exp(\eta_i + \sigma_u s_{r_i} \sqrt{2})\}^{n_i}} \right] \dots (2.26)$$

Los valores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k, \hat{\sigma}_u$ que maximizan esta expresión, o su logaritmo, pueden ser determinados numéricamente. Los errores estándar de estos parámetros también están disponibles como producto del proceso de optimización (Collet, 2003).

2.5.3.-Modelos con más de un efecto aleatorio

El modelo en la ecuación (2.16) puede ser extendido para incluir dos o más componentes aleatorios. Por ejemplo, supongamos que en adición a u_{1i} , hay otro efecto aleatorio u_{2i} . Un modelo de regresión logística para la probabilidad de respuesta está entonces dado por

$$\text{logit}(\pi_i) = \eta_i + u_{1i} + u_{2i} \dots (2.27)$$

Las variables aleatorias U_{1i} y U_{2i} se distribuyen $N(0, \sigma_{u_1}^2)$ y $N(0, \sigma_{u_2}^2)$ respectivamente y no necesariamente son independientes, entonces (U_{1i}, U_{2i}) pueden tener una distribución normal bivariada con correlación ρ .

Un modelo general puede incluir m efectos aleatorios, cada uno teniendo un coeficiente conocido, dando

$$\text{logit}(\pi_i) = \eta_i + a_{1i}u_{1i} + a_{2i}u_{2i} + \dots + a_{mi}u_{mi} \dots \dots (2.28)$$

donde los valores $a_{1i}, a_{2i}, \dots, a_{mi}$ son conocidos y $u_{1i}, u_{2i}, \dots, u_{mi}$ son realizaciones de variables aleatorias $U_{1i}, U_{2i}, \dots, U_{mi}$, con distribución normal multivariada m -dimensional, con media $\mathbf{0}$ y matriz de varianzas covarianzas Σ .

2.5.4.- Interpretación de los parámetros

La interpretación de los parámetros con efectos fijos se puede realizar de la misma manera que en la sección 1.3.3, por supuesto recordando que todo es condicional a la realización de la parte aleatoria.

Supóngase para ejemplificar que se tiene un modelo logístico donde la variable respuesta es Y , con una sola variable explicativa $X = 0,1$ con efecto fijo y que las observaciones están conglomeradas de acuerdo a k niveles (que fueron elegidos aleatoriamente) de un factor U . El efecto de U debe ser tomado entonces como aleatorio debido a la posible correlación que crea el agrupamiento de los datos. El modelo a ajustar sería:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_i + u_j \quad j = 1, \dots, k. \quad u_j \text{ es una realización de } U \sim N(0, \sigma_u^2)$$

Ahora supóngase que los parámetros estimados son $\hat{\beta}_0 = 1.3$, $\hat{\beta}_1 = -2.1$ y $\hat{\sigma}_u^2 = 2.3$ además de que son todos significativos.

En la escala del logit el intercepto para el grupo j sería en el ejemplo $\hat{\beta}_0 + u_j = 1.3 + u_j$ donde u_j es el efecto de agrupación j (el cual se estima después de ajustar el modelo). Entonces se puede decir que se tiene un modelo con intersección aleatoria y $\hat{\beta}_0$ es el intercepto para un grupo de observaciones (las que comparten el nivel j de U) para las cuales se estimó $u_j = 0$.

La interpretación de $\hat{\beta}_1 = -2.1$ es: Las posibilidades de que $Y = 1$ cuando $X = 1$ son $e^{-2.1} = 0.12$ veces las posibilidades de cuando $X = 0$.

El efecto aleatorio $\hat{\sigma}_u^2 = 2.3$ se interpreta como la varianza que existe en los efectos de los diferentes niveles de agrupación. También se puede calcular el coeficiente de partición de varianza (VPC por sus siglas en inglés) o también llamado coeficiente de correlación intraclase por algunos autores (Collet, 2003; Fielding y Pillinger) de acuerdo a:

$$VPC = \text{correlación intraclase} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \frac{\sigma_u^2}{\sigma_u^2 + \pi^2/3} \approx \frac{\sigma_u^2}{\sigma_u^2 + 3.29}$$

Donde σ_e^2 es la varianza de nivel 1 o dentro de grupos (within-group variation). En el caso de los modelos logit mixtos, este error se supone logístico estándar (Rodríguez Germán, 2003) (en vez de normal como en el caso de los lineales mixtos), por lo tanto tiene varianza aproximada a 3.29.

En el ejemplo $VPC=0.411$ el cual se interpreta como la variabilidad debida a los niveles de agrupación o como la correlación que existe entre cualquier par de observaciones (con respecto a la respuesta binaria Y) dentro de un mismo grupo. Ésta es muy alta, con lo cual se justifica la aplicación de un modelo mixto.

CAPÍTULO 3. LEPTOSPIROSIS: MARCO TEÓRICO

Este capítulo detalla cuestiones teóricas acerca de la leptospirosis así como de la bacteria que la causa. Contiene términos técnicos que fueron imposibles de evadir e importantes de presentar para algún lector con conocimientos del área biológica. Si algún lector ajeno a estos términos desea leer este capítulo las palabras marcadas con un superíndice (*) estarán presentadas de manera relativamente sencilla en el glosario al final de este trabajo para su mejor entendimiento. Las secciones necesarias para la continuidad de este estudio son 3.1, 3.3, 3.5 y 3.6.

3.1.-Generalidades.

La leptospirosis es una enfermedad re-emergente, infecto-contagiosa con carácter zoonótico* de importancia global y es causada por espiroquetas del género *Leptospira* interrogans. Anualmente, millones de casos en humanos ocurren en todo el mundo y las tasas de mortalidad estimadas hasta el momento fluctúan entre el 20% al 25% en algunas regiones del mundo. La transmisión de la infección a la especie humana es a través de la exposición al medio ambiente contaminado con la orina de los animales enfermos en cuyo caso son colonizados crónicamente por las diferentes serovariedades de *Leptospira* spp. La exposición al agente patógeno se produce tanto en países industrializados como aquellos en vías del desarrollo, pero es más prevalente en las zonas tropicales donde las condiciones ambientales y socioeconómicas son particularmente favorables (Kupek y cols., 2000; Pate y cols., 2000; Tangkanakul y cols., 2000; Jansen y cols., 2005). Las leptospirosis pueden sobrevivir en ambientes con temperatura cálida y húmeda es decir en el suelo o en el agua por semanas o hasta meses. En las áreas urbanas y rurales bajo condiciones de hacinamiento se puede facilitar la exposición humana a la orina de los animales infectados con cepas nativas identificadas en el ganado vacuno, animales de compañía, roedores o animales salvajes (Atxaerandio y cols., 2005). En ciertas ocasiones cuando existen plagas de roedores que florecen en las zonas urbanas pueden ser una importante fuente de infección aumentando el nivel de riesgo para las personas que ahí habitan (Ganoza y cols., 2006). Además, los casos recientes en todo el mundo entre atletas, personal militar, civiles y viajeros ponen de relieve el riesgo de contagio de la leptospirosis en los entornos ecológicos. Los factores de riesgo inherentes a la infección incluyen: el consumo de agua contaminada, la exposición a los residuos sólidos conducidos por arroyos o ríos y la presencia de perros, gatos y ratas en las casas que deambulan en los sitios de impacto ambiental. El riesgo de contagio y transmisión también parece coincidir con la estación meteorológica cálida y la mayor ocurrencia durante los tiempos de severas inundaciones que lava el suelo de los residuos sólidos contaminados con orina de animales enfermos.

3.2.-Taxonomía.

División: Procariotes; Clase: Schizomicetes; Orden: Spirochaetales; Familia: Leptospiraceae; Género: *Leptospira spp.*; Especies: *L. interrogans*, *L. biflexa*.

La familia Leptospiraceae comprende tres géneros: *Leptospira*, *Leptonema* y *Turneria* (Faine y cols., 1999). Hasta 1989, el género *Leptospira* fue dividido en dos especies, *Leptospira interrogans* (cepa patogénica) y *Leptospira biflexa* (cepa saprofita). Los organismos tienen forma espiral, de 5 a 18 micras de longitud y de 0.1 a 0.2 micras de ancho. Las cepas de *L. interrogans* utilizadas para el diagnóstico serológico son cultivadas a la temperatura óptima de 30°C y el tiempo para obtener una nueva generación del cultivo es de 7 a 10 días. Sin embargo, el aislamiento bacteriológico es difícil de recuperar mediante cultivos in vitro (McDonough, 2001).

En la actualidad, los dos principales tipos en el sistema de clasificación del género *Leptospira spp.* se basan en las características fenotípicas y genotípicas (Faine y cols., 1999; Levett 2001). Según la clasificación fenotípica, *Leptospira interrogans* se divide en más de 200 serotipos patógenos y *Leptospira biflexa* en más de 60 serotipos que se definen por aglutinación con antígeno homólogo.

De acuerdo con el sistema de clasificación del genotipo, el género *Leptospira spp.* está dividido en 18 geno-especies basadas en la homología ADN-ADN. La recientemente organización de 18 genomo-especies son: *L. alexanderi*, *L. biflexa*, *L. borgpetersenii*, *L. fainei*, *L. inadai*, *L. interrogans*, *L. kirschneri*, *L. meyeri*, *L. noguchii*, *L. santarosai*, *L. weilii*, *L. wolbachii*, *L. parva* genomo-especies 1, 3, 4 y 5 (Brenner y cols., 1999; Faine y cols., 1999). Sin embargo, ésta clasificación no es útil para fines clínicos debido a que no corresponde al sistema de serogrupos y ambas cepas de leptospirosis tanto patógenas y no patógenas ocurren dentro de la misma especie. De los más de 200 serotipos patogénicos que ocurren en todo el mundo, sólo veintidós pertenecen a dos especies de leptospirosis y se reconocen como endémicas en México.

3.3.-Reservorios y vías de transmisión de leptospirosis.

Las principales fuentes de infección por *Leptospira interrogans* en humanos son los mamíferos salvajes o domésticos y se sugiere que todas las especies conocidas de mamíferos pueden ser reservorios de leptospirosis (Faine et al, 1999). Desde una perspectiva epidemiológica, es importante reconocer la relación existente en la naturaleza de los huéspedes de *Leptospira interrogans* o el estado de reservorio de mantenimiento, la especie animal multiplicadora y el huésped aberrante.

Un reservorio de mantenimiento es aquella especie animal que mantiene la infección perpetuando al patógeno en la población y en ciertas ocasiones, o bien no contrae la enfermedad o sólo existe una leve respuesta serológica, y sólo los animales jóvenes y recién nacidos muestran infección clínica mientras que los adultos son inmunes con manifestaciones subclínicas. Un huésped amplificador es aquel que elimina grandes concentraciones de microorganismos a través

de la orina o de las heces, si la leptospira es expuesta a los animales susceptibles en las excretas la transmisión de la infección es inminente, aunque no pueden mantener la infección en la nueva especie por largo tiempo a menos que la exposición sea constante o intermitente reemplazando la infección a otra especie reservoria. Es así que si se elimina el intercambio de la infección entre los huéspedes, la infección, tarde o temprano desaparecerá de los huéspedes amplificadores de efectos colaterales. Es común que estos huéspedes sufran de la enfermedad mucho más grave que un reservorio de mantenimiento afectando a más grupos de edad como ocurre en la oveja. En el medio rural puede haber varias especies animales amplificadoras de la infección o dentro de esas al menos un animal vertiendo la infección a un segundo animal susceptible y así sucesivamente. Sin embargo, si se elimina la fuente amplificadora de los reservorios, la cascada de efectos colaterales finalmente se eliminará con medidas contra epizooticas de control. En el corto plazo, el control de la leptospirosis en los huéspedes secundarios sólo funcionará si se impide la transmisión a más de un reservorio a partir de vacunación y otras medidas contra epizooticas tales como limpieza y desinfección, disposición de excretas y basuras, reducción de la carga bacteriana en abrevaderos, y el mejoramiento reproductivo que garantice el reemplazo de animales jóvenes. Un huésped aberrante es uno que sólo rara vez es infectado, comúnmente sufre una enfermedad grave y el patógeno por lo general no es excretado para transmitirse a otros huéspedes. El huésped aberrante, por lo tanto carece de importancia en la epidemiología de la enfermedad, pero pueden verse gravemente afectados. La especie humana es un huésped aberrante para la leptospirosis.

Los reservorios de mantenimiento son vitales para la persistencia de leptospiras en el medio ambiente y la presentación de los casos se acentúa durante el ciclo de infección cuando un huésped amplificador transmite la infección a un animal susceptible. En esta accesibilidad, la infección se observa en los animales jóvenes y la perpetuación de la enfermedad o cronicidad de excreción urinaria aumenta con la edad. Los nuevos anfitriones adquieren la infección por contacto directo o indirecto de las especies huéspedes y mientras que la infección a menudo puede ser grave en estos nuevos anfitriones la duración de la excreción de leptospiras suele ser auto limitado en algunas especies excretoras. Asimismo, es posible que algunas especies suelen ser reservorios de serovariedades específicas de *Leptospira interrogans* y los efectos colaterales en los animales enfermos sean debido a la presencia de otros patógenos oportunistas (Faine y cols., 1999; Levett 2001).

Los pequeños mamíferos silvestres son a menudo importantes reservorios y una fuente de infección para los humanos y las especies domésticas como la bovina, porcina y cánidea también lo son, aunque en mayor grado (Cox y cols., 2005). El ganado lechero es un reservorio de serovariedades Hardjo, Pomona y Grippotyphosa (Arteaga y cols., 1991), en los cerdos Pomona, Tarassovi o Bratislava y en perros Canicola (Hathaway, 1981; Faine y cols., 1999). La percepción general de que las ovejas no se consideran reservorios naturales para cualquiera de las serovariedades de *L. interrogans* radica en la carencia de la sintomatología clásica de la infección durante la conjunta explotación pecuaria con bovinos (Blackmore y cols., 1982; Hathaway y cols., 1982; Tooloei y cols., 2008), aunque otros estudios han mostrado algunos indicios de que las

ovejas actúan como huésped multiplicadores para al menos la serovariedad Hardjo (Gordon, 1980; Primos y cols., 1989; Gerritsen y cols., 1994).

Independiente de los ciclos de transmisión entre los animales domésticos, se han descrito también relaciones ecológicas entre los roedores, animales de granja y humanos, donde las serovariedades que normalmente influyen en los animales domésticos son distintos de los presentes en los roedores y otros animales que conviven con los pequeños rumiantes (Matthias y cols., 2008). La tasa de transmisión está influenciado por muchos factores, tales como el grado de contacto entre las especies huéspedes y anfitriones amplificadores, la densidad de población, el tamaño de los rebaños, la edad, la duración y concentración de leptospiras excretadas en la orina y las condiciones ambientales que favorecen el mantenimiento de estos microorganismos, tales como: clima, temperatura, humedad, pH y su composición química del suelo en donde las leptospiras son excretadas por la orina. La mayoría de los brotes de leptospirosis en animales (Smith y Armstrong, 1975) y humanos (Trevejo y cols., 1998; Kupek y cols., 2000; Ashford y cols., 2000; Ganoza y cols., 2006) están asociados con las lluvias y las inundaciones de la superficie terrestre.

La transmisión puede ser directa o indirecta. La transmisión directa de la infección se produce a través del contacto transplacentario, sexual, lactancia de una madre infectada (Ellis y cols., 1986; Faine y cols., 1999) y semen o embriones infectados que son utilizados para la inseminación artificial de vacas o a partir de la fertilización in vitro (Bielanski y Surujballi 1998; Faine y cols., 1999). La transmisión directa es la ruta para la contaminación de los productos lácteos y la infección de los ganaderos, veterinarios, inspectores de carne y los trabajadores de los rastros a través de la manipulación directa de los animales, de sus productos o subproductos. La transmisión indirecta de la infección suele producirse por contacto con aguas superficiales contaminadas, aguas residuales, drenaje de líquidos del matadero, el agua de drenaje y barro y suelo. La principal vía de entrada de las leptospiras es a través de cortes, heridas o abrasiones en la piel y la conjuntiva o las mucosas del tracto digestivo o respiratorio de los animales o humanos. En los deportistas que practican el triatlón, la infección también puede estar mediada a través de la piel intacta después de la inmersión prolongada en aguas contaminadas por leptospiras (Levett, 2001).

3.4.-Patogénesis.

La patogénesis de la leptospirosis es similar, sino idéntica, para cada especie que se sabe de infección, incluida la humana. Las leptospiras no se localizan en el sitio de entrada extendiéndose rápidamente a través de los vasos linfáticos a la circulación sanguínea desde donde se distribuyen a todos los tejidos. La motilidad de estos organismos puede ser un importante aspecto de la patogénesis debido a la capacidad de rapidez, movimiento oscilatorio y proliferación que observan las leptospiras desde el punto de entrada de la infección hasta los órganos afectados tales como el pulmón, hígado, riñones, ojos y cerebro. Las leptospiras se replican de forma exponencial en la sangre y en estos órganos con tiempos de duplicación de aproximadamente 8 horas. Las

leptospiras avirulentas se eliminan del torrente sanguíneo por fagocitosis, mientras que las leptospiras patógenas pueden eludir este proceso (Faine y cols., 1999).

En el reservorio, las especies patógenas de *Leptospira spp.* se diseminan por la vía hematológica al riñón, donde colonizan la superficie apical del túbulo contorneado proximal, lo cual permite que las leptospiras sean vertidas en la orina con la inminente transmisión a nuevos huéspedes (Cheville y cols., 1980; Sterling y Thiermann, 1981; Manktelow y Thompson, 1989; Faine, 1994). El riñón es también el principal órgano blanco en el proceso de la enfermedad, especialmente en los huéspedes hospedadores accidentales como el hombre. Las lesiones primarias implican daños a las paredes de los pequeños vasos sanguíneos dando lugar a la rotura de la integridad de las membranas de las células endoteliales, extravasación de células y hemorragias.

Las hemorragias y defectos de la coagulación son signos característicos de las formas graves de leptospirosis, debido a la trombocitopenia y la fibrinólisis independiente de la coagulopatía intravascular diseminada. Los casos fatales de leptospirosis humana también muestran hemorragia pulmonar, trombocitopenia como resultado de la agregación plaquetaria y la adhesión del patógeno al endotelio vascular, donde puede haber evidencia o no de la presencia del antígeno de *Leptospira interrogans* (Pereira da Silva y cols., 2002).

La desaparición de los organismos de la sangre coincide con la producción de anticuerpos durante la respuesta inmune. Un factor que influye en la severidad de los síntomas proviene del proceso inflamatorio asociado con la producción de complejos inmunes. La infección generalizada en humanos y caballos puede causar uveítis autoinmune. La autoinmunidad también influye en la progresión de las lesiones renales en perros (Faine y cols., 1999). Estudios experimentales en cobayos, se demostró que el antígeno de leptospira es localizado en el intersticio del riñón, mientras que las inmunoglobulinas G (IgG) y complemento 3 (C3) son depositados en los glomérulos y en las paredes de los pequeños vasos sanguíneos renales (Yasuda y cols., 1986). Algunas de las personas sensibilizadas al antígeno muestran albergue de leptospiras en el puerto de la cámara anterior del ojo (sitio aislado e inmunológicamente privilegiado) causando la inflamación por complejos auto-inmunes del tracto uveal, una condición conocida como "luna de la ceguera" de los caballos (Faine y cols., 1999).

