

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

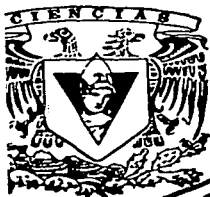


FACULTAD DE CIENCIAS

"ESTIMACIÓN DE LA PROBABILIDAD DE UN PARTO MÚLTIPLE EN BORREGAS PELIBUEY CON UN MODELO DE REGRESIÓN LOGÍSTICA"

T E S I S
QUE PARA OBTENER EL TITULO DE:
A C T U A R I O
P R E S E N T A :
CLAUDIA ESTHER ALCALÁ ESCAMILLA

DIRECTORA DE TESIS: E en E.A. MARIBEL JADALUPE SÁNCHEZ GONZÁLEZ



TESIS CON FALLA DE ORIGEN

DIVISION DE ESTUDIOS PROFESIONALES



FACULTAD DE CIENCIAS SECCION ESCOLAR



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.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**M. EN C. ELENA DE OTEYZA DE OTEYZA**  
Jefa de la División de Estudios Profesionales de la  
Facultad de Ciencias  
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

"Estimación de la probabilidad de un parto múltiple en borregas Pelibuey con un modelo de regresión logística".

realizado por Alcalá Escamilla Claudia Esther.

con número de cuenta 09337758-4, quién cubrió los créditos de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis  
Propietario

E. en E.A. María Guadalupe Sánchez González

Propietario

M. en C. José Antonio Flores Díaz

Propietario

M. en A.P. María del Pilar Alonso Reyes

Suplente

Act. Lucio Gerardo Chávez Heredia

Suplente

Act. Jaime Vázquez Alamilla

Consejo Departamental de Matemáticas

M. en C. JOSE ANTONIO FLORES DIAZ

FACULTAD DE CIENCIAS  
CONSEJO DEPARTAMENTAL  
DE  
MATEMÁTICAS

*A mi Mamá,  
que con su esfuerzo por facilitarnos la vida  
me ha dado la oportunidad de llegar hasta aquí; y*

*a E. E. A. María Guadalupe Sánchez González,  
por su invaluable apoyo en la realización de esta Tesis  
y por su enorme paciencia,*

**MI MÁS SINCERO Y PROFUNDO  
AGRADECIMIENTO.**

---

# CONTENIDO

---

<b>INTRODUCCIÓN</b>	<b>2</b>
<b>CAPÍTULO I</b>	<b>4</b>
1.1. PRODUCCIÓN OVINA	4
1.2. DESCRIPCIÓN DEL BORREGO PELIBUEY	7
1.3. DEFINICIÓN DE LA POBLACIÓN	14
1.3.1. DESCRIPCIÓN DE LAS VARIABLES	14
<b>CAPÍTULO II</b>	<b>16</b>
2.1. MODELO SIMPLE DE REGRESIÓN LOGÍSTICA	16
2.2. MODELO MÚLTIPLE	32
2.3. AJUSTE DEL MODELO	40
2.3.1. BONDAD DE AJUSTE	40
2.3.2. RESIDUOS	45
2.3.3. MEDIDA DE ASOCIACIÓN MÚLTIPLE ENTRE LAS VARIABLES INDEPENDIENTES Y LA VARIABLE DEPENDIENTE.	47
<b>CAPÍTULO III</b>	<b>50</b>
3.1. RESULTADOS Y DISCUSIÓN	50
<b>CONCLUSIONES</b>	<b>61</b>
<b>BIBLIOGRAFÍA</b>	<b>62</b>

# **INTRODUCCIÓN**

La estadística como herramienta para extraer información de un conjunto de datos es ampliamente utilizada en una gran diversidad de campos de estudio; en particular, los métodos de regresión proporcionan una metodología útil en el estudio y en la búsqueda de las relaciones existentes entre características de una población, lo cual ayuda en la toma de decisiones o aporta opciones en la solución de problemas.

En este estudio, lo que se busca es brindar información acerca de un problema que frecuentemente se encuentra en las explotaciones de animales: producir más, este caso se refiere a los borregos Pelibuey. *El problema es saber cuáles son los factores que influyen principalmente en la cantidad de borregos que se obtienen del rebaño productor, es decir, el tamaño de camada por oveja*, para este fin se tienen variables como el tipo de parto anterior, el intervalo entre partos, el tipo de nacimiento de la madre o el número de partos.

Los métodos de regresión son útiles al momento de buscar relaciones entre variables, en el caso de esta tesis la regresión logística se consideró la herramienta adecuada, al ser la variable de respuesta estudiada de tipo categórico (tamaño de la camada).

En el primer capítulo se ofrece una breve introducción de lo que es la producción ovina junto con una descripción del borrego Pelibuey, además de que se presentan las variables independientes del estudio y la información referente a la explotación de la cual proceden los datos utilizados para el análisis.

En el segundo capítulo se da una introducción a la regresión logística, tanto simple como múltiple, así como algunas medidas que son utilizadas para determinar la calidad del ajuste del modelo.

Finalmente, en el tercer capítulo, se muestran los resultados obtenidos al aplicar el modelo propuesto junto con la discusión de estos resultados.

# *CAPÍTULO I*

---

## **1.1. PRODUCCIÓN OVINA**

---

La reproducción ejerce una marcada influencia sobre la eficiencia productiva de cualquier explotación ganadera, en ovinos cuanto mayor sea la tasa de reproducción tanto mayor será el número de animales que se puedan destinar a la venta, lo que implica una mayor producción de carne, lana, piel y otros productos. Los porcentajes de calidad de estos productos que se logran en una explotación son la consecuencia de una diversidad de factores que influyen sobre cada individuo del rebaño. Por otro lado, la disponibilidad de un mayor número de reemplazos hace posible aumentar los criterios de selección, lo que conduce a una mejoría genética y al consiguiente aumento de la potencialidad productiva de los animales estudiados[7].

Se considera que el número de corderos criados hasta la edad o el peso de mercado, por oveja y por año es un factor que influye en los beneficios de explotaciones comerciales de ovinos[6]. Con el aumento en la demanda de productos de origen animal, las explotaciones se han visto en la necesidad de incrementar su producción[7].

Como consecuencia de lo anterior se tiene la necesidad de introducir cambios en el manejo reproductivo a fin de obtener mayor número posible de corderos durante la vida reproductiva de una borrega. Esto significa que hay que adelantar tanto como sea posible la edad del primer parto, aumentar el número de corderos vivos al nacimiento y acortar los intervalos entre partos, al mismo tiempo se busca incrementar las ganancias de peso por animal y



reducir su mortalidad[7]; en ciertos casos el número óptimo de corderos puede depender de la disponibilidad de alimentos u otras variables que afectan el número de corderos[6].

El número de corderos nacidos (tipo de parto, simple o múltiple) se puede aumentar de distintas maneras [6][8]:

- Empleando razas de ovejas que tengan muchas crías. Existen razas que son populares por su gran prolificidad, por ejemplo el Pelibuey. En el primer parto, algunas ovejas paren un solo cordero (simple), otras dos y unas cuantas tres (múltiple). En los siguientes partos las hembras tienden a repetir el tipo de parto ocurrido en los anteriores, a esta característica se le ha dado el nombre de "repetibilidad"[8].
- Empleando razas no estacionales. La oveja es un animal poliéstrico<sup>1</sup> estacional, es decir, su época natural de apareamiento se limita al otoño[7], cuando la relación entre las horas luz y oscuridad comienza a disminuir, lo cual se conoce como fotoperíodo [7]; esto conlleva a que la oveja sólo tenga un parto por año, influyendo en la producción de corderos. Es posible aumentar el número de partos de una oveja por año o reducir el intervalo entre partos controlando el fotoperíodo o seleccionando razas no estacionales, por ejemplo, el Pelibuey.
- Controlando la edad en la que alcanzan la pubertad las corderas, ya que, teniendo en cuenta la naturaleza estacional de la

---

<sup>1</sup> El estro es la etapa de receptividad de la hembra hacia el macho. Poliéstrico significa que tiene varios estros.

reproducción en la oveja, la dificultad de alcanzar la pubertad a una edad determinada puede retrasar hasta doce meses el inicio de la actividad reproductiva efectiva.

El número de corderos por hembra cubierta aumenta con la edad hasta alcanzar un máximo a los 3-5 años y comienza a descender rápidamente a partir de los 8-10 años.[8] Rojas *et al.* encontraron que, en ovejas Blackbelly, el tamaño de camada para animales de 1 año es de  $(1.32, 1.48)^2$  corderos por parto, para borregas jóvenes; de 2 a 4 años, se reporta  $(1.61, 1.73)$  crías, para ovejas de entre 5 a 7 años se observó un intervalo de  $(1.68, 1.82)$  y para borregas viejas, de 8 a 10 años, se registraron  $(1.51, 2.09)$  crías. Sin embargo muy pocas ovejas se mantienen en el rebaño reproductor hasta edad tan avanzada[8][18].

Además del incremento en la eficiencia reproductiva que puede lograrse, se obtienen las siguientes ventajas utilizando una raza como el borrego Pelibuey :

- I. La producción de corderos es más uniforme a lo largo del año, lo que supone una ventaja para los productores.
- II. Mejoría en el aprovechamiento de instalaciones, como zonas de partos, de alimentación, etc.
- III. Reparto de los riesgos provocados por las malas condiciones o los mercados desfavorables a lo largo del año.

---

<sup>2</sup> Todos los intervalos descritos en esta tesis son de la forma  $(\bar{x} \pm \sigma)$  con  $\bar{x}$  el promedio y  $\sigma$  la desviación estándar.

- IV. Proporcionar a los productores ingresos uniformes a lo largo del año.

---

## 1.2. DESCRIPCIÓN DEL BORREGO PELIBUEY

---

El borrego Pelibuey es una raza de ovinos productores de carne que por su rusticidad<sup>3</sup> y su fácil manejo que favorece su estancia en pequeñas explotaciones familiares, se ha difundido enormemente en el territorio mexicano[14][19][21]. El nombre de Pelibuey es de carácter popular regional y se origina del parecido que tiene con el pelo del ganado bovino; algunos autores lo denominan borrego Tabasco[5][13]. Dada precisamente su rusticidad es un animal que se adapta fácilmente a cualquier tipo de praderas; existen datos referentes a su explotación bajo condiciones intensivas y de obtener rendimientos de carne por hectárea favorables[16]. Se encuentra ampliamente distribuido en la región sureste de México, principalmente en los estados de la costa del Golfo de México y península de Yucatán, aunque recientemente ha tenido mayor difusión en los estados de la costa del Pacífico[16][21].

El Pelibuey es de cabeza mesocéfala, frente ancha, redondeada, sin cuernos aunque en algunos machos es fácil notar pequeñas yemas sobre el cráneo[3], presenta dos depresiones notables atrás de los arcos orbitales, sus órbitas son salientes, su cara es de longitud y anchura medianas, tiene grandes glándulas sebáceas debajo del ángulo comisural interno de los ojos, su perfil es rectilíneo, cubierto de pelo corto y fino, su piel también es fina y

---

<sup>3</sup> Rusticidad: adaptabilidad a diferentes ecosistemas.

adherente. Las orejas son cortas y lanceoladas, cubiertas de pelo corto, fino y suave que llevan horizontalmente; los ojos son grandes, poco prominentes, de color café o verde claro; su boca es pequeña, con labios fuertes siendo el superior hundido en la parte media, las mucosas ocular, nasal y bucal se pueden presentar en rosado y pigmentadas. Su cuello es corto, fuerte y redondeado; el macho presenta en la mayoría de los casos pelo largo desde la protuberancia occipital hasta la región de la cruz e inferiormente desde la región faríngea hasta la entrada del pecho. Este pelo no se encuentra en las hembras, en las que el cuello es más fino y delgado; en ocasiones presentan pedúnculos epiteliales (mamellas) en la zona faríngea[13][16][21] .

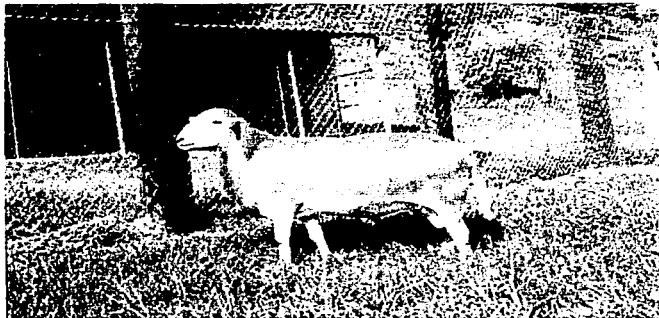


Su cuerpo es cilíndrico con la cruz prominente, línea dorsal o ensillada, grupa recta o ligeramente caída, la cola delgada de inserción baja y de una longitud aproximada de 20 cm. Generalmente con la porción terminal de color blanco. Las costillas son arqueadas anchas y con gran capacidad, abdomen voluminoso y caderas fuertes y redondeadas. La piel ligeramente

---

## CAPÍTULO I

adherida se encuentra cubierta de pelo y una capa inferior corta de lana. Las extremidades son de tamaño medio, delgadas y finas, bien aplanadas cubiertas de pelo corto, presentan una voluminosa glándula sebácea en el espacio interdigital la cual es típica de los ovinos; sus pezuñas son de color claro o pigmentadas.



El Pelibuey puede presentar 2 tipos de color: café, y blanco y la mezcla de estos colores, el pinto (marca grande delimitada). [13][16][21]

Los borregos Pelibuey presentan su pubertad a una edad de (239, 361) días teniendo un peso de (15.51, 30.5) kg. La duración del estro o celo es de (29.3, 30.1) horas, aunque se puede encontrar un rango más amplio de 24 a 36 horas[13] ó de 24 a 48 horas[22]; tiempo en el que la hembra acepta al macho. El período comprendido entre la presentación de un estro y otro es de (16,19) días[21][22], también se habla de un rango de 16 a 20 días con una media de 17[13], Cruz *et al.* reportan una duración del ciclo estral de

(16.9,18.7) días[5]. El servicio por concepción (el número de montas o servicios que una borrega necesita para quedar preñada) es de 1.1 a 1.5 montas por concepción.[21][22]

La raza Pelibuey ha demostrado una excelente eficiencia reproductiva, del 80.3% con un promedio de 1.23 corderos por hembra, según el estudio de Segura *et al.*[20] Esto repercute en forma directa en una mayor producción de corderos al año, en comparación con otras especies[16][22]. Las ovejas Pelibuey presentan estro o celo durante la mayor parte del año, esto quiere decir que se pueden obtener hasta tres partos cada dos años, ya sea que programen los empadres o se mantengan las ovejas en empadre continuo[11].

Sin embargo, los resultados de Cruz *et al.* mostraron que la presentación de celos durante un año varía del 81.25% de las ovejas en abril al 100% en agosto, en general se observó una disminución del estro entre los meses de marzo a junio. Esto puede ser debido a la calidad de los forrajes presentes en la región entre febrero y mayo, lo que ocasiona pérdidas de peso y deterioro en las condiciones corporales de los borregos[5]. Según Morales, de aproximadamente 100 borregas, 85 paren y de éstas se pueden obtener de 110 a 119 corderos nacidos[16].

Los corderos nacidos de un parto múltiple pesan (1.76, 2.66) kg., lo cual es menos que los de un parto simple, (2.31, 3.27)kg. Los machos al nacimiento tienen un peso superior al de las hembras, (2.19, 3.29) kg. y (1.92, 2.96) kg. respectivamente[16][22]. Se ha notado una relación entre el vigor al nacer, con la ganancia de peso, es decir, los corderos con mejor peso se desarrollan más rápido; también los corderos que no alcanzan los 1000 grs. de peso al nacer difícilmente sobreviven en la lactancia, sobre todo en

condiciones de pastoreo. Esto debe tomarse en cuenta cuando el parto sea múltiple, para dar cuidados especiales durante los primeros 10 o 15 días[13].

Los corderos se pueden destetar a partir de los 75 días de nacidos, cuando las ovejas están en buenas condiciones, la edad del destete puede prolongarse, puesto que las borregas pueden quedar preñadas aún estando en lactancia[16].

Se recomienda que la oveja se cubra entre los 40 y 60 días después del parto[13], aunque también se menciona un ciclo de 25 a 60 días[21][22].

La duración promedio de la gestación es de aproximadamente 5 meses[13], Morales menciona un período de (147.1 ,151.7) días; sin embargo, puede observarse una variación en este período de acuerdo al tipo de parto, en el Cuadro 1 se presentan los promedios de duración de gestación[16]:

TIPO DE PARTO	DURACIÓN EN DÍAS (promedio $\pm$ desviación estándar)
Simple	149.5 $\pm$ 2.3
Doble	149.3 $\pm$ 2.4
Triple	148.9 $\pm$ 1.6

Cuadro 1. Duración promedio de la gestación según el tipo de parto.

Aunque existe poca variación en la duración de gestación de acuerdo entre los partos simples, dobles, triples, existe la tendencia a disminuir el periodo a medida que el número de corderos aumentan[16]. Otros

investigadores toman como promedio de la duración de gestación (146.6, 152.8) días[21], valor que corresponde al de otras razas ovinas[22].

El intervalo entre partos es el tiempo comprendido entre un parto y otro[13], y es probablemente el mejor parámetro para evaluar la eficiencia reproductiva. El intervalo ideal para la raza Pelibuey es de 7.5 meses[16] (216 días), pero puede considerarse de hasta 8.2 meses[21] (255 días), o también se estima en 248 días[13], esto considerando que el rebaño se encuentre en empadre continuo y no existan influencias de clima o alimentación[16].

La prolificidad o tamaño de camada se define como el número de corderos nacidos por parto en la raza pelibuey. Se ha observado que la incidencia de parto múltiple, es decir, más de un cordero, se encuentra dentro de un rango de 18.8 a 39.9 % con media de 19.6%[22], este valor representa de 118 a 140 corderos nacidos por 100 ovejas paridas[16].

Por otro lado, Segura *et al.* encontraron las siguientes cifras: 81.9% para partos simples, 18.1% para gemelos, 0.5% para triates[20]. Según González *et al.* el tamaño promedio de camada es de (0.8,1.6) crías[10], el modelo de Rojas *et al.* reporta una media de 1.63 corderos por parto en ovejas Blackbelly[18]. Segura *et al.* reportaron una media de 1.23 corderos por parto[20]. Campos *et al.*, al realizar una comparación de la productividad de las 3 variedades de color de pelibuey reportaron los siguientes porcentajes de tipo de parto[4]:



COLOR DE LA HEMBRA	PARTO SIMPLE	PARTO DOBLE	PARTO TRIPLE O SUPERIORES
Blanco	46.68	47.39	5.93
Canelo	41.34	51.93	6.73
Pinto	47.52	46.69	5.79

Cuadro 2. Porcentajes de tipo de parto.

El manejo alimenticio al que se someten los animales antes y después de la época de servicio tienen influencia sobre el número de corderos nacidos por borregas paridas[16].

El tipo de nacimiento de la madre también puede ser un factor que influya en el tamaño de camada, en el estudio de Rojas *et al.* se encuentra que las ovejas de raza Blackbelly que provienen de madres de parto doble y triple tuvieron significativamente ( $p < 0.01$ ) un mayor tamaño de camada que aquellas que provenían de parto simple[18]:

TIPO DE PARTO DE LA MADRE	TAMAÑO DE CAMADA (media $\pm$ desviación estándar)
Simple	1.457 $\pm$ 0.08
Doble	1.705 $\pm$ 0.08
Triple	1.795 $\pm$ 0.11
Cuádruple	1.636 $\pm$ 0.20

Cuadro 3. Tamaño de camada según el tipo de parto de la madre

Según el estudio de Segura *et al.*, las borregas con pesos superiores a los del promedio del rebaño obtienen mayor productividad, prolificidad y tamaño de camada que aquellas que se encuentran debajo del promedio. Es posible que las hembras con peso inferior al promedio de su grupo se vean restringidas en la obtención de energía lo cual tiene un efecto directo en la prolificidad[20].