Principalmente, la respuesta inmunológica a la leptospirosis es mediada por los anticuerpos siendo específica para serovariedades homólogas o serotipos estrechamente relacionados. Por lo general, la respuesta humoral es de larga duración y de protección, inclusive con títulos serológicos bajos. La inmunidad a la leptospirosis se debe a la interacción contra el antígeno LPS, sin embargo, la relativa eficacia de la vacunación en el ganado bovino y porcino ha demostrado que los anticuerpos contra el LPS por sí sola no son suficientes para la protección de la infección y la enfermedad. La respuesta inmune mediada por las células T, que aunque no se detecta en bovinos infectados naturalmente con la serovariedad Hardjo, puede ser un componente crítico en la protección inducida por la vacuna en el ganado (Faine y cols., 1999).

3.5.-Signos clínicos en ovejas y cabras.

Las epizootias en estas especies son muy raras, especialmente en el ganado caprino. Muchos de los animales afectados aparecen muertos, aparentemente por septicemia (Davidson y Hirsh, 1980). Los animales enfermos presentan los signos y síntomas clínicos tales como: fiebre, anorexia, disnea, alguna ictericia, hemoglobinuria, palidez de las mucosas, infertilidad, nacimiento de crías débiles o muertas y aborto (Perdomo y Garin, 2002). El período de incubación varía desde 4 hasta 6 días. Los signos clínicos observados comúnmente en la leptospirosis ovina están marcados por anorexia, depresión, disnea y fiebre (0.5°C a 2°C por arriba de los 38.5°C) con una duración de 4 a 5 días. La hemoglobinemia y hemoglobinuria con orina de color rojo oscuro maloliente se puede observar en algunos animales (Vermunt y cols., 1994b; Faine y cols., 1999). Los abortos y las muertes prematuras de los productos de la concepción en hembras gestantes y la mortalidad neonatal y de los corderos recién nacidos se producen tras la infección por *L. hardjo* (Beamer 1953; McCaughan y cols., 1980; Ellis y cols., 1984) y *L. pomona* (Smith y Armstrong, 1975; Davidson y Hirsh, 1980). La anemia hemolítica aguda que acompañada por la depresión, disnea, taquicardia y la alta mortalidad de los corderos también han sido reportados durante la infección por *L. pomona* (Smith y Armstrong, 1975; Davidson y Hirsh, 1980; Vermunt y cols., 1994a; Vermunt y cols., 1994b). Desde los estudios por Morse (1957), la fiebre, anemia hemolítica y hemoglobinuria han sido los signos clínicos observados en los estudios experimentales de la leptospirosis ovina. La leptospirosis clínica en el ganado ovino es comúnmente debida a la serovariedad Pomona en Australia (Sullivan, 1974), en América (Beamer 1953; Davidson y Hirsh 1980) y en Nueva Zelanda (Vermunt y cols., 1994a; Vermunt y cols., 1994b). El signo de agalactia también ha sido asociado con la infección de la leptospirosis ovina por la serovariedad Hardjo (McKeown y Ellis, 1986).

La infección subclínica con seroconversión, leptospiruria y nefritis intersticial focal ha sido reportada en la infección experimental de ovejas con las serovariedades Balcanica o Hardjo (Durfee y Presidente, 1979; Hathaway y Marshall, 1979). En un estudio similar, cuando las ovejas fueron infectadas con *L. hardjo*, las manifestaciones clínicas fueron lo suficientemente leves para pasar desapercibidas (Andreani y cols., 1983). Las ovejas infectadas experimentalmente mostraron fiebre por 2 a 6 días y leptospiremia con duración de 2 a 6 días después de 4 a 8 días de inoculado el antígeno la cual fue observada en todos los animales infectados. Las leptospiras fueron aislados a partir de la orina de dos ovejas en el día 35 después de la inoculación, aunque no hubo propagación de la infección entre el grupo de ovejas control. Los corderos de esas ovejas fueron clínicamente normales con altos títulos serológicos por la prueba de aglutinación microscópica. En Nueva Zelanda (Vermunt y cols., 1994b) durante los brotes de la serovariedad Pomona el comportamiento de la enfermedad en corderos fue similar a la observada en el modelo experimental. Las infecciones agudas en ovinos y caprinos se presentan con relativa frecuencia (Faine y cols., 1999).

3.6.-Las ovejas y su relación con otros reservorios.

Los principales reservorios de la leptospira en ambiente urbano son los perros y las ratas, así como los bovinos, porcinos y equinos en el campo. La leptospira se adaptó a "huéspedes reservorios primarios", los cuales comúnmente son animales salvajes. Estas mismas especies de *Leptospira* también se presentan en casi cualquier otro huésped mamífero como "huéspedes reservorio incidentales o accidentales" (McDonough, 2001).

El perro es reservorio primario para la *L. canicola* y la *L. bataviae*. Los perros, además, pueden infectarse con varios serovares como *L. icterohaemorrhagiae* y *L. georgia*, y servir como huéspedes accidentales. Los bovinos son reservorios primarios para la *L. hardjo* y *L. pomona*, y huéspedes accidentales para la *L. grippotyphosa*. Los porcinos son reservorios primarios para la *L. bratislava*, al igual que los equinos y accidentales para la *L. autumnalis*, a diferencia de los equinos que lo son para la *L. pomona*. Los roedores son reservorios primarios para la *L. icterohaemorrhagiae* y accidentales para la *L. pomona*. También pueden ser reservorios primarios algunos animales silvestres como los zorrillos, cabras, conejos y murciélagos, mientras que el hombre es un mal reservorio; hecho que se explica porque los primeros tienen el pH de la orina alcalino, favoreciendo la sobrevivencia de la leptospira. Se reconoce que 1 mL de orina proveniente de bovino infectado por leptospirosis puede contener hasta 100 millones de microorganismos, aunque en el hombre debido a que tiene una orina relativamente ácida para la leptospira esta relación no ha sido estimada adecuadamente sugiriendo que este es un huésped accidental. La excreción de leptospirosis en la orina de los reservorios puede ocurrir por tiempo prolongado en periodos definidos y contaminar el ambiente (Acosta y cols., 1994).

La leptospira no se multiplica fuera del huésped y su supervivencia depende de las condiciones ambientales en las que se encuentren, por ejemplo, condiciones del suelo y agua. El agente patógeno es altamente susceptible a la desecación y a los cambios de pH, mientras un pH menor de 6 y mayor a 8 son inhibidores, las temperaturas por debajo de 10°C (50°F) y las temperaturas mayores a 34°C a 36°C (93-96°F) son nocivas. Los organismos sobreviven hasta 180 días en suelos húmedos, por varios meses en superficies acuosas y sobreviven aún mejor en agua estancada que en movimiento (McDonough, 2001; Acosta y cols., 1994). La mayor prevalencia de la enfermedad ocurre en suelos con pH alcalino, durante las estaciones húmedas (áreas de alta precipitación), en áreas bajas donde es susceptible que la lluvia corra, climas cálidos y húmedos, y áreas con abundante superficie de agua generando campos pantanosos y áreas barrosas.

La infección en humanos y otros reservorios es por contacto directo con orina infectada, material o fluidos fetales y placentarios, descargas uterinas, o por contacto indirecto con el ambiente contaminado (McDonough, 2001; Acosta y cols., 1994). La población a riesgo de leptospirosis comprende la que habita zonas endémicas de los países tropicales subdesarrollados, mientras que en los países desarrollados suele ser una enfermedad profesional de los que trabajan con animales o sus productos, o en medios contaminados especialmente por roedores (veterinarios, ganaderos, tamberos, carniceros, trabajadores de frigoríficos, agricultores, trabajadores de la red de saneamiento, limpiadores de alcantarillas) (Carneiro y cols., 2004). El hombre también puede

infectarse en actividades recreativas al entrar en contacto con agua estancada contaminada (nado, pesca, deportes acuáticos) y por contacto con su mascota. Aerosoles inhalados pueden transportar microorganismos directamente a los pulmones. Los perros en patios cercados pueden exponerse a orina de animales enfermos, incluyendo roedores. Los perros que son ejercitados mediante caminatas en parques, perros de caza y aquellos que deambulan en el campo o nadan en estanques, lagunas y en arroyos con poco y lento caudal están en un riesgo mayor a la exposición a leptospirosis (McDonough, 2001) y por lo tanto los dueños también se encuentran en riesgo, siendo introducidos en la cadena epidemiológica de modo accidental (Carneiro y cols., 2004).

3.7.-Manifestaciones clínicas en humanos.

La leptospirosis se caracteriza por ser una enfermedad febril aguda en humanos y animales y su manifestación varía de la forma subclínica a manifiesta y grave. Independientemente de la manifestación, el estado de portador renal crónico es inevitable y el huésped accidental y el reservorio desempeñan un papel vital en el mantenimiento de leptospiras en el medio ambiente y la epidemiología de la enfermedad. Las leptospiras pueden persistir en los riñones mediante su capacidad para adherirse a las células epiteliales tubulares renales (zona proximal) por períodos de semanas a años siendo eliminadas por la orina de forma continua (en ciertos periodos) o intermitente. En general, el periodo de incubación en animales esta en el rango de 3 a 7 días en la forma aguda. Esta se caracteriza por fiebre, lasitud, pérdida de apetito, irritabilidad, congestión de la conjuntiva y algunas veces diarrea. Después del período de incubación, los síntomas aparecen bruscamente y las manifestaciones de la infección son desde la enfermedad subclínica (común entre los veterinarios y cuidadores de animales) con un cuadro anictérico leve que ocurre en la mayoría de los pacientes (90 a 95%) hasta su forma ictérica severa llamada también enfermedad de Weil en el 5 a 10% de los casos (Acosta y cols., 1994; Songer y Thiermann, 1995; Lilenbaum y cols., 2002).

Forma anictérica presenta 2 fases clínicas:

- Fase 1 o infecciosa: Dura de 4 a 9 días. Existe inicio abrupto de cefalea, fiebre alta, escalofríos y mialgias severas sobre todo de los miembros inferiores, en ocasiones acompañadas de hiperestesias. La anorexia, las náuseas y el vómito ocurren en la mitad o más de los pacientes. Hay compromiso del sensorio (25%), inyección conjuntival, conjuntivitis y exantema que puede ser maculopapular, macular o urticarial. La hepatoesplenomegalia puede ocurrir, pero es más bien rara. Esta fase coincide con la presencia de la leptospira en el líquido cefalorraquídeo (LCR) y en la sangre.
- Fase 2 (inmune). Se presenta después de 2 a 3 días con disminución importante de los síntomas generales y la fiebre (ésta fase corresponde a la aparición de leptospira en la sangre y del LCR) y permanece de 1 a 3 días. Las manifestaciones clínicas son más acentuadas que en la fase 1. Existe la reaparición de fiebre, casi siempre más baja que la fase 1, o puede faltar, y se agrega delirio.

Existe cefalea intensa persistente que no responde a analgésicos comunes. Se encuentran mialgias severas que comprometen en especial los gastronemios, la región paraespinal, el abdomen y el cuello. También hay adenopatías, hepatomegalia, esplenomegalia en 15-25% de los casos y erupción. A nivel ocular y en forma característica se presenta fotofobia, dolor, hemorragia conjuntival e iridociclitis.

Forma ictérica o Síndrome de Weil:

Por lo general su curso clínico es continuo y no bifásico y los síntomas iniciales son similares a la forma anictérica, aunque en el día tercero o sexto post-infección aparece ictericia progresiva con niveles de bilirrubina no mayores de 20 mg/dl y fosfatasa alcalina moderadamente elevada. Existe dolor a la palpación en el hipocondrio derecho, hepatomegalia leve o moderada y elevación de las aminotransferasas que raramente excede 5 veces los valores normales (Acosta y cols., 1994).

La hemorragia generalizada constituye una de las manifestaciones clínicas más notorias de esta forma de enfermedad, epistaxis, sangrado del tracto gastrointestinal, hemorragias pulmonares que resultan en infiltrados pulmonares hasta en el 40% de los casos (Bal, 2005). Rara es la presentación ictérica sin insuficiencia renal que desarrollan neumonía bilateral grave seguida de distress respiratorio (Seijo y cols., 2002). La tendencia hemorrágica se puede explicar por la vasculitis generalizada a nivel de las glándulas suprarrenales y del sistema nervioso central, la trombocitopenia presente hasta en 50% de los casos y en menor grado por la hipotrombinemia. El compromiso de la función renal debido principalmente a una necrosis tubular aguda puede llevar al paciente a una azoemia severa, recurriéndose en la mayoría de estos casos a diálisis peritoneal o hemodiálisis (Acosta y cols., 1994).

El compromiso miocárdico es una complicación frecuente en la fase tardía del síndrome de Weil, pero con pocas repercusiones hemodinámicas. Watt y colaboradores (1990) estudiaron 38 pacientes con síndrome de Weil y observaron 39% de anomalías electrocardiográficas, bloqueo atrio ventricular (AV) de primer grado y cambios sugestivos de pericarditis aguda como lo más frecuente. También se han informado alteraciones más raras, como fibrilación auricular con falla aguda del ventrículo izquierdo. El colapso cardiovascular por lo general es la causa de muerte en estos pacientes. En el desarrollo de esta complicación se han implicado factores tipo endotoxinas aún no estudiados (Acosta y cols., 1994).

3.8.-Diagnóstico diferencial.

En general se deben considerar los cuadros de hepatitis viral, meningitis aséptica, síndromes virales, fiebres de origen desconocido, fiebre amarilla, malaria o dengue. Otras enfermedades pueden eventualmente ser confundidas con la leptospirosis como colecistitis, infecciones respiratorias, sarampión, rubéola, pielonefritis, brucelosis con ictericia y endocarditis, por lo que el examen clínico epidemiológico será de gran ayuda en el diagnóstico definitivo (Laguna, 2000; Bal, 2005).

La presencia de mialgias en los gastronemios, la inyección conjuntival, el dolor ocular, la presencia de aumento en la creatininfosfoquinasa (CPK) y el valor de enzimas hepáticas elevadas, en promedio 5 veces de lo normal (aunque puede ser más), son signos a tener en cuenta, así como la historia epidemiológica, el tipo de ocupación, la evolución del cuadro, el gran compromiso general, una ictericia y la sospecha clínica, hacen pensar en el diagnóstico, lógicamente descartando las otras entidades antes mencionadas.

Los diagnósticos diferenciales de enfermedad peraguda o aguda en el perro incluyen enfermedad por parásitos cardíacos (dirofilariosis), anemia autoinmune hemolítica, bacteremia (debido a heridas por mordedura, prostatitis, enfermedad dental), hepatitis infecciosa viral canina, neoplasia hepática, trauma, lupus, fiebre maculosa de las Montañas Rocosas, ehrlichiosis, toxoplasmosis, neoplasia renal y cálculos renales. Los diagnósticos diferenciales de enfermedad crónica, por ejemplo, aborto, síndrome del cachorro débil, incluyen brucelosis canina, infección canina por herpesvirus y distemper (McDonough, 2001).

CAPÍTULO 4. LEPTOSPIROSIS EN OVEJAS: METODOLOGÍA Y DESCRIPCIÓN DEL ESTUDIO

En este capítulo se detalla el estudio del cual este trabajo toma su nombre, así como la metodología, objetivos y las hipótesis que se plantean.

4.1.-Planteamiento del problema y justificación del estudio

Los cambios climáticos y las enfermedades emergentes son los nuevos retos para el siglo XXI. La leptospirosis es una enfermedad infecciosa de los animales, de origen bacteriano con implicaciones zoonóticas (Faine, 1982). La infección humana es causada por miembros del género *Leptospira spp.*, un grupo de espiroquetas que subsisten en animales domésticos y silvestres (especialmente ratas y ocasionalmente perros). Millones de personas son infectadas cada año a partir del estrecho contacto con los animales enfermos y con el agua, alimentos o tierra contaminada de orina de estos animales pudiendo penetrar al huésped a través del tejido intacto o por inoculación traumática (Bharti y cols., 2003). Primariamente, la leptospirosis ha sido considerada como una enfermedad ocupacional, pero en años recientes relacionada con actividades de ocio, tales como el ecoturismo y la natación, y con desastres naturales asociados con un aumento de la precipitación pluvial y el estancamiento de residuos sólidos con materia orgánica contaminada (Bovet y cols., 1999; Everard y cols., 1992). Se han descrito más de 200 serovariedades diferentes del género *Leptospira*, cada una de ellas siendo mantenida por largos periodos de tiempo en hospederos específicos o bajo condiciones ambientales que les permitan sobrevivir fuera del huésped (Levett, 2001). La mayoría de los casos humanos no manifiestan la enfermedad clínica y otros desarrollan un tipo de influenza o "gripe" que desaparece en pocos días. Sin embargo, del 5 al 10% de esas personas infectadas progresa a una segunda fase, algunas veces fatal. Es normalmente caracterizada por ictericia, problemas renales y un alargamiento de la proyección anatómica del bazo (enfermedad de Weil), pero también puede involucrar a los pulmones. Inclusive, la infección por leptospiras ha sido sugerida como una posible causa de insuficiencia renal crónica (Scott y Coleman, 1996).

Además de ser un problema de salud pública, la leptospirosis animal es una enfermedad reproductiva que ha sido considerada en todo el mundo como responsable de estar involucrada con la disminución de la producción de leche y carne, y en la productividad pecuaria, debido a la infertilidad y el aborto, así como de la reducción del número de animales para reemplazo que constituyen el inventario ganadero nacional (Genovez y cols., 2001; Guitián y cols., 1999).

Si bien los principales reservorios de *Leptospira interrogans* en los entornos urbanos son los perros y ratas, así como los bovinos, porcinos y equinos, pero no las ovejas en las áreas rurales, la leptospira se adapta a "huéspedes reservorios primarios", los cuales comúnmente son animales

silvestres. Además las serovariedades de *L. interrogans* también se presentan en casi cualquier otro huésped mamífero como "huéspedes reservorio incidentales o accidentales" (McDonough, 2001; Perret y cols., 2005).

El perro es reservorio primario para *L. canícola* y *L. bataviae* y puede servir como huésped accidental de otras serovariedades tales como: *L. icterohaemorrhagiae* y *L. georgia*. Los bovinos son reservorios primarios para la *L. hardjo* y *L. pomona*, aunque huéspedes accidentales para la *L. grippotyphosa*. Los porcinos son reservorios primarios para la *L. bratislava*, al igual que los equinos y accidentales para *L. autumnalis*, a diferencia de los equinos que lo son para la *L. pomona*. Los roedores son reservorios primarios para la *L. icterohaemorrhagiae* y accidentales para la *L. pomona*. Los ovinos y caprinos son reservorios primarios para *L. ballum* y *L. hardjo*, y accidentales para *L. pomona*. La cepa de *L. hardjo*, una vez introducida en la población ovina se establece con diferentes niveles de infección que pueden persistir durante largos periodos de tiempo. La especie humana no actúa como hospedero de mantenimiento de ninguna serovariedad de *L. interrogans* y la infección en ésta especie será siempre considerada accidental (Sullivan, 1974; Heath y Johnson, 1994).

También pueden ser reservorios primarios algunos animales silvestres como los zorrillos, cabras silvestres, conejos y murciélagos, mientras que el hombre es reservorio no adecuado de mantenimiento; hecho que se explica porque las especies no humanas tienen el pH de la orina alcalino, favoreciendo la sobrevivencia de la leptospira. Así se sabe que 1 ml de orina de bovino puede contener hasta 100 millones de microorganismos, a diferencia del hombre que tiene una orina relativamente ácida para la bacteria, constituyéndose como un huésped accidental. La excreción de la leptospira en la orina de los reservorios naturales puede ocurrir por periodos prolongados y contaminar el ambiente (Acosta, 1994).

Es por ello que, algunas prácticas laborales, tales como el trabajo agrícola en los cultivos de arroz, así como en la ganadería (veterinarios, ganaderos, trabajadores de rastro), y bajo ciertas actividades recreativas que implican contacto con aguas posiblemente contaminadas (balnearios en medios naturales y deportes acuáticos), suponen importantes escenarios de alto riesgo de contagio. Además del riesgo sanitario, se deberá tomar en cuenta la vertiente económica derivada de los gastos originados por el cuidado médico de los pacientes, las bajas laborales, la pérdida de productividad y capacidad de trabajo, vigilancia y control de los lugares de trabajo, ropas especiales de protección, y los seguros médicos para el personal de alto riesgo (Faine, 1991).

Se espera que los resultados de este estudio serológico de campo en las ovejas puedan proporcionar la valiosa información sobre la distribución geográfica de los numerosos focos y la posible propagación de la enfermedad sirvan de base a los médicos veterinarios oficiales en la planificación y ejecución de los programas de control de la leptospirosis en el Estado de México y en otros lugares. La leptospirosis es una grave amenaza para la salud pública. La población humana es propensa a la infección por muchas enfermedades de animales llamadas infecciones zoonóticas. La leptospirosis se presenta tanto en climas templados y tropicales, así como los entornos urbanos y rurales, por lo que es la enfermedad re-emergente y zoonótica más extendida.

4.2.-Objetivos del estudio

1. Estimar la prevalencia de anticuerpos frente a determinadas serovariedades de *Leptospira interrogans* entre las ovejas de un área de producción pecuaria del Estado de México.
2. Determinar las serovariedades de *Leptospira interrogans* y la distribución de los títulos serológicos presentes en los animales.
3. Investigar las relaciones entre la respuesta serológica a este patógeno y los factores climáticos regionales, tamaño y condiciones de rebaño, el acceso de los animales a los suministros de agua contaminada, las políticas de reemplazo, las características del manejo de rebaño y la historia de abortos.

4.3.-Hipótesis de investigación

1. La estimación de la seropositividad anti-*Leptospira* en la población ovina de una región montañosa del Estado de México estará dentro o será cercana al rango estimado en otras especies productivas de diferentes regiones de México (19-47%).
2. La distribución de los títulos serológicos en los animales permitirá establecer a la especie ovina como reservorio primario o accidental para ciertas serovariedades de *Leptospira interrogans*.
3. Algunos de los factores presentados en la literatura como posibles "factores de riesgo" estarán relacionados con la presencia de anticuerpos anti-*Leptospira*.

4.4.-Población objetivo

Ovejas hembras en edad reproductiva que pertenezcan a un rebaño dentro de la región en estudio (que será definida previamente).

4.5.-Tipo de estudio

En todo estudio es importante plantear la metodología que se usará, en base a los objetivos de la investigación y siempre con la finalidad de obtener información y resultados que nos conduzcan al apoyo o rechazo de nuestras hipótesis. La elección de alguna o algunas metodologías o procesos influye en los métodos cuantitativos o cualitativos que pueden ser considerados para ser usados en el estudio.

Las diferentes investigaciones se clasifican de acuerdo al propósito, la evolución de ésta, la fuente de información y el control del investigador sobre las unidades por estudiar. De acuerdo a estos criterios esta investigación puede ser clasificada de la siguiente manera:

- *De acuerdo al propósito: DESCRIPTIVA*. Uno de los propósitos es “medir” la seroprevalencia de anticuerpos para la *Leptospira interrogans* en ovejas hembras de cierta región del Estado de México, con lo cual sólo se trata de describir una de las características de la población en cuestión. En el segundo objetivo se plantea estudiar si existe o no relación entre la presencia de seropositividad y algunas variables. Como consecuencia de lo anterior, de que sólo se cuenta con una población y de que no hay una hipótesis de causalidad la investigación es de carácter descriptiva. Al tener un estudio de este tipo no se presentan factores de confusión ya que no estamos comparando alguna característica entre distintas poblaciones y por lo tanto no hay contaminantes de nuestra respuesta.
- *De acuerdo a la evolución: TRANSVERSAL*. El periodo de estudio fue entre noviembre de 2003 a marzo del 2005. Las variables sólo fueron medidas en una sola ocasión, entonces la evolución de las mismas no puede ser estudiada, es decir, este análisis es un tipo de “fotografía”, sólo describe las características de la población en el lapso de la investigación.
- *De acuerdo a la fuente de información: PROSPECTIVA*. La información se obtuvo mediante la aplicación de un cuestionario a los propietarios de los rebaños y del análisis del suero extraído de cada una de las ovejas seleccionadas.
- *De acuerdo al control del investigador: OBSERVACIONAL*. Los investigadores no tuvieron control alguno sobre las variables en cuestión, es decir éstas no pudieron ser asignadas aleatoriamente a cada uno de los elementos; por ejemplo a una oveja no se le puede asignar al azar su raza, sino que ésta ya se posee y uno sólo se puede limitar a observar y registrar esta característica del animal.

En resumen, tomando en cuenta los cuatro criterios de clasificación anteriores, al presente estudio de prevalencia de anticuerpos anti-*Leptospira* se le da el nombre de *encuesta descriptiva*. En este tipo de investigación no hay factores de confusión y por lo tanto la variable respuesta no estará contaminada, de ahí que el concepto de validez interna se cumple, sin embargo esto no garantiza que la validez externa (representatividad, capacidad de extrapolar los resultados) sea alcanzada, para esto hay que recurrir a las técnicas de muestreo. Al ser un trabajo de tipo prospectivo aumenta la posibilidad de que la información obtenida sea de utilidad y además que ésta sea de mejor calidad que si hubiera sido obtenida de forma retrospectiva. Un punto muy importante es que al ser un estudio observacional no se puede hablar de causalidad, es decir que si se

encuentran relaciones entre los factores considerados y la seropositividad, éstas no serán de tipo causal, sino estadísticas (Méndez 2002).

4.6.-Lugar de estudio

La investigación fue realizada en un área de 845.2 km² en la región sureste del Estado de México, que comprende tres regiones geográficas: Valle, intermontaña y montaña (19°05'-19°15' N, 99°20'-99°35' W). Las características topográficas del lugar de estudio mostraron un extenso terreno llano a 2670 metros sobre el nivel medio del mar (m.s.n.m.m), además de colinas, montañas y pendientes que van desde 2890 a 3273 m.s.n.m.m. Las fuentes de agua y humedad son pequeños lagos, abrevaderos y canales de irrigación. La temperatura mínima y máxima fueron 3.7°C y 20.2°C respectivamente (temperatura media de 11.9°C) con una precipitación media anual de 991.9 mm.

En este lugar existe una amplia distribución de rebaños trashumantes en donde un fuerte impacto ambiental por la población humana no existe. El sistema de producción extensiva ovina trashumante representa una población de animales de especial interés para la economía mexicana debido a la falta de recursos económicos designados para desarrollar sistemas adecuados de la producción ganadera rural. Como resultado, los productores de ovejas u ovino cultores se ven obligados a movilizar a sus rebaños en función de los alimentos y la disponibilidad de agua, donde los pastos se comparten con otras manadas y animales domésticos y silvestres. La imagen 4.1 es una muestra del lugar de estudio.

4.7.-Diseño muestral

La información sobre el total de animales y de rebaños fue obtenida del censo ovino del año 2000 el cual fue facilitado por la Asociación Local de Trabajadores de Ovejas. Éste fue considerado como el marco de muestreo. En el censo estaban registrados 3.762 ovejas y 3.818 carneros y corderos. Las ovejas eran principalmente raza Hampshire o de un cruce con Suffolk y Pelibuey, todos éstos distribuidos en setenta y cinco rebaños.

El esquema de muestreo utilizado es probabilístico, bietápico y estratificado. El tamaño de rebaño fue la variable sobre la cual se formaron los estratos. Éstos se muestran en la tabla 4.1.

Tabla 4.1 Definición de estratos. Rebaños

Estrato 1	Menos de 50 animales
Estrato 2	51-140 animales
Estrato 3	Más de 141 animales

Las unidades primarias y secundarias de muestreo fueron los rebaños y las ovejas respectivamente.

Imagen 4.1 Muestra del lugar de estudio.



Tamaño de muestra

Debido a la falta de información concerniente a la seroprevalencia de *Leptospira interrogans* en los rebaños, un estudio preliminar se realizó para estimar puntos de corte, la estimación de estos se muestra en la tabla 4.2.

Tabla 4.2 Porcentaje de ovejas seropositivas por estrato en el estudio piloto.

Estrato	Porcentaje
1	27.8
2	39.4
3	44.4

Los setenta y cinco rebaños que estaban registrados en el censo fueron tomados en cuenta. La asignación de tamaño de muestra a cada estrato fue de manera proporcional. Las manadas fueron seleccionadas al azar utilizando un generador de números aleatorios informatizado (Epistat 1987, 1989, EE.UU.). Para un nivel de confianza del 95% y el establecimiento de límites de error del 5%, aproximadamente un décimo de los animales (o todas las ovejas en rebaños con menos de 10 elementos) se muestrearon aleatoriamente utilizando una calculadora de números aleatorios en cada manada.

Los 35 rebaños incluidos en la muestra estuvieron distribuidos en toda la zona en estudio; 8 en el estrato 1, 17 en el estrato 2 y 10 en el estrato 3. El número de unidades últimas (secundarias) de muestreo fue de 367 animales.

4.8.-Registro de los participantes

Debido a que el diagnóstico serológico inicial de la enfermedad clínica se atribuyó a enfermedades bacterianas zoonóticas, intensos esfuerzos se hicieron para maximizar la participación de los propietarios. El objetivo del estudio y su confidencialidad fueron descritos a la Asociación local de ovino cultores. Una carta de invitación junto con una solicitud para la toma de muestras fue proporcionada a cada uno de los propietarios seleccionados. El formato de entrevista, con base en una revisión de la literatura, incluyó preguntas sobre la leptospirosis, además de otras enfermedades de las ovejas con carácter zoonótico como brucelosis y aborto enzoótico ovino (Leal-Castellanos y cols., 2003;).

4.9.-Selección de la Muestra

Al ser el diseño de muestreo bietápico, primero se seleccionaron aleatoriamente las unidades primarias (Rebaños) dentro de cada estrato. A los propietarios de los rebaños seleccionados se les envió la invitación al proyecto como se indicó anteriormente. Todos aceptaron participar en el estudio.

Las unidades secundarias de muestreo (ovejas) también fueron seleccionadas aleatoriamente mediante el mismo generador de números aleatorios. Los criterios de inclusión y exclusión fueron los siguientes:

Criterios de inclusión

Las ovejas hembra sanas clínicamente, en edad reproductiva, de las razas hampshire o su cruce genético con suffolk y pelibuey, así como todos los ovinos que estuvieron distribuidos en los treinta y cinco rebaños seleccionados.

Criterios de exclusión

Animales enfermos o aquellos destinados para sacrificio.

4.10.-Pruebas serológicas

A las ovejas seleccionadas que cumplieron con los criterios de inclusión se les extrajo una muestra de sangre (5 ml). Los sueros fueron evaluados para la detección de anticuerpos para las serovariedades de *Leptospira* usando la prueba de aglutinación microscópica (MAT por sus siglas en inglés). Se utilizaron Antígenos vivos de serovariedades de *Leptospira interrogans* como Bratislava, Pyrogenes, Grippotyphosa, Pomona, Wolffi, Tarassovi y las cepas de campo Icterohaemorrhagiae, Hardjo, y Canicola que fueron aisladas de animales en diferentes regiones de México. Los cultivos vivos de *Leptospira* requeridos para ser usados como antígenos se mantuvieron en un medio Ellinghausen-McCullough-Johnson-Harris, con un 10% de albúmina de suero bovino. La concentración de antígeno se ajustó a 1 a 2×10^8 organismos /mL por espectrofotometría (400 nm de longitud de onda). El MAT fue leído por microscopía de campo oscuro. El punto final fue la mayor dilución del suero a los cuales el 50% de aglutinación se produce.

Imagen 4.2 Momento de toma de pruebas.



4.11.-Análisis estadístico

Para la determinación de factores de riesgo se ajustó un modelo de regresión logística múltiple con efectos mixtos. La estimación de la prevalencia de *Leptospira interrogans* en la población objetivo fue estimada tomando en cuenta la estructura del diseño de muestreo. Todos los análisis estadísticos en este trabajo fueron realizados con el paquete estadístico STATA 11, los cuales se detallarán en el siguiente capítulo.

CAPÍTULO 5. LEPTOSPIROSIS EN OVEJAS: APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA DE EFECTOS MIXTOS PARA LA DETECCIÓN DE FACTORES DE RIESGO

En este capítulo se detallan los análisis estadísticos realizados. En primer lugar se describen los datos obtenidos mediante la metodología reportada en el capítulo anterior, las características de las variables respuesta y explicativas. Posteriormente se analizarán los procedimientos de estimación de la seroprevalencia de *Leptospira interrogans* en las ovejas seleccionadas y los factores de riesgo asociados a ésta. Finalmente se verificará si el modelo es adecuado y se interpretarán los coeficientes estimados.

5.1.-Descripción de la base

La base de datos cuenta con 42 variables cada una de las cuales corresponde a una pregunta del cuestionario que se realizó a los propietarios de los rebaños o son resultado de las pruebas realizadas a las ovejas.

Debido al diseño del estudio (bietápico) y a que la mayoría de los datos son resultado del cuestionario, todas las ovejas de un rebaño comparten el mismo valor en ciertas variables; por ejemplo, cada uno de los animales de la manada X se encuentran localizados en el municipio Y. A este tipo de factores se les llamará de nivel II o de rebaño. Cuando cierta característica cambia de acuerdo al elemento de estudio (por ejemplo la raza), se le considerará como una variable de nivel I o de oveja.

Cabe señalar que las variables de nivel I y II corresponden a las unidades secundarias y primarias de muestreo respectivamente. Este cambio de notación (que se espera no sea motivo de confusión) es hecho para quedar de acuerdo a la usada en los modelos multinivel (Fielding, Pillinger) (los GLMM son miembros de esta familia aunque a veces son llamados indistintamente).

La tabla 5.1 muestra el total de variables por nivel y tipo de escala de medición. Se observa que la mayoría de ellas son de nivel II (37 en total). Los factores de tipo binario son los que aparecen con mayor frecuencia en la base de datos.

Las variables incluidas en la base de datos corresponden a las preguntas o pruebas de la siguiente manera:

- Rebaño: Identificador de la manada a la que pertenece la oveja.
- Estrato en el cual se encuentra clasificado el rebaño.
- Leptospira: Estado de seropositividad antileptospira en la oveja.

- Región en la que habita el rebaño.
- Chlamy: Estado de seropositividad de *Chlamydophila abortus*.
- Municipio en el que habita el rebaño.
- Origen de las ovejas.
- Objetivo de la explotación.
- Raza
- Sistema d explotación.
- Matcorral: Material del que está hecho el corral donde habitan los animales.
- Piso: Tipo de piso del corral.
- Cama: ¿Utilizan cama los animales?
- Luparición: ¿En dónde paren las ovejas?
- Disagua: Suministro de agua.
- Pastoreo: ¿Pastorean los animales?
- Donde: ¿Dónde pastorean?
- Separto: ¿Separan a los animales durante el parto?
- Encierro: Número de animales por corral de encierro.
- Limpcorr: ¿Limpian los corrales?
- Freclimp: Frecuencia de limpieza.
- Excremento: ¿Qué sucede con el excremento que se recoge?
- Conviven: ¿Durante el pastoreo los animales conviven con animales de otros rebaños?
- Introduce: ¿Ha introducido animales en los últimos 6 meses?
- Cuarentena: ¿Realiza cuarentena de los animales que introduce?
- Sistrepr: Sistema de reproducción.
- Auxparto: ¿Auxilian a las ovejas durante el parto?
- Dxgest: ¿Realiza diagnóstico de gestación?
- Guantes: ¿Utiliza guantes cuando auxilia a las ovejas?
- Quien: ¿Quién realiza el auxilio a las ovejas?
- Ayudan: ¿Le ayudan a atender los partos?
- Placenta: ¿Qué sucede con la placenta?
- Sepantpar: ¿Separan las ovejas antes del parto?
- Aborto: ¿Han abortado sus ovejas?
- Contacto: ¿Ha estado usted en contacto con abortos?
- Partprem: ¿Han presentado partos prematuros las ovejas?
- Placabort: ¿Qué sucede con las placentas y los fetos abortados?
- Mujeres: ¿Hay mujeres trabajando en la explotación?
- Actmujeres: ¿Cuáles son las actividades que realizan las mujeres?
- Embamujer: ¿Estando embarazadas continúan realizando estas actividades?
- Bovis: Estado de seropositividad de la bacteria *Brucella ovis*.
- Babortus: Estado de seropositividad de la bacteria *Brucella abortus*.

Tabla 5.1 Total de variables por tipo de escala y nivel

NIVEL	TIPO	NÚMERO DE VARIABLES
I	Numérica	0
	Catagórica binaria	4
	Catagórica 2+	1
II	Numérica	2
	Catagórica binaria	31
	Catagórica 2+	4

Tabla 5.2 Descripción de las variables

Variable	Nivel	Tipo	Categorías	Codificación	% de muestra
Rebaño	II	Numérica	1-35.	1-35.	-
Estrato	II	Categorica	1	1	14.71
			2	2	53.95
			3	3	31.34
Leptospira	I	Binaria	Negativo	0	46
			Positivo	1	54
Región	II	Categorica 2+	Valle	1	64.31
			Inter montaña	2	26.43
			Montaña	3	9.26
Chlamy	I	Categorica 2+	Negativo	0	74.93
			Positivo	1	21.53
			Sospechosos	2	3.54
Municipio	II	Categorica 2+	Xalatlaco	1	63.49
			Santiago	2	10.9
			Calpuluah	3	20.71
			Chapultepec	4	4.9
Origen	I	Categorica 2+	Propio	1	14.71
			Importadas	2	9.26
			Mercado rural	3	76.02
Objetivo	II	Binaria	Reproducción	0	80.65
			Engorda	1	19.35
raza	I	Categorica 2+	Peliquey	1	6.27
			Hampshire	2	46.59
			Sulfock-peliquey	3	4.9
			Peliquey-Hampshire	4	16.62
			Sulfock	5	25.61

Tabla 5.2 Descripción de las variables (continuación).