---

### 1.3. DEFINICIÓN DE LA POBLACIÓN DE ESTUDIO

---

Los datos se obtuvieron del rancho "San Jacinto", en Chalma, Estado de México. El clima de la zona es semi-cálido sub-húmedo con lluvias en verano ((A)Ca(wz)(w)ig), con temperatura promedio anual de 19.8° C y una precipitación pluvial de 1,191.9 mm[9].

Se utilizaron los registros de tipo de parto, en total 2250, de 642 borregas Pelibuey que se encontraban entre el 2° y 15° parto. El manejo de los rebaños consiste en pastoreo diurno, las hembras y sus corderos en un potrero y los machos en otro. Las praderas son de Estrella de Surinam (*Cyndon nlemfuensis*). Por la tarde se encierran a los animales en corrales en donde se les da un concentrado con 16% de proteína y sales minerales a libre acceso.

#### 1.3.1. DESCRIPCIÓN DE LAS VARIABLES

---

Las variables utilizadas en el estudio para explicar el tipo de parto fueron: intervalo de parto anterior, el número de parto anterior, el tipo de

parto anterior y los grupos de años, de las cuales sólo el intervalo al parto es variable continua, las demás son categóricas.

**Las variables independientes utilizadas en este trabajo fueron:**

- ∴ Tipo de Nacimiento de la Madre. Los registros de todas las borregas fueron divididas en tres grupos y la clasificación de las hembras en cada uno de estos grupos se hizo dado el criterio de selección empleado en la explotación para las hembras de reemplazo: en el Grupo 1 las hembras provenientes de parto simple entre los años 1987 y 1990; en el Grupo 2, las nacidas de parto simple o múltiple entre 1991 y 1995; y en el Grupo 3 aquellas de parto múltiple únicamente, entre 1996 y 2001.
- ∴ Número de Parto. Se cuentan con ovejas desde el 2° parto al 15<sup>avo</sup>. Los registros de éstas se dividieron en tres grupos: aquellos de animales entre el 2° y 5° parto, entre 6° al 10° y 11<sup>avo</sup> al 15<sup>avo</sup>.
- ∴ Tipo de Parto Anterior. Se codifica 0 en caso de simple (la borrega parió un sólo cordero) o 1 si es múltiple (el animal parió 2 o más crías).
- ∴ Intervalo de Parto. Se encuentra en el rango de 173 a 551 días, con una media de 245.144 y una desviación estándar de 46.87 días.

Como variable dependiente se tiene el Tipo de Parto, múltiple o simple, éstos fueron codificadas con 0 en caso de parto simple y 1 para parto múltiple.

# *CAPÍTULO II*

---

## **2.1. MODELO SIMPLE DE REGRESIÓN LOGÍSTICA**

---

Los métodos de regresión son un componente substancial de cualquier análisis de datos concerniente a la descripción de la relación entre la variable respuesta y una o más variables explicativas. Es frecuente el caso en el que la variable dependiente es discreta, tomando dos o más posibles valores. El objetivo de la regresión logística es encontrar un modelo sencillo y razonable que describa la relación entre un resultado denominado variable dependiente y un conjunto de variables independientes llamadas predictoras o explicativas[12].

La diferencia entre regresión logística y regresión lineal es que la variable dependiente del modelo logístico es categórica, la cual puede tener dos o más niveles o categorías[1][12]. Además, en regresión logística, la función que describe la distribución del número de “éxitos” es la binomial y no la normal como en el caso de regresión lineal. Se supone  $Y$  la variable respuesta dicotómica, entonces ésta se codifica con los siguientes dos valores: 0 ó 1. La suma de todos los resultados es la suma de unos. El promedio como se conoce, la suma de todas las variables entre el número total de la muestra, no es más que la proporción de unos.

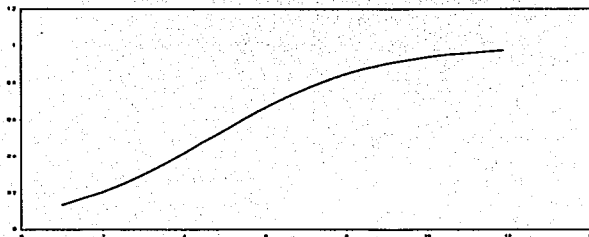
En la regresión lineal se modela la media de la variable dependiente, en el caso de variables dicotómicas se modela la proporción de respuestas de una de las dos categorías de la variable respuesta[2].

Sea entonces  $\Pi = E[Y]$ , la proporción de unos en la población, de esta manera  $\Pi$  representa la probabilidad de que un sujeto seleccionado aleatoriamente tome como valor de respuesta un 1, es decir, la probabilidad de un éxito. Esta probabilidad es el parámetro  $p$  de la distribución binomial y puede modificarse de acuerdo a los valores de la variable independiente. Para una sola variable explicativa, el modelo simple sería:

$$E[Y] = \Pi = \alpha + \beta X$$

esta expresión implica que  $\Pi$  puede tomar cualquier valor puesto que el rango en el que  $X$  puede tomar valores es de  $-\infty$  a  $+\infty$ .

Debido a que la variable de respuesta es dicotómica la media de ésta es una proporción y por lo tanto  $\Pi$  debe ser mayor igual que cero o menor igual que 1, por lo que la gráfica de  $\Pi$  más que lineal parecerá la gráfica de una distribución de una variable aleatoria es decir, la curva típica para este tipo de casos es en forma de S, entre 0 y 1, como se puede observar en la siguiente Gráfica.



GRAFICA 1.

Debido a esto, la ecuación de regresión debe estar formulada de tal manera que  $\pi$  se encuentre acotado entre 0 y 1, esta relación suele describirse con la fórmula[2]:

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X$$

la función  $\log\left(\frac{\pi}{1-\pi}\right)$  se llama transformación logística ó logit y es el modelo al que se le conoce como de regresión logística.

Es común usar logaritmos naturales en estos modelos, conforme  $\pi$  se incrementa de 0 a 1, el logit va de  $-\infty$  a  $\infty$ . La probabilidad  $\pi = \frac{1}{2}$  corresponde a un logit de 0 y los valores de  $\pi$  por arriba o por debajo de  $\frac{1}{2}$  pertenecen a un logit positivo o negativo respectivamente.

La relación entre  $\pi$  y  $X$  es, generalmente, no lineal, es decir, un cambio en  $X$  puede tener menos impacto cuando  $\pi$  está cerca de 0 ó 1, que cuando  $\pi$  esta cerca de la mitad de su rango.

Al despejar  $\pi$  de la ecuación logística se tiene:

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X$$

$$\frac{\pi}{1-\pi} = e^{\alpha+\beta X}$$

$$\pi = e^{\alpha+\beta X} - \pi e^{\alpha+\beta X}$$

$$\pi + \pi e^{\alpha+\beta X} = e^{\alpha+\beta X}$$

$$(1 + e^{\alpha+\beta X})\pi = e^{\alpha+\beta X}$$

$$\pi = \frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}} \dots\dots\dots ec.1$$

donde  $\pi$  se interpreta como la probabilidad de obtener un éxito.

Sea una muestra de  $n$  observaciones independientes de pares  $(x_i, y_i)$ , donde  $i = 1, 2, \dots, n$ , con  $y_i$  el valor de la variable dicotómica dependiente y  $x_i$  el valor de la variable independiente en la observación  $i$ . Para ajustar el modelo de regresión logística a los datos se requiere estimar los parámetros  $\alpha$  y  $\beta$  que son desconocidos. En lugar de utilizar el método de mínimos cuadrados se utilizan técnicas de máxima verosimilitud, para maximizar el valor de la función log-verosimil.

La probabilidad de que  $Y = 1$  dado el resultado  $X$  es  $\pi$ , como se definió anteriormente en *ec.1*, la probabilidad de que  $Y = 0$  es entonces  $1 - \pi$ , por lo tanto, la función de densidad de probabilidad de  $Y$  es:

$$f_Y(y) = \pi^y (1 - \pi)^{1-y}; \quad y = 0, 1$$

y la función de verosimilitud es:

$$\ell(y, \pi) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

y el logaritmo natural de esta función es:

$$\begin{aligned} L(y, \pi) &= \text{Ln} [\ell(y, \pi)] = \text{Ln} \left[ \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \right] \\ &= \sum_{i=1}^n \text{Ln}(\pi_i^{y_i}) + \text{Ln}(1 - \pi_i)^{1-y_i} \\ &= \sum_{i=1}^n \{y_i \text{Ln}[\pi_i] + (1 - y_i) \text{Ln}[1 - \pi_i]\} \end{aligned}$$

además

$$\pi_i = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}$$

considerando esto y haciendo álgebra se tiene que:

$$\begin{aligned} \text{Ln}[\ell(y; \pi_i)] &= \sum_{i=1}^n y_i \text{Ln} \left( \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \right) + (1 - y_i) \text{Ln} \left( 1 - \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \right) \\ &= \sum_{i=1}^n y_i \text{Ln} \left( \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \right) + (1 - y_i) \text{Ln} \left( \frac{1 + e^{\alpha + \beta x_i} - e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \right) \end{aligned}$$



$$\begin{aligned}
&= \sum_{i=1}^n y_i \operatorname{Ln} \left( \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \right) + (1 - y_i) \operatorname{Ln} \left( \frac{1}{1 + e^{\alpha + \beta x_i}} \right) \\
&= \sum_{i=1}^n \left[ y_i (\operatorname{Ln}(e^{\alpha + \beta x_i}) - \operatorname{Ln}(1 + e^{\alpha + \beta x_i})) + (1 - y_i) (\operatorname{Ln}(1) - \operatorname{Ln}(1 + e^{\alpha + \beta x_i})) \right] \\
&= \sum_{i=1}^n \left[ y_i \operatorname{Ln}(e^{\alpha + \beta x_i}) - y_i \operatorname{Ln}(1 + e^{\alpha + \beta x_i}) - \operatorname{Ln}(1 + e^{\alpha + \beta x_i}) + y_i \operatorname{Ln}(1 + e^{\alpha + \beta x_i}) \right] \\
&= \sum_{i=1}^n \left[ y_i \operatorname{Ln}(e^{\alpha + \beta x_i}) - \operatorname{Ln}(1 + e^{\alpha + \beta x_i}) \right] \\
&= \sum_{i=1}^n \left[ y_i (\alpha + \beta x_i) - \operatorname{Ln}(1 + e^{\alpha + \beta x_i}) \right]
\end{aligned}$$

para encontrar el valor de los estimadores se maximiza la última ecuación con respecto a  $\alpha$  y con respecto a  $\beta$ .

Así para  $\alpha$  se obtiene:

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \operatorname{Ln}(\ell(y; \pi)) &= \frac{\partial}{\partial \alpha} \sum_{i=1}^n \left[ y_i (\alpha + \beta x_i) - \operatorname{Ln}(1 + e^{\alpha + \beta x_i}) \right] \\
&= \sum_{i=1}^n \left[ y_i - \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \right]
\end{aligned}$$

$$= \sum_{i=1}^n [y_i - \pi_i] \dots\dots\dots ec.2$$

y con respecto a  $\beta$  se obtiene:

$$\begin{aligned} \frac{\partial}{\partial \beta} \text{Ln}(\ell(y; \pi_i)) &= \frac{\partial}{\partial \beta} \sum_{i=1}^n [y_i(\alpha + \beta x_i) - \text{Ln}(1 + e^{\alpha + \beta x_i})] \\ &= \sum_{i=1}^n \left[ y_i x_i - \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} x_i \right] \\ &= \sum_{i=1}^n x_i [y_i - \pi_i] \dots\dots\dots ec.3 \end{aligned}$$

para verificar que sea máximo se calcula la segunda derivada de ambas ecuaciones. Derivando nuevamente con respecto de  $\alpha$  se tiene:

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2} \text{Ln}(\ell(y; \pi_i)) &= \frac{\partial}{\partial \alpha} \sum_{i=1}^n [y_i - \pi_i] \\ &= \frac{\partial}{\partial \alpha} \sum_{i=1}^n y_i - \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \\ &= \sum_{i=1}^n \left[ - \frac{e^{\alpha + \beta x_i} (1 + e^{\alpha + \beta x_i}) - e^{\alpha + \beta x_i} (e^{\alpha + \beta x_i})}{(1 + e^{\alpha + \beta x_i})^2} \right] \end{aligned}$$

$$= - \sum_{i=1}^n \frac{e^{\alpha+\beta x_i} + (e^{\alpha+\beta x_i})^2 - (e^{\alpha+\beta x_i})^2}{(1 + e^{\alpha+\beta x_i})}$$

$$= - \sum_{i=1}^n \frac{e^{\alpha+\beta x_i}}{(1 + e^{\alpha+\beta x_i})^2}$$

como la función exponencial siempre es mayor que cero, se tiene que

$$\sum_{i=1}^n \frac{e^{\alpha+\beta x_i}}{(1 + e^{\alpha+\beta x_i})^2} > 0$$

y por lo tanto

$$- \sum_{i=1}^n \frac{e^{\alpha+\beta x_i}}{(1 + e^{\alpha+\beta x_i})^2} < 0$$

es decir

$$\sum_{i=1}^n [y_i - \pi_i]$$

es máximo.

Análogamente para la ec.3 se tiene:

$$\sum_{i=1}^n x_i [y_i - \pi_i]$$

calculando la segunda derivada con respecto a  $\beta$  se tiene

$$\begin{aligned} \frac{\partial^2}{\partial \beta^2} \text{Ln}(\ell(y_i; \pi_i)) &= \frac{\partial}{\partial \beta} \sum_{i=1}^n x_i y_i - x_i \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \\ &= - \sum_{i=1}^n x_i \frac{(e^{\alpha + \beta x_i}) x_i (1 + e^{\alpha + \beta x_i}) - (e^{\alpha + \beta x_i}) x_i (e^{\alpha + \beta x_i})}{(1 + e^{\alpha + \beta x_i})^2} \\ &= - \sum_{i=1}^n \frac{x_i^2 (e^{\alpha + \beta x_i}) (1 + e^{\alpha + \beta x_i}) - x_i^2 (e^{\alpha + \beta x_i})}{(1 + e^{\alpha + \beta x_i})} \\ &= - \sum_{i=1}^n \frac{x_i^2 (e^{\alpha + \beta x_i}) (1 + e^{\alpha + \beta x_i} - 1)}{(1 + e^{\alpha + \beta x_i})} \\ &= - \sum_{i=1}^n x_i^2 \frac{(e^{\alpha + \beta x_i})^2}{(1 + e^{\alpha + \beta x_i})^2} \end{aligned}$$

los sumandos siempre van a ser positivos, por lo tanto

$$- \sum_{i=1}^n x_i^2 \frac{(e^{\alpha + \beta x_i})^2}{(1 + e^{\alpha + \beta x_i})^2} < 0$$

y por lo tanto

$$\sum_{i=1}^n x_i [y_i - \pi_i]$$

es máximo.

A las *ec.2.* y *ec.3* se les llaman ecuaciones de verosimilitud. Para encontrar los valores de  $\alpha$  y  $\beta$  se iguala a cero y se trata de despejar  $\alpha$  y  $\beta$ :

$$\sum_{i=1}^n [y_i - \pi_i] = 0$$

$$\sum_{i=1}^n y_i - \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} = 0$$

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \dots\dots\dots ec.4$$

por otro lado

$$\sum_{i=1}^n x_i [y_i - \pi_i] = 0$$

$$\sum_{i=1}^n x_i y_i - x_i \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} = 0$$

$$\sum_{i=1}^n x_i y_i = \sum_{i=1}^n x_i \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \dots\dots\dots ec.5$$

No se puede despejar a  $\alpha$  o  $\beta$  de *ec.4* o *ec.5*. A diferencia de los estimadores de los parámetros de mínimos cuadrados que pueden obtenerse directamente, debido a que los parámetros son lineales, en la regresión logística los

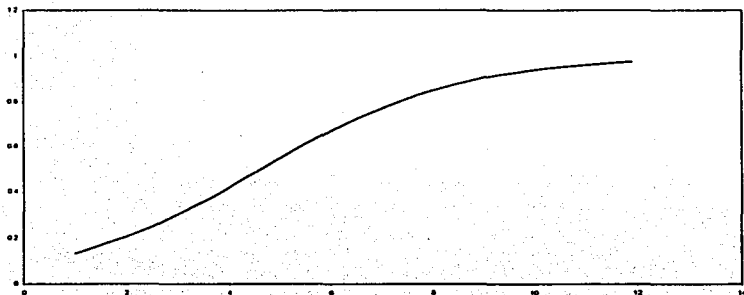
parámetros no son lineales, por lo que las ecuaciones se resuelven suponiendo una solución tentativa, revisando si puede ser mejorada y repitiendo el proceso hasta que el cambio en la función log-verosímil de un paso del proceso a otro sea despreciable, es decir, se hace a través de un proceso iterativo. Una vez que la función no cambia más, se dice que la solución converge[15].

Si  $\beta = 0$ , coeficiente de la ecuación  $\alpha + \beta X$  del modelo logístico, el logit (y por esto  $\pi$ ) no cambia conforme  $X$  cambia, por lo que la curva se aplana quedando una línea recta horizontal, es decir

$$\frac{\exp(\alpha + \beta X)}{1 + \exp(\alpha + \beta X)}$$

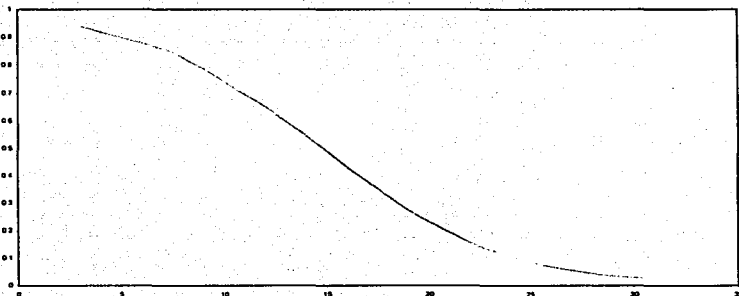
$$= \frac{e^{\alpha}}{1 + e^{\alpha}}$$

el cual evidentemente permanece constante para cualquier  $\alpha$  y la probabilidad  $\pi$  es igual para toda las  $X$ . El parámetro  $\beta$  determina la razón de incremento y decremento de la curva en forma de S; para  $\beta$  positiva  $\pi$  crece conforme  $X$  crece, como se puede ver en la Gráfica 2.



GRAFICA 2. Ejemplo de la curva de regresión logística cuando  $\beta > 0$

Para  $\beta$  negativa  $\pi$  decrece conforme  $X$  crece, es decir, la probabilidad de una respuesta 1 tiende a cero para valores grandes de  $X$ , como se muestra en la Gráfica 3.