Variable	Nivel	Tipo	Categorías	Codificación	% de muestra
Sistema	II	Binaria	Intensivo	0	88.56
			Trashumante	1	11.44
Matcorral	II	Binaria	Tabique	0	32.43
			Malla	1	67.57
Piso	II	Binaria	Tierra mas cemento	0	18.26
			Tierra	1	81.74
Cama	II	Binaria	No	0	83.38
			Sí	1	16.62
Luparición	II	Binaria	Pradera	0	89.37
			Paridero	1	10.63
Disagua	II	Binaria	Tapwater	0	43.32
			Lago	1	56.68
Pastoreo	II	Binaria	No	0	7.63
			Sí	1	92.37
Donde	II	Binaria	Pecebre	0	7.63
			Pradera	1	92.37
Separto	II	Binaria	No	0	16.89
			Sí	1	83.11
Encierro	II	Numérica	1-20.	1-20.	
Limpcorr	II	Binaria	No	0	6.54
			Sí	1	93.46
Freclimp	II	Binaria	1 vez por semana	0	63.22
			más de 1 por semana	1	36.78
Excremento	II	Binaria	Lo riega como abono	0	16.08
			Lo aglomera cerca del corral	1	83.92
Conviven	II	Binaria	No	0	63.49
			Sí	1	36.51
Introduce	II	Binaria	No	0	74.11
			Sí	1	25.89
Cuarentena	II	Binaria	No	0	5.72
			Sí	1	94.28
Sistrepr	II	Binaria	Empadre no controlado	0	25.34
			Empadre controlado	1	74.66
Auxparto	II	Binaria	Sí	0	16.08
			No	1	83.92

Tabla 5.2 Descripción de las variables (continuación).

Variable	Nivel	Tipo	Categorías	Codificación	% de muestra
Dxgest	II	Binaria	No	0	96.19
			Sí	1	3.81
Guantes	II	Binaria	No	0	12.26
			Sí	1	87.74
Quien	II	Binaria	Encargado	0	63.22
			Dueño	1	36.78
Ayudan	II	Binaria	No	0	30.52
			Sí	1	69.48
Placenta	II	Binaria	La dejan en la pradera	0	18.26
			La recogen	1	81.74
Sepantpar	II	Binaria	No	0	20.71
			Sí	1	79.29
Aborto	II	Binaria	No	0	17.17
			Sí	1	82.83
Contacto	II	Binaria	No	0	23.98
			Sí	1	76.02
Partprem	II	Binaria	No	0	30.25
			Sí	1	69.75
Placabort	II	Binaria	No	0	79.29
			Sí	1	20.71
Mujeres	II	Binaria	No	0	58.86
			Sí	1	41.14
Act mujeres	II	Binaria	Limpieza	0	52.04
			Pastoreo	1	47.96
Embamujer	II	Binaria	No	0	92.1
			Sí	1	7.9
Bovis	I	Binaria	Negativo	0	96.1
			Positivo	1	3.9
Babortus	I	Binaria	Negativo	0	91.15
			Positivo	1	8.85

La tabla 5.2 muestra la descripción de cada una de las variables, el nombre que se le asignó, nivel, tipo y categorías, así como la codificación que se usará en el análisis y el porcentaje que cada valor de los factores ocupó de la muestra. Únicamente las variables Bovis y Babortus presentan datos faltantes, 59 y 62 respectivamente. La variable "Donde" no será tomada en cuenta ya que todos los rebaños donde se respondió que las ovejas pastaban también se contestó que lo hacían en la pradera, siendo entonces información redundante.

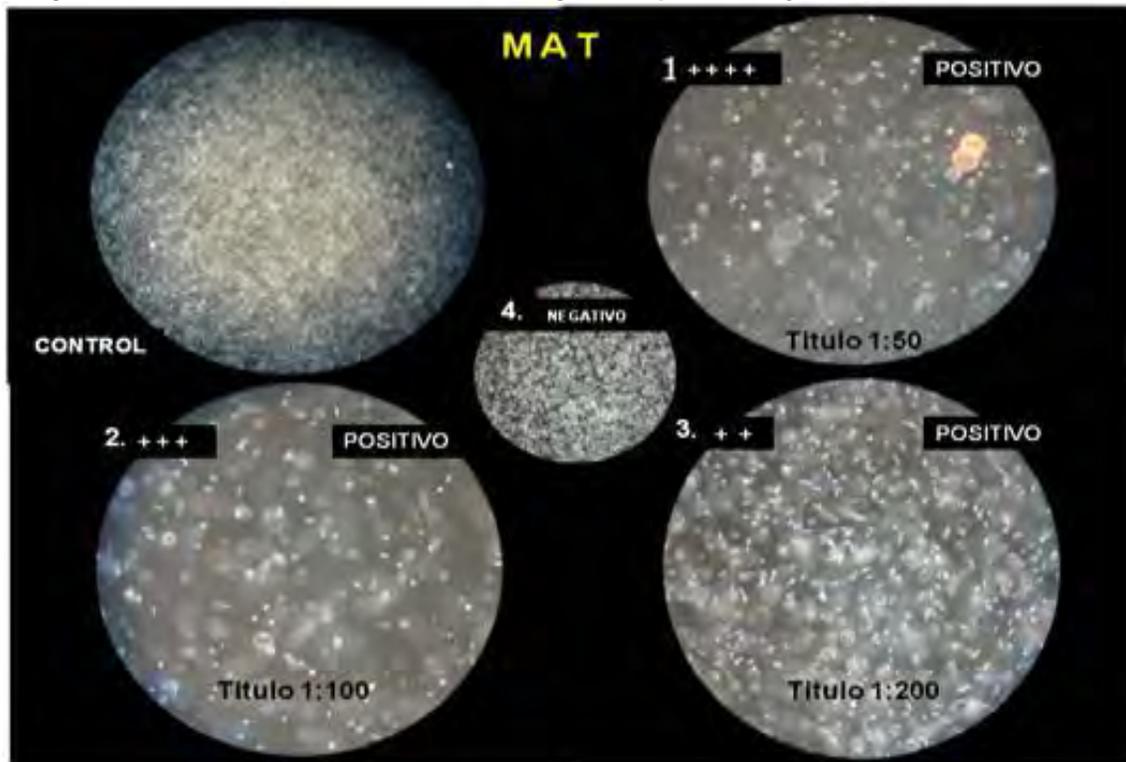
De entre todas las variables hay que destacar tres: Estrato, Rebaño y Leptospira, siendo estas las indicadoras del diseño del estudio: manera de estratificar (tabla 4.1), criterio de agrupamiento y variable respuesta respectivamente. Estos tres factores formaran parte fundamental en esta investigación.

5.2.-Variable Respuesta o de interés

Tanto para la estimación de seroprevalencia como para la construcción del modelo de regresión logística, la variable de interés es Leptospira, ésta es una indicadora de la presencia o ausencia de anticuerpos de la bacteria en cuestión. Una oveja era considerada positiva si bajo la prueba de aglutinación microscópica MAT (sección 4.10) se encontraban títulos mayores o iguales a 1:100 para cualquiera de las serovariedades.

Leptospira $\begin{cases} 0 \text{ negativa bajo la prueba MAT} \\ 1 \text{ positiva bajo la prueba MAT} \end{cases}$

Imagen 5.1 Estado de las muestras cuando son negativas o positivas bajo MAT.



En la imagen 5.1 se muestran los diferentes niveles de aglutinamiento y el estado serológico de diferentes muestras.

5.3.-Estimación de la seroprevalencia

En la tabla 5.2 se observa que el 54% de las ovejas fueron positivas. Éste número puede ser una estimación de la seroprevalencia, sin embargo no se está tomando en cuenta el esquema de muestreo. Usualmente el cálculo de las cantidades de interés no difiere mucho si despreciamos el diseño (puede ocurrir que sí), sin embargo el valor de las varianzas de los estimadores sí puede cambiar drásticamente. En el paquete STATA se puede definir el tipo de muestreo mediante el comando *Svysset* y así las proporciones y las varianzas estimadas estarán tomando en cuenta a éste.

Tabla 5.3 Estimación de seroprevalencia (%) de *Leptospira interrogans* respetando el diseño de muestreo.

Valle			Intermontaña		Montaña	
52.5 (44-61)			55.6 (46-65)		57 (42-71)	
Estrato 1	Estrato 2	Estrato 3	Estrato 1	Estrato 2	Estrato 1	Estrato 2
37.4 (18-57)	57.2 (42-72)	50.7 (39-61)	24.5 (13-35)	60.2 (51-59)	45.2 (11-79)	62.2 (45-78)
seroprevalencia general: 53.4 (47-60)						

Los Intervalos al 95% de confianza se encuentran entre paréntesis

En la tabla 5.3 se indica que la estimación puntual de la proporción de seropositivos es de 0.53, aproximadamente 1% de diferencia de la cantidad calculada cuando se ignora el diseño. En el extremo inferior del intervalo al 95% de confianza se observa que la prevalencia está casi en el límite de los valores observados en otras regiones del país (47%) y tal vez hasta muy por encima (60%).

Mediante el comando *estat effects*, se calculó el efecto del diseño (DEFF) (Cochran 1976). Éste indica cuan mayor o menor es la varianza del estimador con respecto a si se hiciera un muestreo aleatorio simple (m.a.s); el DEFF fue de 1.62295, el cual dice que el diseño es menos eficiente (por lo regular esto siempre ocurre en los diseños por conglomerados), sin embargo no hubiera sido posible realizar un m.a.s debido a la cantidad de tiempo que se invertiría. Este efecto podría ser ocupado para corregir las pruebas de hipótesis en los modelos que no toman en cuenta el esquema de muestreo, sin embargo no es el objetivo de este trabajo.

El cálculo de seroprevalencia para las subpoblaciones de las diferentes partes del lugar de estudio son importantes ya que indican que la *Leptospira interrogans* se está adaptando a altitudes en las cuales no se creía que podía sobrevivir (3273 m.s.n.m en algunas zonas) y que en las tres regiones está distribuida de manera más o menos homogénea (52%, 55%,57% para valle, intermontaña y montaña respectivamente). Para el estrato 2 se encontraron estimaciones puntuales más altas.

5.3.1.- Determinación de las serovariedades y títulos anti *Leptospira* presentes en los animales

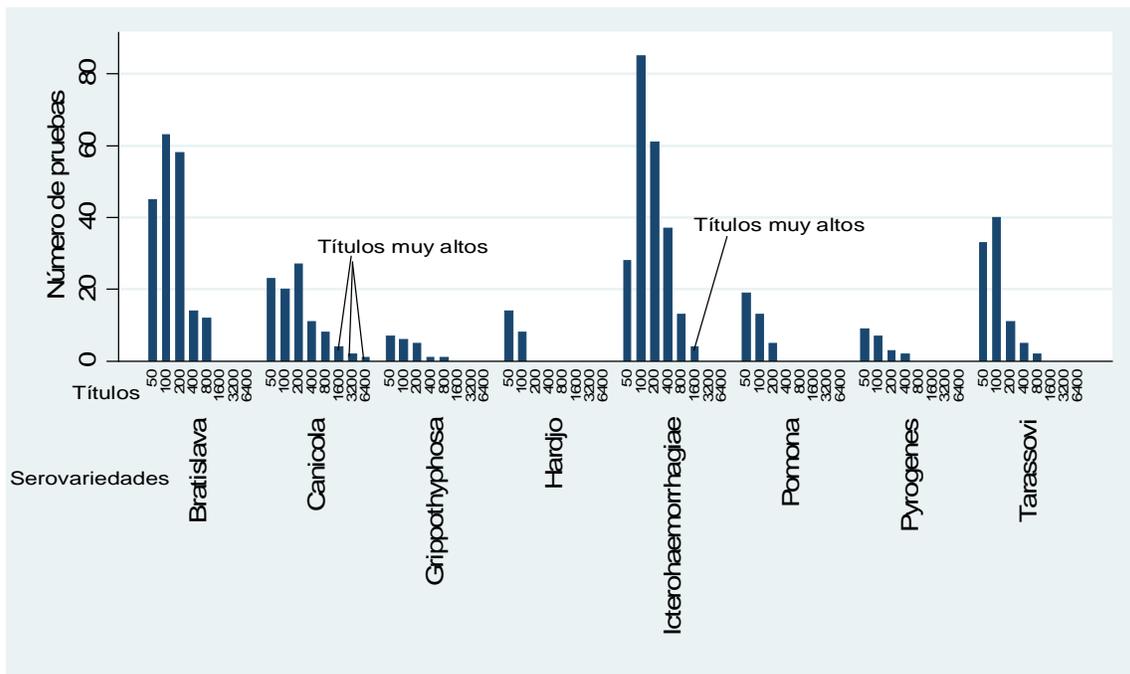
La determinación de las serovariedades de *Leptospira interrogans* presentes en la población de estudio es relevante, ya que es hacia éstas a las que se deben dirigir las medidas preventivas, por ejemplo, la creación de una vacuna que ataque específicamente a ellas.

En las ovejas seleccionadas se encontraron 8 diferentes serovariedades: Bratislava, Canícola, Grippothyphosa, Hardjo, Icterohaemorrhagiae, Pomna, Pyrogenes y Tarasovi; en todas ellas los títulos más frecuentes fueron 1:50, 1:100 y la falta de reacción (no seropositividad).

La serovariedad Canícola generalmente identificada en perros e Icterohaemorrhagiae en ratas son las únicas que mostraron títulos muy altos (hasta 1:6400 y 1:1600 respectivamente) lo cual sugiere la adaptación al nuevo huésped. Además, la presencia de títulos altos en un reducido número de ovejas indica que se puede clasificar a la especie ovina (por lo menos en esta población) como huésped accidental de las serovariedades Canícola e Icterohaemorrhagiae.

Por otro lado, las serovariedades Bratislava, Grippothyphosa, Hardjo, Pomona, Pyrogenes y Tarassovi que mostraron títulos bajos y en algunos casos como con *L. bratislava* en una cantidad considerable mayor que las otras serovariedades, sugieren que estos serogrupos de *Leptospira* se han instalado en la especie ovina, por lo tanto las ovejas ahora deberán ser consideradas en la epidemiología de la zona. La gráfica 5.1 muestra la distribución de los títulos serológicos anti-*Leptospira*.

Gráfica 5.1.- Distribución de los títulos en las diferentes serovariedades.



5.4.-Factores de riesgo. Modelo de regresión logística con efectos mixtos

El siguiente paso después de haber estimado la proporción de animales que mostraron reacción serológica contra *Leptospira*, es el estudio de los factores que pueden estar relacionados con la presencia o ausencia de dicha característica. Un método común en estudios de este tipo es la regresión logística (capítulo 1) debido a la interpretación epidemiológica natural de los coeficientes en términos de razón de momios (odds ratio). Un supuesto muy importante en el modelo logit clásico es el de la independencia de la variable respuesta en los diferentes elementos en estudio. Al permanecer las ovejas dentro de las unidades de agrupamiento (rebaño) es muy probable que los animales que compartan el mismo sitio estén correlacionados unos con otros, con lo cual no se cumple la condición deseada y al ser esto ignorado, conduce a errores estándar subestimados, por lo tanto, a identificar relaciones “falsas” o no adecuadas. Los modelos de efectos mixtos (capítulo 2) son una alternativa para manejar el problema de las observaciones correlacionadas. En este estudio la variable rebaño (de agrupación) debe tener efecto aleatorio sobre la respuesta, controlando así la posible correlación.

5.4.1.-Construcción del modelo

Modelo nulo

El primer paso es ver si el nivel rebaño tiene un efecto aleatorio significativo en la ausencia de cualquier otra variable explicativa. El modelo por ajustar es:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + u_j \quad u_j \text{ es una realización de } u \sim N(0, \sigma_u^2)$$

Donde π_{ij} es la probabilidad de que la oveja i en el rebaño j sea seropositiva y u_j es el efecto aleatorio del rebaño j . Los parámetros por estimar son β_0 y σ_u^2 . Mediante el comando *xtnlogit* se ajustó este modelo arrojando los resultados mostrados en la tabla 5.4.

Tabla 5.4 Modelo nulo con efecto aleatorio.

Parte fija						
Parámetro	Coefficiente estimado	Error estándar	z	P> z	Intervalo al 90% de confianza	
β_0	0.1563239	0.1323877	1.18	0.0852	0.1031512	0.4157989
Parte aleatoria						
Parámetro	Coefficiente estimado	Error estándar	χ^2	P> χ^2	Intervalo al 90% de confianza	
σ_u^2	0.1797179	0.1617517	2.4	0.0608	0.0408934	0.7898219

En la parte inferior de la tabla 5.4 se observa la prueba de razón de verosimilitudes para la hipótesis nula $H_0: \sigma_u^2 = 0$. La estadística de prueba tiene una distribución mixta y el valor-p es la mitad del que se obtiene normalmente (sección 2.4.4). El programa STATA muestra el valor correcto. Si se acepta un nivel de significancia del 7% la hipótesis nula se rechaza, entonces existe efecto aleatorio significativo de rebaño ($p= 0.061$).

El coeficiente $\hat{\beta}_0 = 0.15$ se interpreta como el logaritmo del momio de ser seropositivo en un rebaño "promedio", es decir con la estimación del efecto aleatorio (sección 2.4.5) correspondiente igual a cero ($\hat{u}_j = 0$). Para cada uno de los rebaños el logaritmo del momio de ser seropositivo es $0.15 + \hat{u}_j$, donde el efecto aleatorio tiene una varianza de 0.17.

La gráfica 5.2 muestra la estimación del efecto de cada rebaño y su intervalo al 90% de confianza, los cuales están ordenados de menor a mayor de acuerdo a su valor puntual. Se puede observar que sólo las manadas 21 y 14 tienen un efecto estadísticamente significativo, siendo éstas las que hacen que la variabilidad sea mayor que cero.

Una manera de medir el porcentaje de variabilidad total que es debida a la diferencia entre grupos es mediante el coeficiente de partición de la varianza (VPC por sus siglas en inglés):

$$VPC = \frac{\text{varianza nivel 2}}{\text{varianza nivel 1} + \text{varianza nivel 2}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

Donde σ_u^2 es la estimada por el modelo y σ_e^2 en el caso de la regresión logística se fija en 3.29^1 (sección 2.5.4) (cuando se usa la liga probit, $\sigma_e^2=1$). En este caso el VPC es igual a $0.1797/(0.1797+3.29)=.05179$, entonces el porcentaje de la variabilidad total debida a los rebaños cuando no se está ajustando por ninguna otra variable es del 5%.

La estimación de σ_u^2 se hace por integración numérica (sección 2.5.2) por lo que sólo es una aproximación y se debe analizar si es adecuada, esto se puede realizar aumentando el número de puntos de integración y ver si existen cambios drásticos en los coeficientes estimados.