GRAFICA 3. Ejemplo la curva logística cuando  $\beta < 0$

Algunas veces es útil saber el valor de  $X$  al cual  $\pi = \frac{1}{2}$ . Sustituyendo en la ecuación logística:

$$\log \left( \frac{\pi}{1-\pi} \right) = \log \left( \frac{\frac{1}{2}}{1-\frac{1}{2}} \right)$$

$$= \log \left( \frac{\frac{1}{2}}{\frac{1}{2}} \right)$$

$$= \log (1) = 0$$

por lo tanto

$$0 = \alpha + \beta X$$

$$\alpha = -\beta X$$

$$X = -\frac{\alpha}{\beta}$$

este valor de  $X$  es frecuentemente llamado nivel mediano efectivo, se denomina  $EML_{50}$  y representa el nivel en el cual cada resultado tiene un 50% de oportunidad. Hay que hacer notar que  $\beta$  no puede ser interpretada como una simple pendiente porque la razón con la que la curva sube o baja cambia de acuerdo al valor de  $X$ . La razón de cambio más grande está en el punto  $\pi = \frac{1}{2}$  y se puede conocer esta razón calculando la pendiente de una recta tangente trazada en ese punto[2], para esto se calcula la derivada:



$$\begin{aligned}
 \frac{d}{dx} \pi &= \frac{d}{dx} \left( \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} \right) \\
 &= \frac{e^{\alpha+\beta x} \cdot \beta \cdot (1+e^{\alpha+\beta x}) - e^{\alpha+\beta x} \cdot \beta \cdot e^{\alpha+\beta x}}{(1+e^{\alpha+\beta x})^2} \\
 &= \frac{e^{\alpha+\beta x} \cdot \beta \cdot (1+e^{\alpha+\beta x} - e^{\alpha+\beta x})}{(1+e^{\alpha+\beta x})^2} \\
 &= \beta \cdot \frac{e^{\alpha+\beta x}}{(1+e^{\alpha+\beta x})^2} \\
 &= \beta \cdot \pi(1 - \pi)
 \end{aligned}$$

sustituyendo el valor de  $X = -\frac{\alpha}{\beta}$ , que es el valor de  $X$  cuando  $\pi = 0.5$ , se tiene:

$$\begin{aligned}
 &\beta \frac{e^{\alpha+\beta(-\frac{\alpha}{\beta})}}{\left(1+e^{\alpha+\beta(-\frac{\alpha}{\beta})}\right)^2} \\
 &= \beta \frac{e^0}{(1+e^0)^2} = \frac{\beta}{4}
 \end{aligned}$$

es decir, una recta tangente trazada en el punto  $\pi = \frac{1}{2}$  tiene una pendiente de  $\frac{\beta}{4}$ , con  $\beta$  el coeficiente de regresión logística.

Una manera importante de interpretar el coeficiente de regresión logística  $\beta$  es como un efecto sobre el momio, el cual es una medida de asociación que está definida por:

$$\text{Momio} = \frac{\pi}{1-\pi}$$

Con  $\pi$  la probabilidad de éxito. El valor del momio siempre es positivo, cuando es mayor que 1 indica que un éxito es más probable que un fracaso. [1]

La razón de momios se define como:

$$\theta = \frac{\text{momio}_i}{\text{momio}_j} = \frac{\frac{\pi_i}{1-\pi_i}}{\frac{\pi_j}{1-\pi_j}}$$

El cociente de momios también es positivo siempre, cuando  $1 < \theta < \infty$  es más probable que ocurra un éxito que un fracaso en el caso del subíndice  $i$  que para el subíndice  $j$ , es decir,  $\pi_i > \pi_j$ . Cuando  $0 < \theta < 1$ ,  $\pi_i < \pi_j$ , la probabilidad de que ocurra un éxito, a diferencia de un fracaso, para el subíndice  $i$  es menor que para el subíndice  $j$ . [1]

Cuando el momio es igual a 1,  $\theta = 1$ , la probabilidad de éxito es la misma,  $\pi_i = \pi_j$ .

Si se toma el antilogaritmo de ambos lados de la ecuación logística

$$\log \left( \frac{\pi}{1-\pi} \right) = \alpha + \beta X$$

se obtiene

$$\left( \frac{\pi}{1-\pi} \right) = e^{\alpha} (e^{\beta})^X$$

esta relación exponencial implica que cada unidad de incremento en  $X$  produce un efecto multiplicativo de  $e^{\beta}$  en el momio, es decir, el momio en la variable  $X + 1$  es igual a aquel en  $X$  multiplicado por  $e^{\beta}$ . Cuando  $\beta = 0$ ,  $e^{\beta} = 1$ , y el momio no cambia conforme  $X$  cambia.

Para determinar la significancia del coeficiente estimado se hace una prueba de hipótesis:

$$H_0 : \beta = 0$$

vs.

$$H_1 : \beta \neq 0$$

Para realizar la prueba se utiliza el estadístico de Wald, que se obtiene comparando el estimador de máxima verosimilitud del parámetro de pendiente  $\beta$  con su estimado de desviación estándar.

$$W_i^2 = \left( \frac{\hat{\beta}_i}{\text{d.e.}(\hat{\beta}_i)} \right)^2$$

que asintóticamente se distribuye ji-cuadrada con 1 grado de libertad, o también

$$W_i = \frac{\hat{\beta}_i}{\text{d.e.}(\hat{\beta}_i)}$$

que en este caso se distribuye normal estándar. En ambos casos la hipótesis nula se rechaza para valores del estadístico grandes, es decir, se rechaza  $H_0$  si:

$$W_i > \chi^2_{1,\alpha}$$

con  $\alpha$  el nivel de significancia elegido.

La desventaja del estadístico de Wald es que para  $\beta$  grandes, el error estándar estimado se infla resultando en falla al rechazar la hipótesis nula cuando ésta es falsa.

---

## 2.2. MODELO MÚLTIPLE

---

Un modelo de regresión logística puede ser utilizado con varias variables independientes las cuales pueden ser cuantitativas, cualitativas o ambas.

Sean  $Y_1, Y_2, \dots, Y_n$ , resultados de conteos binomiales, las variables dependientes (o componente aleatorio). Asociada a cada respuesta  $Y_i$  están las variables independientes (componente sistemático)  $x_{i1}, x_{i2}, \dots, x_{ip}$ , el logit de la probabilidad  $\pi$  para una respuesta de  $Y = 1$  ó un éxito es:

$$\begin{aligned} \text{logit}(\pi_i) &= \log\left(\frac{\pi_i}{1-\pi_i}\right) \\ &= \alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}; \quad i = 1, \dots, n \end{aligned}$$

es decir:

$$\begin{aligned} \text{logit } \pi_1 &= \log\left(\frac{\pi_1}{1-\pi_1}\right) \\ &= \alpha + \beta_1 x_{11} + \dots + \beta_p x_{1p} \end{aligned}$$

$$\begin{aligned} \text{logit } \pi_2 &= \log\left(\frac{\pi_2}{1-\pi_2}\right) \\ &= \alpha + \beta_1 x_{21} + \dots + \beta_p x_{2p} \end{aligned}$$

$$\begin{aligned} \text{logit } \pi_n &= \log\left(\frac{\pi_n}{1-\pi_n}\right) \\ &= \alpha + \beta_1 x_{n1} + \dots + \beta_p x_{np} \end{aligned}$$

que en forma matricial se expresa como:

$$\begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \mathbf{X}\beta$$

por lo tanto

$$\text{logit}(\pi) = \mathbf{X}\beta$$

con  $\text{logit}(\pi)$  vector de  $n \times 1$  que contiene los valores estimados por el modelo,  $\mathbf{X}$  matriz diseño o matriz modelo de  $n \times p+1$ , que contiene las variables explicativas  $x_1, \dots, x_p$  para las  $n$  observaciones y  $\beta$  vector de parámetros de  $(p+1) \times 1$ . A las variables explicativas numéricas se les llama *covariables* y cada una se representa en una sola columna de  $\mathbf{X}$ . Las variables explicativas que son categóricas son llamadas *factores*, si hay  $k$  niveles de factores entonces habrá  $k-1$  columnas de 1 y 0 en  $\mathbf{X}$  para ese factor.

El método de estimación de  $\beta' = (\alpha, \beta_1, \beta_2, \dots, \beta_p)$  es también el de máxima verosimilitud. El valor de  $\pi_i$  se obtiene despejando de la ecuación:

$$\begin{aligned} \text{logit}(\pi_i) &= \log\left(\frac{\pi_i}{1-\pi_i}\right) \\ &= \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \end{aligned}$$

$$\frac{\pi_i}{1-\pi_i} = e^{\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}$$

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

$$\pi_i + \pi_i e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}} = e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}$$

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

La función de verosimilitud es en este caso:

$$\ell(y; \pi) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}; \quad y = 0, 1$$

Aplicando logaritmo natural a la ecuación:

$$\text{Ln}(\ell(y; \pi)) = \sum_{i=1}^n y_i \text{Ln}(\pi_i) + (1 - y_i) \text{Ln}(1 - \pi_i)$$

Sustituyendo el valor de  $\pi_i$

$$= \sum_{i=1}^n y_i \text{Ln} \left( \frac{e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}} \right) + (1 - y_i) \text{Ln} \left( \frac{1}{1 + e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}} \right)$$

Calculando las derivadas parciales para cada parámetro:

$$\frac{\partial}{\partial \alpha} \text{Ln}(\ell(y; \pi)) = \frac{\partial}{\partial \alpha} \sum_{i=1}^n \left[ y_i (\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) - \text{Ln}(1 + e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}) \right]$$

$$= \sum_{i=1}^n y_i - \frac{e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}$$

$$= \sum_{i=1}^n y_i - \pi_i = 0$$

Obteniendo la derivada parcial con respecto a  $\beta_j$

$$\frac{\partial}{\partial \beta_j} \text{Ln}(\ell(y; \pi)) = \frac{\partial}{\partial \beta_j} \sum_{i=1}^n \left[ y_i \text{Ln} \left( \frac{e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}} \right) + (1 - y_i) \text{Ln} \left( \frac{1}{1 + e^{\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}} \right) \right]$$



$$\begin{aligned}
&= \sum_{i=1}^n \left[ y_i \left( x_{ij} - \frac{e^{\alpha + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}} x_{ij} \right) - (1 - y_i) \left( \frac{e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}} x_{ij} \right) \right] \\
&= \\
&\sum_{i=1}^n \left[ x_{ij} \left( y_i - \frac{e^{\alpha + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}} y_i \right) - \frac{e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}} + \left( \frac{e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}} y_i \right) \right] \\
&= \sum_{i=1}^n \left[ x_{ij} \left( y_i - \frac{e^{\alpha + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}}{1 + e^{\alpha + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}}} \right) \right]
\end{aligned}$$

$$\sum_{i=1}^n x_{ij} (y_i - \pi_i) = 0; \quad j = 1, 2, \dots, p$$

Habr  entonces  $p+1$  ecuaciones de verosimilitud, las cuales se obtienen derivando la funci n de log-verosimilitud respecto a cada uno de los  $p+1$  coeficientes e igualando a cero. Nuevamente se requiere de paquetes de computaci n especiales para encontrar la soluci n de las ecuaciones. En el caso de esta tesis se utiliz  el paquete SPSS v.8 para calcular los estimadores de los par metros.

Los estimadores de varianza y covarianza de los coeficientes estimados se obtienen a partir de segundas derivadas parciales del logaritmo de la función de verosimilitud. Estos tienen la forma general:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i)$$

y

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_u} = - \sum_{i=1}^n x_{ij} x_{iu} \pi_i (1 - \pi_i)$$

para  $j, u = 0, 1, 2, \dots, p$  y con  $\pi_i$  como se definió anteriormente. [12] La matriz de  $p+1 \times p+1$  de los términos de las ecuaciones anteriores se denota  $I(\beta)$  y se llama matriz información. Las varianzas y covarianzas de los coeficientes estimados se obtienen con la inversa de la esta matriz que se denota  $\Sigma(\beta) = I^{-1}(\beta)$ . Se utiliza la notación  $\sigma^2(\beta_j)$ , la varianza de  $\beta_j$ , para denotar al elemento  $j$ -ésimo de la diagonal y  $\sigma^2(\beta_j, \beta_u)$ , covarianza de  $\beta_j$  y de  $\beta_u$ , para denotar cualquier término fuera de la diagonal.

Los estimadores de la varianza y covarianza, denotados por  $\hat{\Sigma}(\beta)$ , se obtienen evaluando  $\Sigma(\beta)$  en  $\hat{\beta}$ . Para denotar los valores de esta matriz se utilizan  $\hat{\sigma}^2(\beta_j)$  y  $\hat{\sigma}^2(\beta_j, \beta_u)$ ,  $j, u = 0, 1, 2, \dots, p$  para denotar los valores en esta matriz.

Una forma de generar la matriz de información es  $I(\beta) = X'VX$  donde  $X$  es una matriz de  $n \times p+1$  que contiene los datos de cada individuo y  $V$  es una matriz diagonal de  $n \times n$  con elementos comunes  $\pi_i(1-\pi_i)$ :

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

y

$$V = \begin{pmatrix} \pi_1(1-\pi_1) & 0 & \cdots & 0 \\ 0 & \pi_2(1-\pi_2) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \pi_n(1-\pi_n) \end{pmatrix}$$

Los parámetros  $\beta_i$  se refieren al efecto de  $x_i$  en el momio controlando las demás  $x$ ; por ejemplo,  $\exp(\beta_i)$  es el efecto multiplicativo en el momio al incremento en una unidad de  $x_i$ , cuando las otras variables están fijas[1].

El estadístico de Wald también puede calcularse en el caso multivariado, se obtiene haciendo los siguientes cálculos[12]:

$$\begin{aligned} W &= \beta'[\Sigma(\beta)]^{-1}\beta \\ &= \beta'(X'VX)\beta \end{aligned}$$

que se distribuye ji-cuadrada con  $p+1$  grados de libertad, el número de parámetros que se están probando, bajo la hipótesis de que cada uno de los  $p$  parámetros son iguales a cero,

$$H_0 : \alpha = \beta_1 = \beta_2 = \dots = \beta_p = 0.$$

vs.

$$H_1 : \beta_j \neq 0, \text{ para alguna } j$$

La regla de decisión es como en el caso simple, rechazar  $H_0$  si

$$W > \chi^2_{(p+1), \alpha}$$

con  $\alpha$  el nivel de significancia elegido.

---

## 2.3. AJUSTE DEL MODELO

---

### 2.3.1. BONDAD DE AJUSTE

---

Al ajustar el modelo de regresión logística se asume que las variables que se encuentran en el modelo son aquellas que aportan información significativa acerca de la variabilidad de la respuesta, en este caso, la probabilidad de éxito. Como siguiente paso se averigua que tan efectivo es el modelo para describir a la variable de respuesta, lo que se llama bondad de ajuste.

Para decidir si el modelo ajusta o no se toman en cuenta ciertos criterios como que las medidas de la distancia entre valores observados y estimados sean pequeñas y que no exista algún tipo de patrón entre los valores observados y estimados.

Cuando se ajusta el modelo de regresión logística se supone que los valores de la variable dependiente son tales que cada uno es único en su configuración de variables de predicción, es decir, el número de patrones de covariables (el número de combinaciones distintas de variables independientes) es igual al número de observaciones de la variable de respuesta.

Sea  $X$  la matriz de diseño de un modelo ajustado con  $p$  variables independientes:

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

en caso de que  $K < n$ ,  $K$  el número de patrones de covariables, se tendrá entonces dos o más renglones  $(1, x_{i1}, x_{i2}, \dots, x_{ip})$  con las mismas entradas.

Se define  $m_k$  como el número de respuestas que tienen el mismo patrón de covariables,  $k = 1, 2, \dots, K$ , entonces  $\sum_{k=1}^K m_k = n$ ; sea ahora  $w_k$  el número total de éxitos que se observan en cada patrón de covariables,  $w_k$  se puede estimar como:

$$\hat{w}_k = m_k \hat{\pi}_k$$

donde  $\hat{\pi}_k$  es la probabilidad de éxito estimada por el modelo de regresión logística.

Para probar la hipótesis nula de que el modelo ajusta se comparan el número total de éxitos observados con los estimados por el modelo para cada patrón de covariables. La hipótesis nula que se prueba es:

$H_0$  : el modelo ajusta de manera adecuada los datos

o

$$H_0 : \text{logit}(\pi) = X\beta$$

vs.

$H_1$  : el modelo no ajusta de manera adecuada los datos

o

$$H_1 : \text{logit}(\pi) \neq X\beta$$

Para realizar esta prueba de hipótesis se pueden utilizar dos estadísticos diferentes, el de Pearson o el de Devianza.

El estadístico  $X^2$  de Pearson se define como:

$$X^2 = \sum_{k=1}^K \frac{(w_k - m_k \hat{\pi}_k)^2}{m_k \hat{\pi}_k (1 - \hat{\pi}_k)}$$

y el de Devianza como:

$$G^2 = -2 \left\{ \sum_{k=1}^K w_k \text{Ln} \left( \frac{w_k}{m_k \hat{\pi}_k} \right) + (m_k - w_k) \text{Ln} \left( \frac{m_k - w_k}{m_k (1 - \hat{\pi}_k)} \right) \right\}$$

Ambos estadísticos se distribuyen ji-cuadrada con  $K - (p + 1)$  grados de libertad, es decir, el número de patrones de covariables menos el número de parámetros estimados. Valores grandes de  $X^2$  y  $G^2$  proveen evidencia de falta de ajuste, por lo que se rechaza  $H_0$  si  $X^2$ ,  $G^2$  son mayores que  $\chi^2_{K-(p+1), \alpha}$ , con  $\alpha$  el nivel de significancia dado.

La distribución de los estadísticos de prueba de bondad de ajuste se sostiene no sólo del hecho de que  $n$  (el número de observaciones de la variable dependiente) es grande, sino también de que  $w_k$  (el número de éxitos en cada patrón de covariables) es grande. En el caso particular de que alguna de las variables de predicción sea continua, al aumentar  $n$  también crece el número de patrones, pero los  $w_k$  permanecen del mismo tamaño, es decir, son pequeños, por lo que los estadísticos  $X^2$  y  $G^2$  no se distribuyen de manera aproximada como una normal.

La manera en que pueden utilizarse estos estadísticos de bondad de ajuste en caso de datos continuos es agrupando los valores observados y estimados, formando categorías (patrones de covariables) en los que el total de éxitos ajustado es la suma de las probabilidades predichas por el modelo, es decir:

$$\hat{w}_k = \sum_{t=1}^{m_k} \hat{\pi}_t$$

el número total de fracasos estimados es entonces:

$$m_k - \hat{w}_k = \sum_{i=1}^{m_k} (1 - \hat{\pi}_i)$$

El número de éxitos observados y estimados por patrón es mayor que cuando no se agrupan los datos y por lo tanto los estadísticos  $X^2$  y  $G^2$  para probar el ajuste del modelo pueden utilizarse.

Se pueden comparar dos modelos distintos utilizando el estadístico  $G^2$  calculando la diferencia de Devianzas del modelo de la hipótesis nula y el modelo de la hipótesis alternativa:

$H_0$  : modelo 1

vs

$H_1$  : modelo 2

el estadístico de prueba es:

$$G^2(H_0 | H_1) = G^2(H_0) - G^2(H_1) = \text{Devianza(modelo1)} - \text{Devianza(modelo2)}$$

el cual se distribuye también ji-cuadrada con

$$g.l. = g.l.(modelo 1) - g.l.(modelo2)$$



un valor grande de  $G^2(H_0 | H_1)$  indica que el modelo 1 ajusta pobremente en comparación al modelo 2, por lo tanto se rechaza la hipótesis nula si  $G^2(H_0 | H_1) > \chi^2_{g.l., \alpha}$ , con  $\alpha$  el nivel de significancia dado.

Este tipo de prueba de hipótesis se utiliza también para verificar la significancia de una cierta variable independiente, en este caso, se iguala a cero el parámetro de la variable que se quiere probar en la hipótesis nula:

$$H_0 : \beta_j = 0; \quad j = 1, 2, \dots, p$$

vs.

$$H_1 : \beta_j \neq 0; \quad \text{para alguna } j$$

el estadístico de prueba es

$$G^2(H_0 | H_1) = \text{Devianza}(\text{modelo con } \beta_j) - \text{Devianza}(\text{modelo sin } \beta_j)$$

los grados de libertad son entonces

$$g.l. = g.l.(\text{modelo con } \beta_j) - g.l.(\text{modelo sin } \beta_j)$$

Nuevamente se rechaza  $H_0$  si  $G^2(H_0 | H_1) > \chi^2_{g.l., \alpha}$ , con  $\alpha$  el nivel de significancia dado.