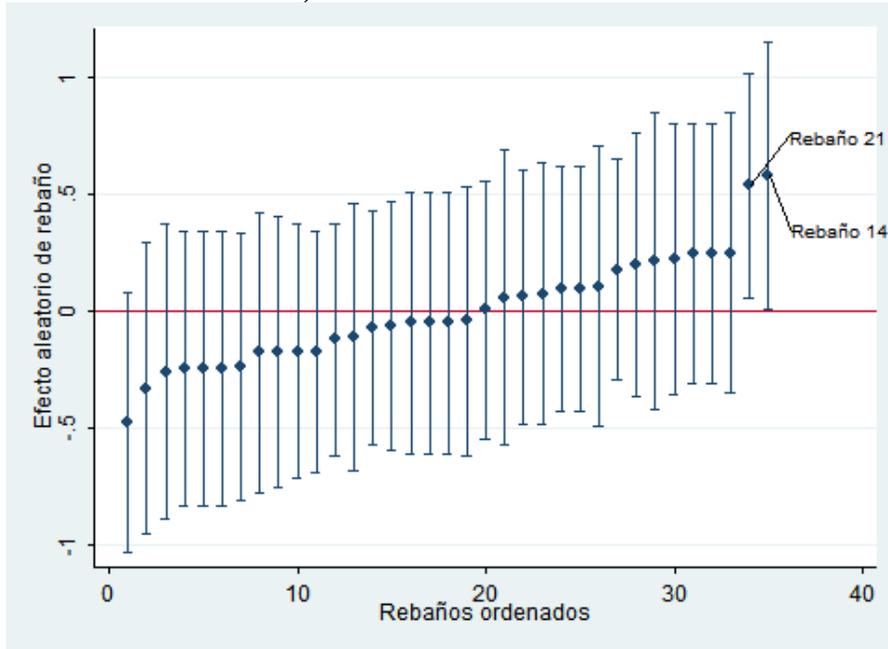
Tabla 5.5 Coeficientes estimados con diferentes puntos de integración.

Parámetro	Número de puntos de integración					
	1	5	7	8	20	50
β_0	0.1574568	0.1563235	0.1563239	0.1563239	0.1563239	0.1563239
σ_u^2	0.1662583	0.1797162	0.1797179	0.1797179	0.1797179	0.1797179
Log verosimilitud	-251.77179	-251.7014	-251.7014	-251.7014	-251.7014	-251.7014

¹ VPC también es llamado coeficiente de correlación intraclass por algunos autores. (Fieldng 2010, Collet 2003).

Los parámetros estimados así como la log-verosimilitud se estabilizan a partir de 7 puntos (tabla 5.5) (número predeterminado en STATA). El cambio en los valores calculados cuando se pasa de 1 a 7 puntos de integración no es muy grande, por lo tanto en los modelos sucesivos se seguirá usando esta cantidad (7).

Gráfica 5.2 Estimación de u_j para cada rebaño.



Modelos con una variable explicativa

La varianza debida a los rebaños puede aumentar o disminuir al incluir una variable explicativa, lo cual depende del comportamiento de dicha variable dentro de los diferentes rebaños. Éste es el primer paso para encontrar los factores relacionados con la respuesta.

La tabla 5.5 muestra únicamente los modelos en los que el valor-p de la prueba de Wald (sección 1.2.4) sobre el factor es menor a 0.25, esto es porque será ocupado en la siguiente sección. En la misma tabla se observa que al incluir una variable explicativa el efecto de rebaño desaparece, es decir la prueba $H_0: \sigma_u^2 = 0$ no se rechaza (excepto en los modelos con las variables Origen y Dxgest con un nivel de confianza del 90%).

El no rechazar la nulidad de la varianza del efecto aleatorio de rebaño indica que se puede regresar al modelo de regresión logística clásico sin embargo esto puede ser peligroso ya que los errores estándar pueden variar y así afectar las pruebas de hipótesis, esto será analizado en la construcción del modelo múltiple.

Tabla 5.5 Modelos mixtos con una sola variable explicativa.

Modelo	Categoría	Coef. estimado	Coef. Valor-p	σ_u^2	Valor-p σ_u^2 *
Estrato	Estrato B	0.9347467	0.007	0.0823839	0.1951
	Estrato C	0.6387093	0.089		
Chlamy	Positivo	0.350223	0.21	0.1362295	0.1327
	Sospechosos	1.183803	0.084		
Municipio	Santiago	-0.5949947	0.121	0.0972589	0.1909
	Calpuluah	-0.5211587	0.085		
	Chapultepec	-0.3651151	0.535		
Origen	Importadas	0.6482389	0.229	0.1504585	0.0803
	Mercado rural	-0.0390898	0.907		
Objetivo	Engorda	-0.3978195	0.215	0.1465514	0.1018
Matcorral	Malla	0.4443822	0.097	0.1250337	0.1461
Luparición	Paridero	0.6471256	0.129	0.0933057	0.2349
Disagua	Lago	0.3107618	0.221	0.1313974	0.1381
Pastoreo	Sí	0.8935239	0.068	0.1358275	0.1108
Encierro	aumento de 1	0.045935	0.161	0.1391758	0.1172
Limpcorr	No	-1.237085	0.026	0.0391488	0.3731
Dxgest	Sí	0.8151076	0.23	0.1634647	0.0708
Contacto	Sí	0.5407131	0.056	0.113314	0.1547
Bovis	Positivo	1.066262	0.178	0.0735916	0.2552
Babortus	Positivo	0.9517853	0.065	0.0363781	0.3652

*Valor-p correspondiente a la prueba $H_0: \sigma_u^2 = 0$

Modelos con múltiples variables explicativas

Los modelos univariados sólo evalúan el impacto de la variable explicativa aisladamente, se necesita estudiar el efecto cuando otra variable es tomada en cuenta, lo que se conoce como "ajustar estadísticamente". Debido a la gran cantidad de factores explicativos es muy difícil computar todos los modelos posibles; por la misma razón y por la complejidad de los cálculos que se hacen cuando se incluyen términos aleatorios no está implementado aún un algoritmo de selección de variables. El procedimiento que se usó para la elección es similar al llamado stepwise (Agresti 2002, Hosmer y Lemeshow 1989, Collet 2003). Los factores que se utilizaron en este método son los que fueron significativos en los modelos univariados (tabla 5.5) y en los que se encontró relación con la respuesta bajo las prueba de independencia de ji-cuadrada de Pearson o la prueba exacta de Fisher. Los pasos que se siguieron fueron los siguientes:

1. La variable estrato fue tomada como base ya que en todas las pruebas preliminares fue significativa además de que es importante en el diseño. Se empezó con el modelo que incluía únicamente a esta variable.
2. Se incluye una variable y se utiliza la devianza para comparar este nuevo modelo con el anterior. La variable recién agregada es retenida si el valor-p de la prueba es menor que 0.10. La significancia y los criterios AIC y BIC² (Agresti 2002) también son evaluados.
3. Si alguna variable del modelo previo ya no es significativa (en presencia de un nuevo factor que si lo es) es sacada del modelo.
4. Se repite el procedimiento hasta que ninguna variable entre o salga.

Los modelos ajustados se muestran en la tabla 5.6. En los que hubo un cambio significativo están marcados con negritas.

Tabla 5.6 Modelos mixtos con más de una variable explicativa.

MODELO	DIF DEVIANZAS	G.L	Valor-p	AIC	BIC
<i>estrato*,chlamy·</i>	5.2936	2	0.0708	503.023	526.45
estrato*,chlamy·,Municipio	4.4387	3	0.2178	504.584	539.7328
estrato*,chlamy·,objetivo	2.269	1	0.1319	502.754	530.0911
estrato*,chlamy·,origen	0.8909	2	0.6405	506.132	537.3752
<i>estrato*,chlamy·,matcorral*</i>	4.2808	1	0.0385	500.74	528.08
estrato*,chlamy·,matcorral·,luparición	2.4322	1	0.1188	500.31	531.5531
estrato*,chlamy·,matcorral·,disagua	0.7819	1	0.3765	501.96	533.2034
estrato*,chlamy·,matcorral·,pastoreo	2.3607	1	0.1244	500.381	531.6246
estrato*,chlamy·,matcorral·,separto	0.1572	1	0.6917	502.585	533.8281
estrato*,chlamy·,matcorral·,encierro	1.1122	1	0.2916	501.63	532.8702
estrato*,chlamy·,matcorral·,municipio	2.0977	3	0.5523	504.649	543.6983
estrato*,chlamy·,matcorral·,objetivo	0.9893	1	0.3198	501.753	532.9961
estrato*,chlamy·,matcorral*,origen	0.6917	2	0.7076	504.05	539.199
<i>estrato*,chlamy·,matcorral*,limpcorr*</i>	6.1993	1	0.0127	496.543	527.786
estrato*,chlamy·,matcorral·,limpcorr*dxgest	0.3523	1	0.5527	498.19	533.339
estrato*,chlamy·,matcorral·,limpcorr*,contacto	0.6993	1	0.4031	497.843	532.992
estrato*,chlamy·,matcorral*,limpcorr·,bovis	100.8039	1	1.02E-23	397.739	431.3101
estrato*,chlamy·,matcorral·,limpcorr·,municipio	2.0396	3	0.5642	500.503	543.4628
estrato*,chlamy·,matcorral·,limpcorr*,objetivo	0.9835	1	0.3213	497.559	532.7079
estrato*,chlamy·,matcorral·,limpcorr*,origen	0.7955	2	0.6888	499.797	538.8512
estrato*,chlamy·,matcorral*,limpcorr·,luparición	0.6889	1	0.7655	498.454	533.6025

² AIC: Akaike information criterion; BIC: Bayesian information criterion.

Tabla 5.6 Modelos mixtos con más de una variable explicativa (continuación).

MODELO	DIF DEVIANZAS	G.L	Valor-p	AIC	BIC
estrato*,chlamy-,matcorral,limpcorr*,disagua	0.4794	1	0.4886	498.063	533.2119
estrato*,chlamy-,matcorral,limpcorr*,pastoreo	2.6204	1	0.105495	495.922	531.0709
estrato*,chlamy-,matcorral,limpcorr*,separto	0.0088	1	0.9249	498.534	533.6825
estrato*,chlamy,matcorral,limpcorr*,encierro	0.79	1	0.3741	497.753	532.9013
estrato*,chlamy-,matcorral*,limpcorr-,babortus-	107	1	4x10-25	391.513	424.9965
estrato*,chlamy-,matcorral-,limpcorr-,babortus-, municipio	1.0323	3	0.7939	396.481	441.1257
estrato*,chlamy-,matcorral-,limpcorr-,babortus-, objetivo	0.3802	1	0.5374	393.133	430.3366
estrato*,chlamy-,matcorral*,limpcorr-,babortus-, origen	0.5603	2	0.7556	394.953	435.8768
estrato*,chlamy-,matcorral*,limpcorr,babortus-, luparición	0.0223	1	0.8811	393.491	430.6945
estrato*,chlamy-,matcorral,limpcorr-,babortus-, disagua	0.5415	1	0.4418	392.922	430.1253
estrato*,chlamy-,matcorral*,limpcorr-,babortus-, pastoreo	0.7395	1	0.3898	392.774	429.9713
estrato*,chlamy-,matcorral*,limpcorr-,babortus-, separto	0.0751	1	0.7831	393.438	430.6416
estrato*,chlamy-,matcorral-,limpcorr-,babortus-, encierro	0.6785	1	0.41	392.835	430.0383
estrato*,chlamy-,matcorral*,limpcorr-,babortus-, dxgest	0.19938	1	0.6552	393.314	430.3175
estrato*,chlamy-,matcorral*,limpcorr-,babortus-, contacto	0.83129	1	0.3619	392.682	429.8856
estrato*,chlamy-,matcorral*,limpcorr-,babortus-, Bovis	1.2017	1	0.2729	392.312	429.5151

Las columnas: Diferencia de devianzas, G.L (grados de libertad) y valor-p de la tabla 5.6 corresponden a la comparación y prueba de hipótesis de modelos anidados (sección 1.2.5 y 2.4.4) del modelo presentado en cada renglón con el mismo modelo pero sin la última variable. Las columnas AIC y BIC corresponden al modelo presentado.

Hay 4 modelos en los que se tienen cambios importantes en la devianza y que tienen variables significativas, los cuales están marcados con negritas en la tabla 5.6 y numerados en la tabla 5.7.

Tabla 5.7 Significancia de σ_u^2 de los principales modelos mixtos.

MODELO	$\hat{\sigma}_u^2$	Valor-p	VPC
1.-estrato	0.0823839	0.1951	0.02442898
2.-estrato,chlamy	0.0399369	0.3435	0.01199329
3.-estrato,chlamy,matcorral	4.32E-15	1	1.31E-15
4.-estrato,chlamy,matcorral,limpcorr	2.06E-14	1	6.26E-15
5.-estrato,chlamy,matcorral,limpcorr,babortus	1.13E-16	1	3.43E-17

Como se puede observar en la tabla 5.7 en ninguno de los modelos fue significativa la varianza del efecto aleatorio de rebaño, más aún al agregar una segunda variable de nivel II, $\hat{\sigma}_u^2$ es estadísticamente igual a cero con un valor-p cercano a 1 y al incluir un tercer factor la variabilidad debida a la agrupación (VPC) es prácticamente nula.

Los coeficientes estimados en el mejor modelo mixto (modelo 5 tabla 5.7) se muestran en la tabla 5.8.

Tabla 5.8 Coeficientes estimados del modelo mixto seleccionado.

Parte fija						
Parámetro	Coefficiente estimado	Error estándar	z	P> z	Intervalo al 90% de confianza	
Estrato 2	0.8651397	0.3628535	2.38	0.017	0.2682988	1.461981
Estrato 3	0.3817941	0.4095182	0.93	0.351	-0.291803	1.055392
Chlamy 1	0.0249366	0.3078975	0.08	0.935	-0.48151	0.531383
Chlamy 2	1.842417	1.076591	1.71	0.087	0.0715822	3.613252
Matcorral 1	0.5928036	0.2717162	2.18	0.029	0.1458702	1.039737
Limpcorral 1	-1.270608	0.6877072	-1.85	0.065	-2.401786	-0.1394308
Babortus 1	0.9017116	0.5248013	1.72	0.086	0.0384903	1.764933
Const	0.6829029	0.7978719	0.86	0.392	-0.62948	1.995285
Parte aleatoria						
Parámetro	Coefficiente estimado	Error estándar	χ^2	P> χ^2	Intervalo al 90% de confianza	
	1.13E-16	3.97E-09	0	1	0	.

Se sabe que los modelos mixtos son condicionales al efecto aleatorio, es decir tanto los valores esperados, los momios y las probabilidades que se estiman se calculan una vez dada una realización de éste:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \text{Predictor lineal estimado} + \Omega_j$$

Al ser $\hat{\sigma}_u^2 = 0$ y cada u_j cero, se llega a que el modelo mixto queda igual que el logístico usual, por esta razón la interpretación de los coeficientes se dará más adelante.

5.5.-Modelo logístico sin efecto aleatorio

Para poder regresar a la regresión logística de un sólo nivel se necesita revisar que los errores estándar de las variables y los modelos finales bajo las dos técnicas no tengan cambios muy "drásticos". A continuación se presenta un resumen de la construcción de este modelo y algunos comentarios importantes:

1. En todos los modelos con dos o más variables explicativas los errores estándar fueron exactamente los mismos que en su similar con efecto aleatorio de rebaño. En los que sólo se incluía una variable, la diferencia en los errores estándar era a lo más de centésimas y al nivel de confianza que se ha estado trabajando (90%) no hubo cambio, con respecto al modelo mixto, en cuál era significativa o no.
2. *Selección de variables mediante stepwise iniciando por modelo nulo y por modelo completo sólo con las variables que se usaron para la construcción del modelo mixto múltiple:* Se utilizó un valor de entrada de menos de 0.10 y de salida mayor o igual que el mismo. La prueba en la que se basó el método fue la razón de verosimilitudes. El modelo bajo los dos criterios (modelo nulo y completo) resultante fue exactamente el mismo que se obtuvo mediante la regresión logística con efecto mixto.
3. *Selección de variables mediante stepwise iniciando por modelo nulo tomando en cuenta todas las variables explicativas:* El modelo resultante es el mismo que el mixto.
4. *Selección de variables mediante stepwise iniciando por modelo completo tomando en cuenta todas las variables explicativas:*
 - Mediante este criterio de selección el modelo resultante incluyó las variables estrato, Chlamy, Matcorral, limpcorral, Aborto y Auxparto.
 - Si al modelo resultante se le añadía la variable Babortus (que es la que sale significativa mediante los anteriores criterios) tanto Aborto, Auxparto y Babortus no eran significativas. Esto se puede deber a la alta correlación Auxparto-Aborto (presentada en la muestra) y Aborto-Babortus (presentada teórica y en la muestra).
 - Se corrieron los modelos en los que se incluían las 4 primeras variables (que en ningún modelo cambian significativamente) más las variables Aborto-Babortus y Auxparto-Babortus, en los cuales solo la variable Babortus fue significativa. Esto nos condujo de nueva cuenta al modelo logístico encontrado con el efecto aleatorio de rebaño.

5. Todos los coeficientes estimados en estos modelos coincidieron con los estimados bajo el modelo de segundo nivel (tabla 5.8).

En este estudio se puede determinar mediante los cinco puntos anteriores que el modelo logístico clásico y mixto coinciden. Por esto y porque se encontró que $\hat{\sigma}_u^2 = 0$ concluimos que el efecto de agrupación en rebaños es nulo, esto nos lleva a que podemos tratar la variable respuesta entre diferentes ovejas como independiente y por lo tanto poder usar la regresión logística usual, lo cual es muy conveniente ya que estos modelos han sido más estudiados y se cuenta con métodos de bondad de ajuste (en los de efectos mixtos no) además de ser más sencillos.

5.5.1.-Bondad de ajuste y verificación del modelo

Antes de interpretar los parámetros estimados, se deberá revisar si hay buen ajuste del modelo a los datos, si la función liga es adecuada y si existen observaciones atípicas o influyentes.

Debido a que las observaciones se encuentran de forma desagrupada (binaria), la devianza no es un método adecuado de bondad de ajuste, en este caso se tiene que usar la estadística de Hosmer-Lemeshow (sección 1.3.4). La tabla 5.9 muestra la prueba con diferentes números de grupos en la cual podemos observar que en ninguno de ellos se rechaza la hipótesis nula de que el modelo ajuste adecuadamente a los datos.

Tabla 5.9 Prueba Hosmer-Lemeshow.

Número de grupos	Estadística	Valor-p
4	0.85	0.6544
6	2.24	0.6915
7	2.26	0.688
10	2.03	0.8448

Existen pruebas para verificar si la elección de la función liga es adecuada, las cuales se basan en las transformaciones de Arnada-Ordaz (Collet 2003). Se crea una nueva variable. Para que la liga sea adecuada esta variable no tiene que tener poder explicativo (ser no significativa). Para el modelo en cuestión, resultó que no se rechaza la hipótesis nula de que la liga Logit es adecuada (comando *linktest* de STATA).

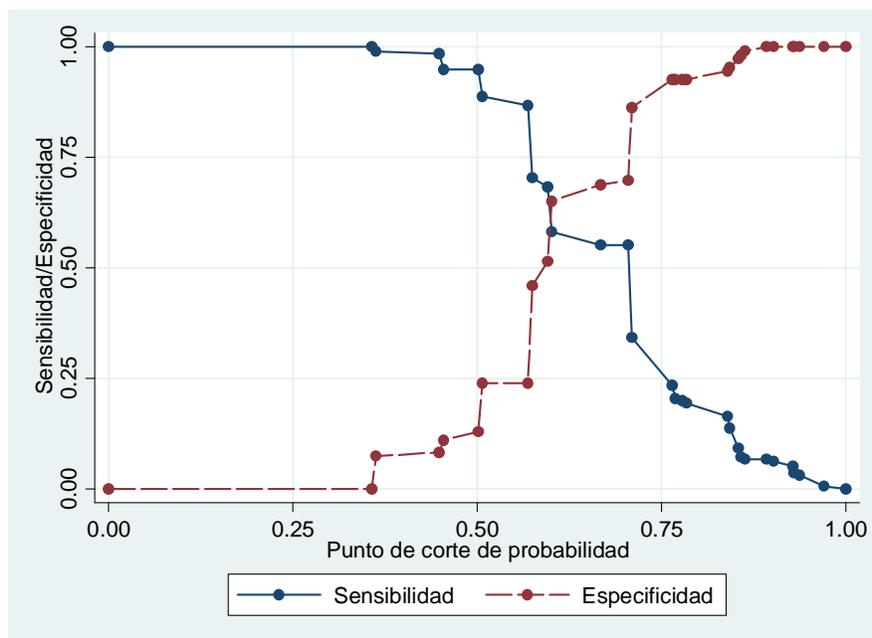
Otra manera de medir cómo ajusta nuestro modelo a los datos es mediante una tabla de clasificación donde se calcula la sensibilidad (probabilidad de que dada una oveja positiva, el modelo la clasifique correctamente con un punto de corte dado) y la especificidad (probabilidad de quedada una oveja no seropositiva sea clasificada como tal). La gráfica 5.3 muestra cómo

cambian la sensibilidad y especificidad con las probabilidades estimadas de los diferentes patrones de covariables (diferentes combinaciones de los niveles de las variables, en este modelo 28 en total) y la tabla 5.10 muestra las mismas estadísticas con ciertos puntos de corte de interés. A pesar de que el modelo no se construyó con fines predictivos el nivel de clasificados correctamente es relativamente alto, con lo cual se puede apoyar la hipótesis de que no hay falta de ajuste.

Tabla 5.10 Algunos puntos de corte para el cálculo de sensibilidad y especificidad.

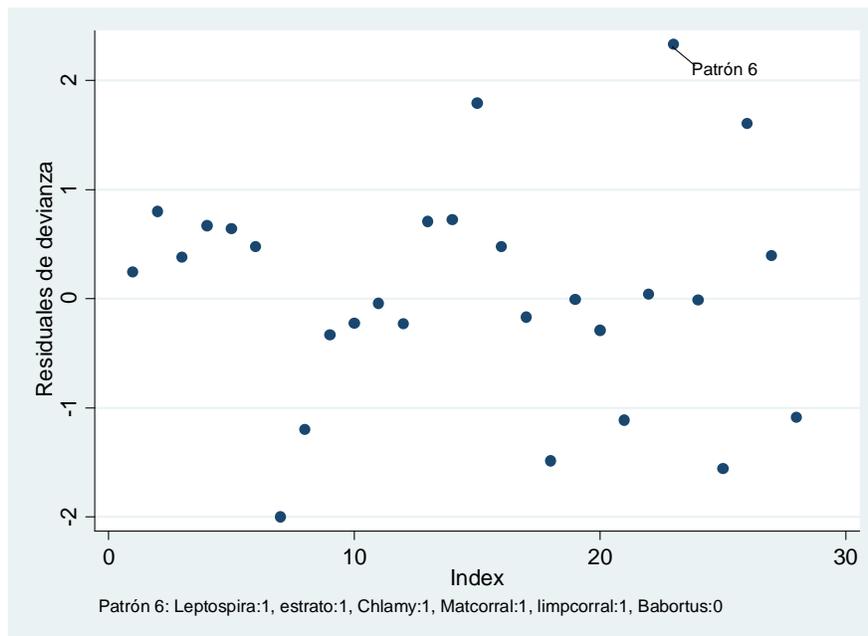
Punto de corte	Sensibilidad	Especificidad	Clasificados correctamente
Prevalencia estimada(.5341)	86.73%	23.85%	64.26%
Probabilidad media estimada(.6426)	55.10%	68.81%	60.00%
Punto "óptimo"(.5955)	68.37%	51.38%	62.30%

Gráfica 5.3 Sensibilidad y especificidad



Para identificar posibles patrones de covarianza que sean "atípicos" o influyentes en la estimación de los coeficientes se utilizan los residuos de la devianza (el programa STATA no lo hace por observación, sino por patrón de covarianza).

Gráfica 5.4 Residuos de devianza.



En STATA se pueden calcular estadísticas que miden el cambio en la devianza y en los coeficientes estimados al eliminar cada patrón de covariables (Manual de referencia STATA 11). Las diferencias más importantes resultaron ser en el patrón 6, el mismo que se observa más alejado en la gráfica 5.4. Los resultados de ajustar el modelo sin las observaciones que están dentro de este patrón (4 en total) se muestran en la tabla 5.11.

Tabla 5.11 Modelo ajustado sin el patrón de covariable 6.

Parámetro	Coefficiente estimado	Error estándar	z	P> z	Intervalo al 90% de confianza	
Estrato 2	1.118334	0.3861099	2.9	0.004	0.48324	1.753428
Estrato 3	0.6339242	0.4291917	1.48	0.14	-0.072033	1.339882
Chlamy 1	-0.1569272	0.3167721	-0.5	0.62	-0.677971	0.3641166
Chlamy 2	1.888369	1.084792	1.74	0.082	0.1040451	3.672692
Matcorral 1	0.56483	0.2732283	2.07	0.039	0.1154095	1.014251
Limpcorral 1	-1.353172	0.6904857	-1.96	0.05	-2.48892	-0.2174243
Babortus 1	0.8785193	0.5244252	1.68	0.094	0.0159167	1.741122
Const	0.5715823	0.8034836	0.71	0.477	-0.750031	1.893195

No se observan cambios drásticos en los coeficientes estimados ni en sus errores estándar, los mismos niveles de las variables que eran significativos anteriormente lo siguen siendo aún.

El coeficiente del nivel 1 de Chlamy cambió de signo, sin embargo sigue siendo estadísticamente cero.

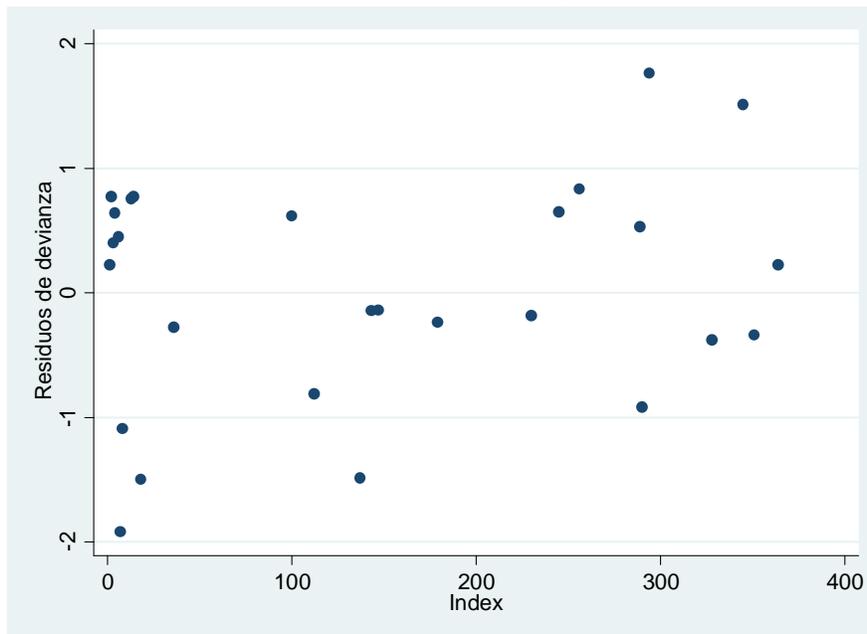
La tabla 5.12 muestra un resumen del modelo una vez eliminado el patrón 6. Sigue habiendo un buen ajuste, la función Logit aún es adecuada y hubo una pequeña ganancia en especificidad y en observaciones clasificadas correctamente.

Tabla 5.12 Resumen de ajuste del modelo después de eliminar el patrón 6.

Prueba o Estadística de clasificación	Estadística	Valor-p
Hosmer-Lemeshow (7 grupos)	2.33	0.802
Linktest	0.0651554	0.753
Sensibilidad(.5955)	66.67%	-
Especificidad(.5955)	55.05%	-
Clasificados correctos	62.46%	-

En la gráfica 5.5 ya no se observan residuos demasiado grandes (todos menores a 2 regla empírica).

Gráfica 5.4 Residuos de devianza del modelo sin el patrón 6.



5.6.-Interpretación de los coeficientes

Una vez que se ha verificado que el modelo es adecuado se pueden interpretar los coeficientes estimados, esto se hará en términos de los odds ratios o razón de momios (sección 1.3.3). Se explicará el modelo sin el patrón de covariables 6 (tabla 5.11).

Tabla 5.13 Odds Ratio (Razón de momios) del modelo ajustado.

Parámetro	Odd Ratio	Intervalo al 90% de confianza	
Estrato 2	3.059753	1.621319	5.774365
Estrato 3	1.884993	0.9304999	3.818592
Chlamy 1	0.8547663	0.507646	1.439242
Chlamy 2	6.608579	1.10965	39.35773
Matcorral 1	1.759149	1.122333	2.757296
Limpcorral 1	0.2584192	0.0829995	0.8045885
Babortus 1	2.407333	1.016044	5.703739

Las posibilidades de una oveja de ser seropositiva a *Leptospira interrogans* en rebaños con 51 a 140 animales (estrato 2) son mayores (de 1.62 a 5.7 veces) que las de otra con las mismas características, pero que conviven en una manada con menos de 50 (estrato 1) o más de 141 animales (estrato 3). Una oveja que co-habita en un rebaño entre 51 a 140 animales: Factor de riesgo.

A pesar que la presencia de seroreactividad a *Chlamydomphila abortus* resulto ser no influyente en la variable respuesta de nuestro modelo, las posibilidades de la infección mixta entre las ovejas seropositivas a *L. interrogans* fueron mayores (1.1 a 39.4 veces) que las observadas en las ovejas seronegativas.

La capacidad económica de los ovino cultores para construir barreras de protección de sus animales es reflejada por el material utilizado en los corrales de encierro nocturno. Cuando el material del corral es de malla o tela ciclónica las posibilidades de que una oveja que habita en un lugar de este tipo sea seropositiva son de 1.12 a 2.75 veces las posibilidades de una oveja en un rebaño que habita en un corral hecho de tabique. Nivel socio-económico: Factor de riesgo.

En los rebaños donde se realiza la limpieza del corral de encierro nocturno se encontró que las posibilidades de que una oveja fuera seropositiva eran de 0.08 a 0.8 veces las posibilidades de una oveja que habita en un rebaño donde no se realiza limpieza. Limpieza del rebaño: Factor de protección.

Las posibilidades de co-infección *Brucella abortus/Leptospira interrogans* en una oveja cuando ésta sea seropositiva a *Brucella abortus* son de 1.02 a 5.7 veces las posibilidades de una oveja en las mismas condiciones pero no seropositiva. Reactividad a *Brucella abortus*: Factor de riesgo.

CONCLUSIONES.

La percepción general es no considerar a la especie ovina como un reservorio para ninguna de las serovariedades de *Leptospira interrogans*. Con los resultados serológicos del estudio se concluye lo siguiente:

1. La prevalencia de anticuerpos anti-*Leptospira* en las ovejas en esta región del Estado de México es por arriba del 50%, es decir que al menos 5 de cada 10 animales estuvieron expuestos al antígeno.
2. Las condiciones ecológicas del área bajo estudio pueden facilitar la sobrevivencia y propagación de la leptospirosis en los animales, aunque existe la creencia general de que *L. interrogans* subsiste fuera del huésped en ecosistemas tropicales y no en altitudes por arriba de los 1500 m.s.n.m.m.
3. Es posible clasificar a la población ovina, de acuerdo a la distribución de los títulos serológicos, como huésped accidental para las serovariedades Canicola e Icterohaemorrhagiae y reservorio de mantenimiento para Bratislava, Grippothyphosa, Hardjo, Pomona, Pyrogenes y Tarassovi.
4. El modelo de regresión logística mostró que los rebaños con un rango de 51 a 140 animales, el nivel socio-económico establecido por los materiales del corral de encierro nocturno, la limpieza del corral y la co-infección *Chlamydophila abortus/Leptospira* y *Brucella abortus/Leptospira* se mantuvieron independientemente asociados con la seropositividad a *Leptospira* lo que sugiere que estos cinco factores en conjunto fueron los predictores más fuertes en la evaluación de los factores de riesgo epidemiológico.

Mientras que los altos ó persistentes títulos de anticuerpos no reflejan necesariamente una infección activa o crónica, la identificación serológica de las ovejas permitirá establecer las posibles medidas de control a tener en cuenta en la epidemiología de la leptospirosis ovina. La identificación y comprensión de los predictores a nivel de rebaño también permitirá la aplicación concreta de las campañas de educación y sensibilización que serán necesarios para asegurar el éxito de una estrategia regional de control de la enfermedad. Esto será particularmente importante para la detección inicial de reemplazos dentro del rebaño antes de la introducción de animales sin antecedentes clínicos del aborto o de fracaso reproductivo, así como limitar el riesgo de transmisión zoonótica.

DISCUSIÓN

Contrariamente a las predicciones hechas en el siglo 20, las enfermedades infecciosas y no las crónico-degenerativas siguen siendo la causa principal de muerte en todo el mundo. Varias características complejas contribuyen a la creciente amenaza de enfermedades re-emergentes y de la gama más amplia de agentes patógenos reconocidos. El sistema de diagnóstico serológico de *Leptospira interrogans* en los animales infectados es un factor crítico en la reducción de la morbilidad en humanos en los países en desarrollo, incluso en las zonas rurales y suburbanas. Aunque el examen serológico es la única prueba presuntiva de la infección en los animales y el peso del resultados de esta se ve aumentado cuando se realizan muestras pareadas (en intervalos de tiempo de dos o tres semanas que muestra un marcado aumento de los títulos de anticuerpos), en este estudio, los títulos de anticuerpos contra *L. interrogans*, *Cp. abortus*, *B. ovis* y *B. abortus* puede sugerir contacto previo con los agentes infecciosos.

La proporción de ovejas seropositivas a *L. interrogans* identificados en la muestra de estudio (53.4%; IC_{95%} 47% a 60%) fue sorprendentemente alta en comparación con lo reportado en otros países que han informado anteriormente sus prevalencias, tales como, Australia (32.3%), Nigeria (23.5%), Irán (18.4%) y Grecia (13.6 %) (Cousins y Robertson, 1986; Agunloye, 2002; Burriel y cols., 2002; Tooloei y cols., 2008). Incluso esta proporción de ovejas seropositivas comparada con otras especies domésticas de diferentes regiones de México (Arteaga y cols., 1991). Este resultado podría explicarse por la estrecha relación entre la exposición de animales infectados (potencialmente ganado vacuno, perros, ratas y/o fauna silvestre) y un gran número de ovejas susceptibles, sus serotipos y los desechos de los animales en la temporada previa de partos.

En el presente estudio, los porcentajes de seropositividad en las regiones valle, intermontaña y montaña no fueron estadísticamente diferentes (52.5 %, 55.6% y 57% respectivamente). Importante es el hecho de la sobrevivencia y propagación de la infección en los animales de esta zona, ya que existe la creencia general de que *L. interrogans* subsiste fuera del huésped en ecosistemas tropicales y no en altitudes por arriba de los 1500 m.s.n.m como las que presenta la región de estudio (3273 m.s.n.m en la región de montaña). Esto puede indicar que la bacteria está adquiriendo una gran capacidad de adaptación fuera del huésped, apareciendo en lugares en los que anteriormente se pensaba no se encontraría. Los factores climáticos y ambientales que pueden influir en la distribución de *L. interrogans* entre las ovejas se determinó con la ayuda de GPS y el sistema de información geográfica del Instituto Nacional de Estadística, Geografía e Informática (INEGI, 2008). Al parecer, el nivel de óptima supervivencia de *L. interrogans* fuera del huésped puede ocurrir bajo condiciones de humedad fría con un pH cercano al neutro en el suelo en la región de valle. Estos hallazgos sugieren que las condiciones de temperatura, precipitación y humedad y las características del suelo representan un peligro considerable para la perpetuación de la infección en esta región. Si, *L. interrogans* se transmite sobre todo a la población susceptible a través del agua o el suelo y por contacto directo con otros animales infectados durante la

convivencia diaria y la infección se expande entre las ovejas, las velocidades de transmisión y el riesgo para los seres humanos también se incrementarían.

Esta población ovina se puede clasificar, de acuerdo a la distribución de los títulos serológicos, como huésped accidental para las serovariedades *Canicola* (19.9%) e *Icterohaemorrhagiae* (54.2%) y reservorio de mantenimiento para Bratislava (38.9%), *Grippothyphosa*, Hardjo, Pomona, Pyrogenes y Tarassovi (15.5%). Los casos seropositivos con los más altos títulos de anticuerpos para *Canicola* (1:6400) e *Icterohaemorrhagiae* (1:1600) fueron identificados en la región del valle, donde la mayoría de ellas convive con diferentes animales domésticos. Todos los animales pueden ser reservorios cuando las leptospirosas persisten en los sitios anatómicos favorecidos para su adhesión, especialmente en el túbulo renal (Hansen, 1976). Sin embargo, el compromiso renal en muchos animales es de naturaleza diferente. En algunos los altos títulos se relacionan con daño renal y la excreción de grandes cantidades de leptospirosas en la orina y en otros la cronicidad sin daño aparente pero con la posibilidad de transmisión (Levett, 2001). Estos reservorios pueden ser las especies silvestres y roedores que viven en zonas altamente pobladas (Sunbul y cols., 2001). La exposición al agua o tierra contaminada con orina de animales infectados puede aumentar el riesgo de infección, y la localización precisa de los rebaños cercanos al lago de Chingnahuapan, donde permanecen las mejores pasturas y se convive con los animales de otros rebaños puede sugerir una zona de mayor contaminación de *L. canicola* y *L. icterohaemorrhagiae* que la zona de la montaña. Además, estos hallazgos sugieren que las ovejas expuestas a la orina de perros, ratas o las especies de vida silvestre en su vida diaria están en riesgo de contraer la infección de una fuente común de exposición (Trevejo y cols., 1998; Vado-Solís y cols., 2002). Con excepción de *L. bratislava*, no se observaron diferencias significativas en el porcentaje de seropositividad para *L. icterohaemorrhagiae*, *L. canicola*, y *L. tarassovi* entre ovejas adquiridos en el mercado de ganado y los nacidos en el rebaño. Estos resultados sugieren que la persistencia de estos serotipos regionales puede existir por dos razones: 1) el comercio de animales infectados y 2) la mezcla de rebaños positivos y negativos a leptospirosis en corrales de engorde (Marshall y Manktelow, 2002).