### 2.3.2. RESIDUOS

---

Los estadísticos  $X^2$  y  $G^2$  ofrecen una medida general de que tan bien se ajusta el modelo a los datos, sin embargo se puede obtener información extra

si se comparan los valores observados y estimados de manera individual. Para la observación k-ésima, la diferencia entre valores observados y estimados es:

$$w_k - \hat{w}_k$$

con  $\hat{w}_k$  el número de éxitos estimado por el modelo en el patrón k. El residuo de Pearson es una estandarización de esta diferencia y está definido por:

$$\text{residuo de Pearson} = r_k = \frac{w_k - m_k \hat{\pi}_k}{\sqrt{m_k \hat{\pi}_k (1 - \hat{\pi}_k)}}$$

El estadístico  $X^2$  es simplemente la suma de residuos de Pearson elevados al cuadrado. Cuando  $m_k$  es grande, el residuo se distribuye aproximadamente como una normal con media 0 y varianza 1.

Una alternativa a los residuos de Pearson son los residuos de Devianza, que están definidos por

$$d_k = \pm \sqrt{2 \left[ w_k \text{Ln} \left( \frac{w_k}{m_k \hat{\pi}_k} \right) + (m_k - w_k) \text{Ln} \left( \frac{m_k - w_k}{m_k (1 - \hat{\pi}_k)} \right) \right]}$$

el signo del residuo es el mismo de la diferencia  $w_k - m_k \hat{\pi}_k$ .

El residuo de Devianza representa la contribución de cada caso al estadístico  $G^2$  y cuando  $m_k$  es grande tiene una distribución aproximada normal con media 0 y varianza 1.

Puesto que ambos residuos se distribuyen aproximadamente normal estándar se puede considerar que si para  $k$

$$|r_k| > 2 \text{ o } |d_k| > 2$$

entonces es posible que haya falta de ajuste, sin embargo, cuando existen datos que se distribuyen como una normal, el 95% de los casos deberían tener valores entre  $-2$  y  $+2$  y el 99% entre  $-2.5$  y  $+2.5$ , es decir, se puede esperar que aproximadamente el 5% de los residuos se encuentren fuera del rango de  $-2$ ,  $+2$  y que el 1% se exceda del  $-2.5$ ,  $+2.5$  aunque valores extremadamente lejos de estos intervalos generalmente son indicadores de que el modelo ajusta pobremente a un caso en particular y sugiere que existe algo extraño en la observación. Sin embargo, encontrar residuos superiores a 2 ó 3 en valor absoluto, no significa necesariamente que hay algo mal en el modelo.

### 2.3.3. MEDIDA DE ASOCIACIÓN MÚLTIPLE ENTRE LAS VARIABLES INDEPENDIENTES Y LA VARIABLE DEPENDIENTE.

Para medir el grado de asociación entre las variables independientes y la dependiente se ha propuesto a  $R^2_L$  como un análogo al coeficiente de determinación ( $R^2$ ) de regresión lineal. Éste se define como

$$R^2_L = \frac{G^2}{D_1} = \frac{G^2}{G^2 + D_2}$$

con  $D_1$  la Devianza del modelo que contiene únicamente al término de intersección y representa la variación total de las observaciones con respecto a la media sin considerar las variables de predicción.  $G^2(H_0 | H_1)$ , el estadístico que se utiliza para comparar la Devianza de dos modelos distintos, es la variabilidad debida al efecto de las variables independientes sobre la de respuesta. Se debe recordar que:

$$\begin{aligned} G^2(H_0 | H_1) &= G^2(H_0) - G^2(H_1) = \text{Devianza}(\text{modelo 1}) - \text{Devianza}(\text{modelo 2}) \\ &= D_1 - D_2 \end{aligned}$$

con  $D_1$  la Devianza del modelo 1 y  $D_2$  la Devianza del modelo 2. Entonces  $D_2$  representa la variación no explicada dadas las variables de predicción. El estadístico  $R^2_L$  es, entonces, la reducción de la variación total de la probabilidad de la variable dependiente dada la inclusión de las independientes.  $R^2_L$  puede definirse también de la siguiente manera:

$$\begin{aligned} R^2_L &= \frac{G^2}{D_1} = \frac{D_1 - D_2}{D_1} \\ &= 1 - \frac{D_2}{D_1} \end{aligned}$$

El valor de  $R^2_L$  se encuentra entre 0 y 1:

$$0 \leq R^2_L \leq 1$$

$R^2_L = 0$  para un modelo en la cual  $G^2 = 0$ ,  $D_1 = D_2$  y las variables independientes son insignificantes para la predicción de la variable Y;  $R^2_L = 1$  para  $D_2 = 0$ , es decir no existe variabilidad que no sea explicada por el modelo, y éste predice a la variable dependiente con perfecta exactitud. Cabe mencionar que  $R^2_L$  no mide la validez del modelo de regresión propuesto, todo lo que puede medir es cuánto se explica de la variación total mediante la ecuación de regresión estimada.

Añadir más variables independientes incrementa  $R^2_L$ , debido a esto se sugiere utilizar una medida modificada que se ajusta al número de variables independientes en el modelo:

$$R^2_{LA} = \frac{G_M - 2p}{D_0}$$

donde  $p$  es el número de variables independientes en el modelo.

Puede utilizarse  $R^2$  (el coeficiente de determinación para el modelo de regresión lineal) como una medida complementaria a la asociación entre las variables predictoras y la de respuesta;  $R^2$  permite la comparación directa de modelos de regresión logística con modelos de análisis de varianza y análisis discriminante cuando la predicción del valor observado, en lugar de la probabilidad de la observación, es lo que interesa. Permite así mismo la comparación el ajuste de distintos modelos al mismo conjunto de datos.

# CAPÍTULO III

---

## 3.1. RESULTADOS Y DISCUSIÓN

---

Contemplando todo lo expuesto anteriormente, el método que se utilizará para hacer el análisis es el de regresión logística, la ecuación a ajustar para estimar la probabilidad de parto múltiple es:

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1\text{NPG}(1) + \beta_2\text{NPG}(2) + \beta_3\text{PG}(1) + \beta_4\text{PG}(2) + \beta_5\text{IP} + \beta_6\text{TPA}$$

donde

- **NPG** = Número de Parto por Grupo. Las ovejas de la población se dividieron en tres grupos respecto al número de parto: **NPG(1)** para las que se encuentra entre el 2° y el 5°; **NPG(2)** para las que se encuentran entre el 6° y el 10°; y **NPG(3)** para las ovejas entre el 11<sup>avo</sup> y el 15<sup>avo</sup>. Esta última se definió como categoría base, por lo que no aparece en la ecuación.
- **PG** = Predisposición Genética. Las ovejas se clasificaron en tres grupos de acuerdo al criterio de selección utilizado en el rancho permitiendo identificar el tipo de nacimiento de la madre: **PG(1)** para ovejas que provienen únicamente de parto simple; **PG(2)** para animales que nacieron en parto simple ó múltiple y **PG(3)** para las que provienen solamente de parto múltiple. Esta última

clasificación se utilizó como la categoría base y tampoco aparece en la ecuación.

- IP = Intervalo entre Parto. El tiempo que pasa entre parto y parto de una oveja. Es variable continua.
- TPA = Tipo de Parto Anterior. Toma el valor de 1 cuando la oveja parió anteriormente más de un cordero y 0 cuando sólo tuvo uno.

$\alpha$  el término constante

$\beta_1$  efecto entre el NPG(1) y NPG(3)

$\beta_2$  efecto entre el NPG(2) y NPG(3)

$\beta_3$  efecto entre PG(1) y PG(3)

$\beta_4$  efecto entre PG(2) y PG(3)

$\beta_5$  efecto del intervalo de parto

$\beta_6$  efecto del tipo de parto anterior

El análisis estadístico se hizo con el paquete SPSS v.8.

La distribución de los animales por los diferentes factores se muestran en los Cuadros 4 y 5:

		NPG		
		1	2	3
PG	1 TPA=0	140	0	0
	TPA=1	10	0	0
	2 TPA=0	186	34	1
	TPA=1	114	77	0
	3 TPA=0	142	42	14
	TPA=1	139	46	18

Cuadro 4. Distribución de los factores incluidos en el modelo para TP = 0

		NPG			
		1	2	3	
IPG	1	TPA=0	41	0	0
		TPA=1	3	0	0
	2	TPA=0	163	61	1
		TPA=1	201	114	2
	3	TPA=0	153	54	25
		TPA=1	284	150	35

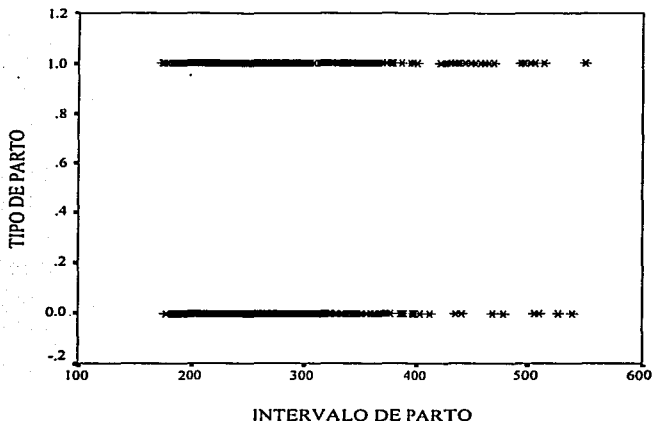
Cuadro 5. Distribución de los factores incluidos en el modelo para TP = 1

En el caso de Intervalo de Parto, que es una variable continua, se obtuvo una media de 245 días con una desviación estándar de 47 días.

La Gráfica 3 muestra la dispersión entre Tipo de Parto, variable de respuesta, e IP, una de las variables explicativas. Para Tipo de Parto existen sólo dos posibles valores 0, 1, por lo que se ven dos rectas horizontales paralelas.

En la gráfica referida, para TP = 1 se observa mayor dispersión que en TP = 0. En ambos casos las observaciones comienzan a dispersarse a partir de un valor de Intervalo entre Partos mayor a 400 días.





**GRAFICA 4. Dispersión entre TP y IP**

En el modelo que incluye todas las variables, Número de Parto (NP), Tipo de Parto Anterior (TPA), Intervalo entre Partos (IP), Tipo de Nacimiento de la Madre (PG); el Número de Parto y el Intervalo de Parto no resultaron significativas con un p-value de 0.0577 y 0.0845 respectivamente, aunque IP se mantendrá dentro del modelo para facilitar la interpretación.