El modelo logístico de efectos mixtos ajustado para la detección de posibles factores de riesgo que influyen en la presencia de anticuerpos anti-*Leptospira* mostró que el efecto de agrupación de las ovejas dentro de los rebaños es estadísticamente nulo, es decir la correlación entre la respuesta de cualquier par de ovejas es prácticamente cero y por lo tanto es posible no considerar el efecto aleatorio del rebaño.

En este modelo de logístico, los rebaños con un rango de 51 a 140 animales (OR 3.1, IC 95% 1.6-5.8, $p < 0.004$), respuesta serológica a *Chlamydomphila abortus* (OR 6.6, IC del 95% 1.1-39.4, $p = 0.08$), materiales del corral de encierro nocturno (OR 1.8, IC 95% 1.1-2.8, $p = 0.04$), limpieza del corral (OR 0.25, IC 95% 0.08-0.8, $p = 0.05$) y aquellos animales con co-infección *Brucella abortus/Leptospira interrogans* (OR 2.4, IC 95% 1.1-5.7, $p = 0.09$) se mantuvieron asociados con la seropositividad a *Leptospira* lo que sugiere que estos cinco variables en conjunto fueron los predictores más fuertes en la evaluación de los factores de riesgo epidemiológico. El hallazgo de que en los rebaños con un rango de 51 a 140 animales existe un efecto significativo en la seropositividad a leptospirosis puede haber sido favorecido por las prácticas de confinamiento de

ovejas en los corrales de encierro nocturno. En el cincuenta y tres por ciento de los rebaños que tuvieron esa cantidad de animales se observó la reclusión de seis ovejas o más en el corral de dichos rebaños, estos resultados son apoyados por otros estudios que muestran que el tamaño de las explotaciones y el alto confinamiento de los animales son importantes factores de riesgo (Ellis, WA, 1994; Lilenbaum y Santos, 1996; Aslanta y Özdemür, 2005). Por lo tanto, es importante hacer hincapié en las prácticas de crianza adecuadas que eviten el confinamiento de un gran número de animales para disminuir el riesgo de transmisión directa de leptospirosis.

Otros predictores relacionados con la seropositividad de leptospirosis ovina fueron la reactividad dual a *Cp. abortus* y *B. abortus*. A pesar que el MAT pudo detectar y diferenciar claramente a 112 de las 200 muestras positivas a *Leptospira* cuando la titulación de los sueros fue mayor a 1:200 es posible considerar la reactividad cruzada con estos antígenos bacterianos cuando los títulos sean menores o iguales a 1:100. Sin embargo, un menor porcentaje de reactividad cruzada no específica se observó entre las 88 ovejas seropositivas a leptospirosis que tuvieron títulos de 1:100 (26.1% de los sueros de ovejas mostraron una reacción no específica en el examen de *Cp. abortus* y 6.8% de los sueros para *B. abortus*). Estos antígenos de reacción cruzada podría incluir proteínas y/o componentes de sacarídicos de la pared celular o los lipopolisacáridos de leptospirosis (Matsuo y cols., 2000). La identificación serológica de las cepas lisas de *Brucella* que incluyen *B. abortus*, *B. melitensis* y *B. suis* en los sueros de las ovejas sugieren reactividad dirigida al lipopolisacárido predominante del antígeno de *Brucella*. Sin embargo, las reacciones cruzadas son sólo conocidas para *Escherichia coli* O157: H7, *Salmonella* serotipo O: 30, *Vibrio cholerae*, *Yersinia enterocolitica* serotipo O: 9, y *Francisella tularensis*, pero no para *Leptospira*, ya que comparten la estructura del lipopolisacárido M (Bundle y cols., 1987). Es posible que los resultados positivos a MAT puedan ser atribuidos al contacto previo con esos antígenos de *Chlamydomphila* y *Brucella*, aunque eso también sobreestimaría la seroprevalencia de leptospirosis encontrada en este estudio.

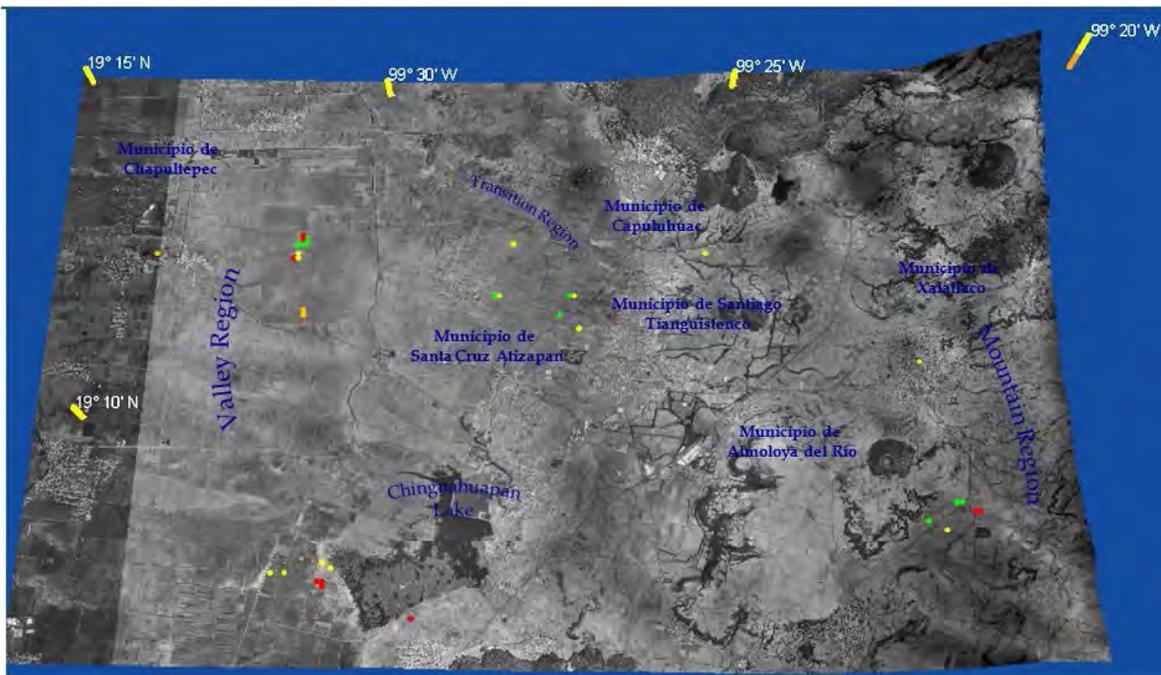
El MAT, ha sido recomendado por la Organización Mundial de la Salud y utiliza como método estándar la aglutinación de las bacterias para detectar los niveles de anticuerpos contra *L. interrogans* con una alta especificidad y sensibilidad. Solamente en el 21.6% de las muestras seronegativas a *Leptospira* tuvieron reacción única para *Cp. abortus* y *B. abortus* (18 sueros positivos a *Cp. abortus* y uno a *B. abortus*, respectivamente). Está más allá del alcance de este documento especular sobre la presencia de dual seropositividad a *Cp. abortus* y/o a *Brucella spp.*, pero estas ovejas pueden presentar un problema importante en el control del aborto enzoótico ovino y brucelosis y sus programas de erradicación.

Componentes del manejo de rebaño como material del corral de encierro nocturno y limpieza del mismo mostraron tener relación con la seropositividad de una oveja, el primero como un elemento socioeconómico que modifica el impacto de la enfermedad en los animales debido a que el material de construcción de los corrales pudiera reflejar una condición de reducida economía y el segundo como factor de protección cuando se realiza las prácticas de limpieza de esos corrales. La exacta explicación de ésta asociación con la presencia de malla ciclónica en los corrales de encierro nocturno requiere de mayor investigación en el modelo social de la región.

DISCUSIÓN

La imagen 6.1 es una fotografía satelital de la área de estudio. Se muestran la posición de los rebaños, éstos están marcados de acuerdo al porcentaje observado de animales seropositivos. Aunque la variable que indicaba si las ovejas convivían con los animales de otros rebaños no mostró tener relación con la probabilidad de seropositividad, la imagen 6.1 muestra que los rebaños con un mayor porcentaje de ovejas seropositivas se encuentran en pequeños conglomerados al igual que los rebaños con un porcentaje medio y bajo, con lo cual se puede pensar que la convivencia y el compartir los campos donde pastar sí incide en la seropositividad.

Imagen 6.1 Distribución geográfica de los rebaños seropositivos a *Leptospira interrogans* por porcentaje de animales dentro del rebaño en el sitio de estudio



Seroprevalencia intra-rebaño (•) < 27.3%, (•) 27.4% a 46.7%, y (•) > 46.8%.

ANEXO A: ESTIMADOR DE MUESTREO BIETÁPICO ESTRATIFICADO

En la sección 5.3 se estimó la seroprevalencia de anticuerpos *anti-Leptospira* en las ovejas de la población en estudio, sin embargo no se describió el procedimiento por el cual se hizo dicha estimación. En este anexo se muestra la notación y las fórmulas necesarias para el cálculo.

El esquema de muestreo es bietápico estratificado en las unidades primarias (UPM), con muestreo simple en las dos etapas. Se desea estimar una proporción, la cual puede ser vista como un promedio de una variable binaria. Como se definió en la sección 5.2, la variable de interés puede ser definida como:

$$y_i = \begin{cases} 0 & \text{la oveja } i \text{ en muestra no es seropositiva} \\ 1 & \text{la oveja } i \text{ en muestra es seropositiva} \end{cases}$$

Una manera razonable de calcular la proporción es mediante el estimador de razón:

$$\hat{y}_i = \frac{\hat{Y}}{\hat{M}}$$

En el cual:

$$\hat{Y} = \sum_{i=1}^m w_i y_i \quad \text{Y} \quad \hat{M} = \sum_{i=1}^m w_i$$

Donde w_i es el peso de muestreo o factor de expansión de la unidad i en la muestra. Éste es el inverso de la probabilidad de selección de cada elemento. Para calcular dicha probabilidad definimos:

N_h = Número de UPM en el estrato h

n_h = Número de UPM en muestra en el estrato h

M_{hj} = Número de unidades secundarias de muestreo (USM), en la UPM j del estrato h

m_{hj} = Número de USM en muestra, en la UPM j del estrato h

Entonces la probabilidad de selección de la oveja i en muestra que pertenece al rebaño j del estrato h es:

$$\pi_i = \frac{n_h}{N_h} * \frac{m_{hj}}{M_{hj}} \quad \therefore \quad w_i = \frac{N_h}{n_h} * \frac{M_{hj}}{m_{hj}}$$

Para calcular los intervalos de confianza se supone normalidad en la distribución del estimador:

$$\hat{y}_i \pm z_{1-\alpha/2} * \widehat{Var}(\hat{y}_i)$$

Debido a que el esquema es estratificado, la muestra (de UPM) es tomada de manera independiente en cada estrato y por lo tanto la varianza total del estimador es la suma de las varianzas del mismo en cada uno de los estratos. Como en cada estrato se tiene un muestreo de dos etapas, la varianza total es una suma de varianzas de muestreos bietápicos (Cochran 1976).

ANEXO B: AIC Y BIC

Cuando se selecciona un modelo, se estará equivocado si se piensa que se ha encontrado el verdadero modelo. Cualquier modelo es una simplificación de la realidad.

Un modelo simple que ajusta adecuadamente tiene las ventajas de la parsimonia. Si un modelo tiene relativamente un pequeño sesgo, describiendo bien la realidad, éste tiende a proveer estimaciones más exactas de las cantidades de interés.

Otro criterio además de las pruebas de significancia y de la devianza, puede ayudar a seleccionar un buen modelo. El más conocido es el criterio de información de Akaike (AIC por sus siglas en inglés). Éste juzga un modelo por qué tanto los valores estimados ajustan a los valores reales. Se sabe que un modelo complejo ajusta mejor a los datos que un modelo más simple, sin embargo, se suele preferir el modelo más simple ya que éste tiende a calcular mejores estimaciones de ciertas características del "verdadero modelo". Así, el modelo óptimo será el que tienda a ajustar mejor a la realidad.

Dada una muestra, Akaike mostro que éste criterio selecciona el modelo que minimiza:

$$AIC = -2(\log \text{verosimilitud maximizada}) + 2p$$

Donde p es el número de parámetros en el modelo. Este criterio toma en cuenta la bondad de ajuste estadística y el número de parámetros que tienen que ser estimados para lograr ese nivel de ajuste, penalizando a los modelos con un mayor número de cantidades estimadas. Valores bajos del AIC indican un mejor modelo, esto es, el que tiene menos parámetros y aún sigue manteniendo un nivel adecuado de ajuste a los datos.

El criterio de información bayesiana, también es usado para comparar modelos, el cual se define:

$$BIC = -2(\log \text{verosimilitud maximizada}) + p * \log(n)$$

Donde p es el número de parámetros y n el tamaño de muestra. Este criterio penaliza con mayor fuerza a los modelos con más cantidades por estimar y mayor tamaño de muestra.

Cuando se usa el AIC o BIC, los modelos por comparar no necesitan estar anidados.

Glosario de términos

Anemia hemolítica aguda: Disminución del caudal hemoglobínico o del número de eritrocitos caracterizada por la destrucción rápida de los mismos con pobre regeneración ontogénica.

Anictérico: Sin coloración amarillenta en piel y mucosas.

Anorexia: Falta de apetito

Avirulento: Falta de toxicidad o infecciosidad del agente etiológico para producir malignidad y acrimonia.

Zoonótico: Estado morboso producido por un patógeno transmitido por animales.

Células endoteliales: Células planas que revisten las cavidades serósas.

Cepas saprofita: Cepa bacteriana no patógena que vive a expensas del organismo.

Cepa patogénica: Cepa bacteriana con capacidad para reproducir los postulados de Robert Koch.

Depresión: Disminución de la actividad vital en una parte o en la totalidad del cuerpo.

Disnea: Dificultad de respiración.

Excretas: Término general asignado al desecho de materia del cuerpo tales como sudor, heces y orina).

Extravasación de células: Salida de hematíes y/o células de los conductos circulatorios.

Fagocitosis: Proceso de ingestión y digestión de bacterias, partículas virales, debris celulares y/o cuerpos extraños por parte de los fagocitos principalmente los macrófagos.

Fenotipo: Conjunto de propiedades físicas y químicas manifiestas por un sujeto heredadas o no de sus antecesores.

Fiebre: Síndrome clínico integrado por hipertermia, taquicardia, taquipnea, estado saburral, quebrantamiento e intranquilidad o estupor.

Fibrinolisis: Disolución de la fibrina por acción de proteínas con actividad enzimática como la fibrinolisisina.

Genotipo: Contenido genómico que es específico de un individuo en forma de ADN y es el conjunto de genes de un organismo.

Hematógena: Producido en la sangre o derivado de ella.

Hemoglobinemia: Presencia anormal de hemoglobina en el plasma debido a la destrucción de eritrocitos.

Hemoglobinuria: Presencia de hemoglobina en la orina sin evidencia de eritrocitos.
Ictericia severa: Coloración amarilla de la piel y mucosas.

Inmunoglobulinas: Anticuerpos.

Lasitud: Debilidad, desfallecimiento, cansancio y/o agotamiento.

Leptospiremia: Infección por leptospiras del riñón y del aparato reproductor.

Leptospiruria: Fase prodrómica caracterizada por una creciente concentración de anticuerpos (aproximadamente después de una semana desde el inicio de la infección), la orina y la corteza renal en el estudio post mortem son el inóculo más adecuado para el aislamiento de leptospiras de los seres humanos. Los animales salvajes y domésticos en el estado de portador puede verter leptospiras intermitentemente durante muchos años o incluso toda la vida, tiempo durante el cual las leptospiras se pueden aislar de la orina y tejido renal.

Medidas contra epizooticas: Los indicadores de riesgo permiten recomendar las medidas contra epizooticas dirigidas a la disminución del riesgo entre la población susceptible, basadas en el ejercicio de un riguroso control sanitario y la identificación y segregación de los animales persistentemente infectados, las que comprenden las siguientes acciones fundamentales: Uso de bio-estimulantes del sistema inmunológico en los animales; selección de animales improductivos y/o longevos para enviarlos al sacrificio; observación y segregación de los animales enfermos y presuntos enfermos para su evaluación clínica; limpieza y desinfección, disposición de excretas y basuras, cuarentena sanitaria en el área afectada y vacunación.

Nefritis intersticial focal: Superficie renal con múltiples nódulos prominentes de coloración blanco-grisáceos (riñón moteado de blanco) debidos procesos inflamatorios proliferativos. Histológicamente es caracterizado por la presencia linfocitos, células plasmáticas e infiltración de histiocitos y fibrosis.

Patogénesis: Origen y desarrollo de la enfermedad, especialmente el modo como el estado morboso afecta al individuo.

Reservorio: El individuo que alberga agente patógenos propagadores de infecciones.

Respuesta humoral: Respuesta inmunológica mediada por anticuerpos y otras proteínas tales como el complemento.

Seroconversión: Estimulación de las células plasmáticas para producir anticuerpos dirigidos hacia el antígeno específico.

Títulos serológicos: Nivel y/o concentración de anticuerpos en el suero.

Transplacentario: Que atraviesa la membrana corio-amniótica invadiendo al feto.

Trombocitopenia: Disminución de la cantidad de plaquetas de la sangre.

Uveítis autoinmune: Inflamación de la cara posterior pigmentada del iris debido a la presencia de anticuerpos y/o complejos autoinmunes.

BIBLIOGRAFÍA.

CITAS DE LOS CAPÍTULOS 1, 2 Y 5

Agresti, Alan. *Categorical Data Analysis*, 2da edición. Wiley Interscience 2002. Segunda edición.

Agresti, Alan. *An Introduction to Categorical Data Analysis*. Wiley Interscience 2007. Segunda edición.

Bruin, J. 2006. newtest: command to compute new test. UCLA: Academic Technology Services, Statistical Consulting Group. <http://www.ats.ucla.edu/stat/stata/ado/analysis/>.

Cochran, William G. *Técnicas de muestreo*. Compañía editorial continental 1976. Sexta edición.

Collet, David. *Modelling Binary Data*. Chapman and Hall/CRC. Segunda edición. 2003.

Dobson, Annette. *An Introduction to Generalized Linear Models*. Chapman and Hall/CRC 2002. Segunda edición.

Fielding, Antony, Pillinger, Rebecca. LEMMA (Learning Environment for Multilevel Methods and Applications, curso en internet. University of Birmingham & Centre for Multilevel Modelling. <http://www.cmm.bris.ac.uk/lemma>.

García Pacheco, Maricela. *Modelos Lineales Mixtos Generalizados*. Tesis de maestría en ciencias matemáticas. UNAM, México D.F. 2006.

Hosmer, David, Lemeshow S. *Applied Logistic Regression*. Wiley Interscience 1989.

Kleinbaum, David. *Logistic Regression, A Self Learning Text*. Springer-Verlang 1994.

McCullagh P., Nelder J. *Generalized Linear Models*. Chapman and Hall 1983. Segunda edición.

McCulloch, Charles E., Serly R. *Generalized, Linear and Mixed Models*. Wiley Interscience. Segunda edición. 2008.

Méndez Ramírez, Ignacio. *El protocolo de investigación*. Trillas. México D.F.

Paul D., Allison. *Convergence Failures in Logistic Regression*. Paper 360-2008. University of Pennsylvania, Philadelphia. <http://www2.sas.com/proceedings/forum2008/360-2008.pdf>

Rincón Rentería, Karina. *Evaluación de la depresión a través de la regresión logística. Encuesta de evaluación del desempeño. México 2002-2003*. Tesis en Actuaría. UNAM, México D.F. 2006.

Rodriguez, Germán. Elo, Irma. Intra-class correlation in random-effects models for binary data. *The STATA Journal* (2003), Número 1, pp.32-46.

STATA Corporation Lakeway Drive, College Station, Texas 77845. *Manual de referencia STATA* 11. STATA Press 2009.

CITAS DE LOS CAPÍTULOS 3 Y 4

Acosta H., Moreno C. y Viáfara D. (1994). Leptospirosis: Revisión del tema. *Colombia Médica* 25: 36-42.

Web: <http://colombiamedica.univalle.edu.co/Vol25No1/leptospirosis.html>

Agunloye, C.A., 2002: Leptospiral agglutinating antibodies in sheep and goat in South-West Nigeria. *Israel J. Vet. Res.* 57, http://www.isrvma.org/article/57_1_6.htm

Andreani E, Tolari F, Farina R. (1983). Experimental infection in sheep with *Leptospira interrogans* serotype hardjo. *British Veterinary Journal* 139; 165-170.

Arteaga, T.G., V.M. Banda, R.I. Mendez, L.P. Moles, J.C. Rosales, J. Torres (1991) Leptospirosis bovina en el complejo agroindustrial Tizayuca, Hgo.: I Posibles implicaciones epidemiológicas, abstr., p. 293-298. XVI Congreso Nacional de Buiatria. Veracruz, Mexico.

Ashford D.A., Kaiser R.M., Spiegel R.A., Perkins B.A., Weyant R.S, Bragg S.L., Plikaytis B., Jarquin C., De Jose Reyes J.O., and Amador, J.J. (2000). Asymtomatic infection and risk factors for leptospirosis in Nicaragua. *Am. J. Trop. Med. Hyg.* 63 (5, 6): 249-254.

Aslanta, Ö. and V. Özdemür (2005). Determination of the Seroprevalence of Leptospirosis in Cattle by MAT and ELISA in Hatay, Turkey. *Turk J. Vet. Anim. Sci.* 29; 1019-1024.

Atxaerandio, R., G. Aduriz, I. Ziluaga, J.I. Esteban, L. Maranda, RC. Mainar-Jaime (2005). Serological evidence of *Leptospira interrogans* serovar Bratislava infection and its association with abortions in cattle in northern Spain. *Vet. Rec.* 156, 376380.

Bal, A. (2005). Unusual clinical manifestations of leptospirosis. Department of Medical Microbiology, Aberdeen Royal Infirmary, Aberdeen, Scotland, United Kingdom. Symposium. Vol. 51-3, p. 179-183.

Web:<http://www.jpgmonline.com/article.asp?issn=00223859;year=2005;volume=51;issue=3;page=179;epage=183;aulast=Bal>

Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, Lovett MA, Levett PN, Gilman RH, Willig MR, Gotuzzo E, Vinetz JM. (2003). Leptospirosis: a zoonotic disease of global importance. *Lancet Infectious Diseases* 3, 757-771.

- Beamer, PDJH, H., Morril, C.C. (1953). Studies on leptospirosis in domesticated animals I. Leptospirosis in sheep. *Veterinary Medicine* 48; 365-366.
- Bielanski AB, Surujballi O. (1998). *Leptospira borgpetersenii* serovar hardjo type hardjobovis in bovine embryos fertilized in vitro. *Canadian Journal of Veterinary Research* 62; 234-236.
- Blackmore, D.K., Bahaman, A.R., Marshall, R.B. (1982). The epidemiological interpretation of serological responses to leptospiral serovars in sheep. *New Zealand Veterinary Journal* 30; 38-42.
- Bovet, P., Yersin, C., Merien, F., Davis, C.E., Perolat, P. (1999). Factors associated with clinical leptospirosis: a population-based case-control study in the Seychelles (Indian Ocean). *Int J Epidemiol.* 28; 583-590.
- Brenner D.J., Kaufmann A.F., Sulzer K.R., Steigerwalt A.G., Rogers F.P., Weyant R.S. (1999). Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for *Leptospira alexanderi* sp. nov. and for 117 new *Leptospira* genomospecies. *International Journal of Systematic Bacteriology* 49; 839-858.
- Bundle, D.R., Cherwonogrodzky, J.W., Caroff, M., and Perry, M.B. (1987). The lipopolysaccharides of *Brucella abortus* and *B. melitensis*. *Ann. Institute Pasteur Microb.* 138; 92-98.
- Burriel, A.R., O.M. Vougiouka, S. Butsini, K. Nomikou, M. Patakakis (2002). A serologic investigation of some causes of reproductive failure among small ruminants in Greece Online *J. Vet. Res.* 6; 53-58.
- Carneiro, M., Giacomini L. y Costa M. (2004). Leptospirosis asociada a la exposición ocupacional: Estudio clínico y epidemiológico. *Revista chilena de infectología.* [online] Vol. 21, No.4, p. 339-344.
Web: http://scielotest.conicyt.cl/scielo.php?script=sci_arttext&pid=S0716-10182004000400008&lng=pt&nrm=iso. ISSN 0716-1018 [20/04/2006].
- Céspedes M., Chun M., Cano E., Huaranca I., Atoche H., Ortiz H., Valentín M., Balda L., Huamán T. (2007). Prevalencia de anticuerpos contra *Leptospira* en personas asintomáticas y en perros de Chancay, Lima 2001. *Rev Peru Med Exp Salud Publica* 24; 343-349.
- Cheville, N. F., Huhn, R., and Cutlip, R.C. (1980). Ultrastructure of renal lesions in pigs with acute leptospirosis caused by *Leptospira pomona*. *Vet. Pathol.* 17:338–351.
- Cox, T.E., Smythe, L.D. y Leung, L.K.-P. (2005). Flying foxes as carriers of pathogenic. *Journal of Wildlife Diseases* 41; 753-757.
- Cousins, D.V., G.M. Robertson, 1986: Use of enzyme immunoassay in a serological survey of leptospirosis in sheep. *Australian Vet. J.* 63, 36-39.

Davidson, J.N., Hirsh, D.C. (1980). Leptospirosis in lambs. *Journal of the American Veterinary Medical Association* 176; 124-125.

Ellis TM, Hustas L, Robertson GM, Mayberry C. (1984). Kidney disease of sheep, associated with infection by leptospire of the Sejroe serogroup. *Australian Veterinary Journal* 61; 304-306.

Ellis, W. A., O'Brien, J. J., Neill, S. D. and Bryson, D. G. (1986). Bovine leptospirosis: experimental serovar hardjo infection. *Veterinary Microbiology* 11, 293-299.

Ellis, W.A. (1994). Leptospirosis as a cause of reproductive failure. *Vet. Clin. North Am.: Food Ani. Prac.* 3; 463-478.

Everard, C.O.R., Bennett, S., Edwards, C.N., Nicholson, G.D., Hassell, T.A., Carrington, D.G., et al. (1992). An investigation of some risk factors for severe leptospirosis in Barbados. *Journal of Tropical Medicine and Hygiene* 95; 13-32.

Faine S. Guidelines for the control of leptospirosis. Faine S. (Ed.). WHO Offset Publication, 67. World Health Organization, Geneva, Switzerland, 1982.

Faine, S. *Leptospira and leptospirosis*. CRC Press, Inc., Boca Raton, Fla., E.U., 1994.

Faine S, Adler B, Bolin C, Perolat P. *Leptospira and Leptospirosis*, 2nd edn. MediSci, Melbourne, Australia, 1999.

Ganoza Ch A., Matthias MA., Collins-Richards D., Brouwer KC., Cunningham CB., Segura ER., Gilman RH., Gotuzzo E., Vinetz JM. (2006). Determining risk for severe leptospirosis by molecular analysis of environmental surface waters for pathogenic leptospira. *PLoS Med* 3(8): e308. DOI: 10.1371/journal.pmed.0030308.

Genovez ME., Oliveira JC., Castro V., Del Fava AC., Ferrari Cl., Pituco EM., Scarcelli E., Cardoso MV., Grasso, LMPS., Santos S.M. 2001. Desempenho reprodutivo de um rebanho Nelore de criação extensiva com leptospirose endêmica: estudos preliminares. *Rev. Bras. Reprod. Anim.* 25; 244-246.

Gerritsen M.J., Koopmans M.J., Paterse D., Olyhoek T. (1994). Sheep as maintenance host for *Leptospira interrogans* serovar hardjo subtype Hardjobovis. *Am. J. Vet. Res.* 55; 1232-1237.

Gordon, L.M. (1980). Isolation of *Leptospira interrogans* serovar hardjo from sheep. *Australian Veterinary Journal* 56, 348-349.

Gutián J., Thurmond MC., Sharon KH. 1999. Infertility and abortion among first-lactation dairy cows seropositive or seronegative for *Leptospira interrogans* serovar hardjo. *J. Am. Med. Assoc.* 215; 515-518.

Hansen, L.E. Pathogenesis of leptospirosis. In *The biology of parasitic spirochetes*. Johnson C.R. Academic Press. Dep. Microbiology. U. Minnesota. USA 295-349. 1976.

Hathaway SC, Marshall RB. (1979). Experimental infection of sheep with *Leptospira interrogans* serovars hardjo and balcanica. New Zealand Veterinary Journal 27; 197-

Hathaway, S.C. (1981). Leptospirosis in New-Zealand - an ecological view. New Zealand Veterinary Journal 29; 109-112.

Hathaway SC, Little TWA, Stevens AE. (1982). Serological survey of leptospiral antibodies in sheep from England and Wales. Veterinary Record 110; 99-101.

Heath, S.E., and Johnson, R. (1994). Leptospirosis. JAVMA 205; 1518-1523.

Jansen A, Schneberg I, Frank C, Alpers K, Schneider T, et al. (2005). Leptospirosis in Germany, 1962-2003. Emerging Infect Dis 11: 1048-1054.

Kupek E, de Sousa Santos Faversoni MC, de Souza Philippi JM (2000). The relationship between rainfall and human leptospirosis in Florianopolis, Brazil, 1991-1996. Braz J Infect Dis 4: 131-134.

Laguna, V. (2000). Leptospirosis. Módulos Técnicos. (Oficina General de Epidemiología / Instituto Nacional de Salud). Serie Documentos Monográficos. No.2, Lima. Web:<http://www.ins.gob.pe/insvirtual/images/otrpubs/pdf/M%C3%B3dulo%20T%C3%A9cnico%20de%20leptospirosis.PDF>. [20/11/2004] 35.

Levett P.N. (2001). Leptospirosis. Clin Microbiol Rev 14; 296-326.

Lilenbaum, M., Santos, M.R.C. (1996). Effect of management systems on the prevalence of bovine leptospirosis. Vet. Rec. 138; 570-571.

Lilenbaum, W., Ristow, P., and Almeida, F. (2002). Evaluation of a rapid slide agglutination test for the diagnosis of acute canine leptospirosis. Revista Latinoamericana de Microbiología. Vol. 44, No.3-4, p. 124 – 128. Web:http://www.medigraphic.com/espanol/e-htms/e-lamicro/e-mi2002/e-mi02-3_4/emmi02-3_4d.htm [03/05/2006]

Manktelow, B.W., and Thompson, J.C., (1989). Pathogenesis of renal lesions in haemoglobinaemic and non-haemoglobinaemic leptospirosis. J. Comp. Pathol. 101: 201-214.

Marshall, R.B., B.W. Manktelow (2002). Fifty years of leptospirosis research in New Zealand: a perspective. New Zealand Vet. J. 50, 61-63.

Matsuo, K., E. Isogai, and Y. Araki (2000). Utilization of exocellular mannan from *Rhodotorula glutinis* as an immunoreactive antigen in diagnosis of leptospirosis. J. Clin. Microbiol. 38; 3750-3754.

Matthias MA, Ricaldi JN, Cespedes M, Diaz MM, Galloway RL, et al. (2008). Human leptospirosis caused by a new, antigenically unique leptospira associated with a *Rattus* species

reservoir in the Peruvian Amazon. *PLoS Negl Trop Dis* 2 (4): e213. doi:10.1371/journal.pntd.0000213

McCaughan CJ, Gordon LM, Rahaley RS, Slee KJ, Presidente PJA. (1980). Evidence for infection of sheep in Victoria with leptospires of the Hebdomadis serogroup. *Australian Veterinary Journal* 56; 201-202.

McDonough, (2001). Leptospirosis en caninos-estado actual. Department of Population Medicine and Diagnostic Science, Diagnostic Laboratory, College of Veterinary Medicine, Cornell University, Ithaca, New York, USA. A0112.0701.ES. Web: http://www.ivis.org/advances/Infect_Dis_Carmichael/mcdonough_es/chapter_frm.asp?LA=203/05/2006

McKeown, J.D., Ellis, W.A. (1986). *Leptospira hardjo* agalactia in sheep. *Veterinary Record* 118; 482-

Pate GE, Hogan MC, FitzSimon N, Mellotte GJ (2000). A review of the epidemiology of leptospirosis in the Republic of Ireland. *Ir Med J* 93: 114-117.

Perdomo, E. y Garin, A. Leptospirosis animal. En: Guía de Control y Manejo de Leptospirosis. OPS/HCP/HCV/URU.ZOO. 24-26. 2002.

Pereira da Silva J.J., Dalston M.O, Manhães de Carvalho J.E., Setúbal S., Chiarino de Oliveira J.M., and Pereira M.M. (2002). Clinicopathological and immunohistochemical features of the severe pulmonary form of leptospirosis. *Revista da Sociedade Brasileira de Medicina Tropical* 35 (4): 395-399.

Perret PC., Abarca VK., Dabanch PJ., Solari GV., García CP., Carrasco LS., Olivares CR., Avalos P. (2005). Prevalencia y presencia de factores de riesgo de leptospirosis en una población de riesgo de la Región Metropolitana. *Rev Méd Chile* 133; 426-431.

Prescott, J.F. Leptospirosis, In: Jubb K.V.F., Kennedy P.C., Palmer N. (Eds.) *Pathology of domestic animals*. Academic Press, Inc, 4th edition. 503-511. 1993

Radostits, O.M., Gay, C.C., Blood, D.C., Hinchcliff, K.W. *Veterinary medicine. A textbook of the diseases of cattle, sheep, pigs, goats and horses*. W.B. Saunders Company Ltd, 1881 pp. London, UK. 1999.

Seijo, A., Coto, H. y San Juan J. (2002). Distres respiratorio debido a hemorragia pulmonar por Leptospirosis. *Medicina (Buenos Aires)*. Vol. 62, p.136-140. Web:<http://www.medicinabuenosaires.com/vol62-02/2/leptospirosis.htm>

Scott,G. and Coleman,T.J. In Cook, G.C. (ed.), *Manson's Tropical Diseases*, 20th edn. B. Saunders Company, London, UK. 1996.

Smith, B.P., Armstrong, J.M. (1975). Fatal hemolytic anemia attributed to leptospirosis in lambs. *Journal of the American Veterinary Medical Association* 167; 739-741.

Songer, G. and Thiermann, A. (1995). Zoonosis Update: Leptospirosis. Department of Veterinary Science, College of Agriculture, University of Arizona, Tucson, US Department of Agriculture, Agriculture Research Service, National Program Staff, Beltsville. Arizona technical paper No. 5022. Web: <http://www.avma.org/noah/resources/zoonosis/znleptos.asp>

Sterling, C.R., and Thiermann A.B. (1981). Urban rats as chronic carriers of leptospirosis: an ultrastructural investigation. *Vet. Pathol.* 18: 628-637.

Sullivan ND. (1974). Leptospirosis in animals and man. *Australian Veterinary Journal* 50; 216-223.

Sunbul, M., Esen, S., Leblebicioglu, H., Hokelek, M., Pekbay, A., Eroglu, C. (2001). *Rattus Norvegicus* acting as reservoir of *Leptospira interrogans* in the middle black sea region of Turkey, as evidenced by PCR and presence of serum antibodies to *Leptospira* strain. *Scan. J. Infec. Dis.* (33) 896-898.

Tangkanakul W, Tharmaphornpil P, Plikaytis BD, Bragg S, Poonsuksombat D, et al. (2000). Risk factors associated with leptospirosis in northeastern Thailand, 1998. *Am J Trop Med Hyg* 63; 204-208.

Tooloei, M., Abdollahpour, G., Karimi, H., and Hasanpor, A. (2008). Prevalence of serum antibodies against six *Leptospira* serovars in sheep in Tabriz, Northwestern Iran. *J. An. Vet. Adv.* 7: 450-455.

Trevejo RT, Rigau-Perez JG, Ashford DA, McClure EM, Jarquin-Gonzalez C, Amador JJ, et al. (1998). Epidemic leptospirosis associated with pulmonary hemorrhage-Nicaragua, 1995. *J Infect Dis* 178: 1457-1463.

Vado-Solis I., M.F. Cardenas-Marrufo, B. Jimenez-Delgadillo, A. Alzina-Lopez, H. Laviada-Molina, V. Suarez-Solis, and J.E. Zavala-Velazquez (2002). Clinical-Epidemiological study of leptospirosis in humans and reservoirs in Yucatan, Mexico. *Rev. Inst. Med. Trop. S. Paulo* 44, 335-340.

Vermunt JJ, West DM, Arthur DG, Marshall RB. (1994a). Leptospirosis in a lamb. *New Zealand Veterinary Journal* 42; 155-

Vermunt, J.J., West, D.M., Cooke, M.M., Alley, M.R., Collinsemerson, J. (1994b). Observations on three outbreaks of *Leptospira interrogans* serovar pomona infection in lambs. *New Zealand Veterinary Journal* 42; 133-136.

Watt G, Padre LP, Tuazon M, Calubaquib C. (1990). Skeletal and cardiac muscle involvement in severe, late leptospirosis. *J Infect Dis* 162: 266-269.

BIBLIOGRAFÍA

Yasuda PH, Hoshino-Shimizu S, Yamashiro EH, de Brito T. (1986). Experimental leptospirosis (*L. interrogans* serovar icterohaemorrhagiae) of the guinea pig: leptospiral antigen, gamma globulin and compliment C3 detection in the kidney. *Experimental Pathology* 29; 35-45.