Redefiniendo el modelo de regresión logística que estima la probabilidad de parto múltiple queda de la siguiente manera:

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 PG(1) + \beta_2 PG(2) + \beta_3 IP + \beta_4 TPA$$

con  $\alpha$  el término constante

### CAPÍTULO III

$\beta_1$  efecto entre PG(1) y PG(3)

$\beta_2$  efecto entre PG(2) y PG(3)

$\beta_3$  efecto del intervalo de parto

$\beta_4$  efecto del tipo de parto anterior

Los estimadores de los parámetros se muestran en el siguiente cuadro:

VARIABLE	B	D.E.	G.L.	SIG	EXP(B)
TPA(1)	0.5912	.0914	1	.0000	1.8062
IP	-0.0018	.0010	1	.0572	.9982
PG			2	.0000	
PG(1)	-1.4512	.1894	1	.0000	.2343
PG(2)	-0.2509	.0917	1	.0062	.7781
CONSTANTE	0.6554	.2445	1	.0073	

Cuadro 6. Estimadores de los parámetros de las variables utilizadas en el modelo

La ecuación final queda definida entonces como sigue:

$$\log\left(\frac{\pi}{1-\pi}\right) = .6554 + .5912 \text{ TPA} - .0018 \text{ IP} - 1.4512 \text{ PG}(1) - .2509 \text{ PG}(2)$$

En el Cuadro 6 la última columna señalada como Exp(B) muestra las razones de momios para cada coeficiente, éstos indican el número por el cual se tendría que multiplicar el momio de parto múltiple por cada incremento en una unidad de la variable independiente. En este caso se tiene que cuando el tipo de parto anterior es múltiple, el momio se incrementa en 80.62%, cuando se tienen registros de PG(1) o PG(2), los momios decrecen en

### CAPÍTULO III

76.57% y 22.19% respectivamente. Es importante aclarar que la razón de momios no es una medida aparte, es la misma información presentada en distinta manera.

Las predicciones para casos individuales se obtienen reemplazando las variables en la ecuación con sus casos específicos, por ejemplo si se supone un valor dado para  $IP = 243$ , se pueden estimar los valores de la probabilidad de parto simple ó múltiple para cada una de las combinaciones de los factores en base a la ecuación final:

1. Para  $TPA = 0$ ,  $PG(1) = 1$ ,  $PG(2) = 0$

$$\text{el logit}(TP) = .6554 - .0018(243) - 1.4512 = -1.2332$$

que corresponde a una probabilidad de parto múltiple de

$$\frac{e^{-1.2332}}{1 + e^{-1.2332}} = 0.2258$$

y el valor del momio es:

$$\frac{\pi}{1 - \pi} = 0.2913$$

2. para  $TPA = 1$ ,  $PG(1) = 1$ ,  $PG(2) = 0$

$$\text{logit}(TP) = .6554 + .5912 - .4374 - 1.4512 = -0.6420$$

$$\text{la probabilidad es } \frac{e^{-.6420}}{1 + e^{-.6420}} = 0.3454$$

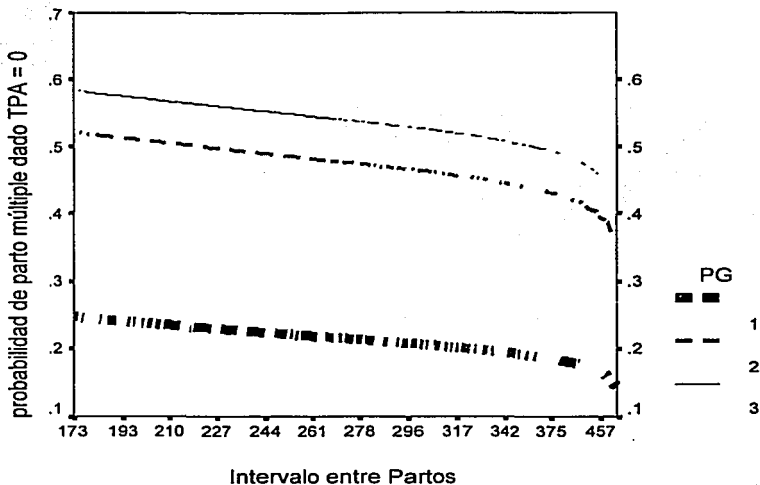
$$\text{y el valor del momio es } \frac{\pi}{1 - \pi} = .5262$$

Haciendo las mismas operaciones para cada una de las categorías de las variables que permanecieron en la ecuación, se obtiene el Cuadro 7 con el valor del logit, la probabilidad y el momio para Tipo de Parto múltiple:

		logit	probabilidad	momio
PG(1)	TPA(0)	-1.2332	0.2258	0.2913
	TPA(1)	-0.6420	0.3454	0.5662
PG(2)	TPA(0)	-0.0329	0.4917	0.9676
	TPA(1)	0.5583	0.6360	1.7476
PG(3)	TPA(0)	0.2180	0.5542	1.2435
	TPA(1)	0.8092	0.6919	2.2461

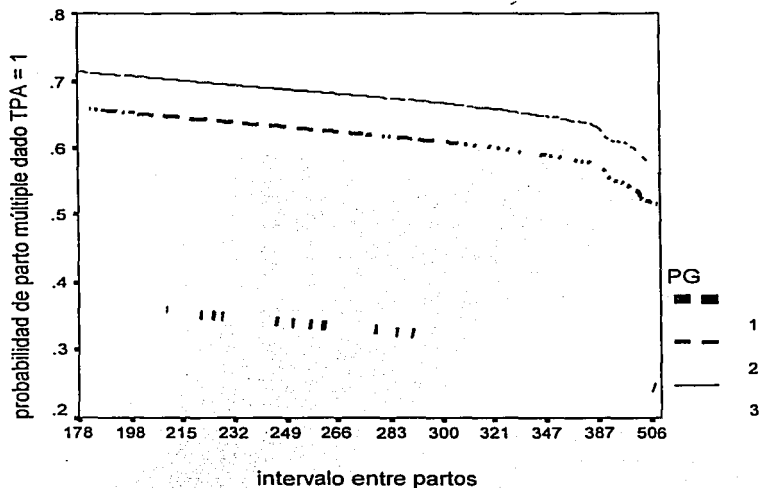
Cuadro 7. Valores del logit, probabilidad y momio cuando IP=243

En este cuadro se puede observar que la probabilidad de parto múltiple va aumentando tanto para TPA = 0 como para TPA = 1, y en cada nivel de PG la probabilidad de parto múltiple es mayor cuando el parto anterior también fue múltiple.



GRAFICA 5. Probabilidad de parto múltiple dados los grupos de años para TPA=0

Esto puede verse más fácilmente en las Gráficas 5 y 6; para PG(3) la probabilidad de parto múltiple alcanza valores mayores a 0.7, cuando el parto anterior fue múltiple, en caso contrario, no supera el 0.6 de probabilidad. Para PG(1) la probabilidad de parto múltiple es pequeña en ambos casos, aunque en las gráficas se puede ver que cuando TPA=1 la probabilidad vuelve a ser mayor al caso cuando TPA=0.



GRAFICA 6. Probabilidad de parto múltiple dados los grupos de años para TPA=1

La razón de momios permite hacer una comparación de las probabilidades observadas entre grupos, éstos se calculan de la siguiente manera:

$$\frac{\text{momios}_{\text{GA2,TPA=1}}}{\text{momios}_{\text{GA1,TPA=1}}} = \frac{1.7476}{.5662} = 3.086$$

$$\frac{\text{momios}_{\text{GA3,TPA=1}}}{\text{momios}_{\text{GA2,TPA=1}}} = \frac{2.2461}{1.7476} = 1.2852$$

$$\frac{\text{momios}_{GA3,TPA=1}}{\text{momios}_{GA1,TPA=1}} = \frac{2.2461}{.5662} = 3.9669$$

Los resultados indican que es 3 veces más probable obtener un parto múltiple que uno simple en PG2 que en PG1, 28% más probable un múltiple que uno simple en PG3 que en PG2 y aproximadamente 4 veces más probable un parto múltiple que uno simple en PG3 que en PG1.

Para verificar el ajuste del modelo que tiene como variables de respuesta a PG1, PG2, TPA, IP se realizó la siguiente prueba de hipótesis:

$H_0$  : el modelo ajusta adecuadamente a los datos

o

$$H_0: \log \left( \frac{\pi}{1-\pi} \right) = \alpha + \beta_1 PG(1) + \beta_2 PG(2) + \beta_3 IP + \beta_4 TPA$$

vs.

$H_1$  : el modelo no ajusta adecuadamente a los datos

o

$$H_1: \log \left( \frac{\pi}{1-\pi} \right) \neq \alpha + \beta_1 PG(1) + \beta_2 PG(2) + \beta_3 IP + \beta_4 TPA$$

Para esto se calculó el estadístico de Devianza y se obtuvo el siguiente resultado:

$$G^2 = 2913.586$$

El cual se compara con una  $\chi^2$  con 1031 grados de libertad y con  $\alpha = .05$ :

$$\chi^2_{1031,.95} = 1107; \text{ p-value} \approx 1$$

por lo que no se rechaza la hipótesis nula y se concluye que el modelo ajusta adecuadamente los datos.

El valor de  $R^2_L$ , que es una medida de cuánto contribuyen las variables independientes a la explicación de la probabilidad de éxito, es:

$$R^2_L = \frac{2913.586}{3072.3437} = 0.9483$$

y

$$R^2_{LA} = \frac{2913.586 - 2(4)}{3072.3437} = 0.9457$$

para el caso ajustado, lo que quiere decir que las 3 variables independientes que permanecieron en el modelo, PG1, PG2, TPA, explican el 94.57% de la variación de éste mismo.

Con referencia a los residuos, de Pearson y de Devianza, éstos se encuentran dentro del intervalo (-1.58176, 2.1458), es decir, no existen valores extremos dentro de los registros.



## ***CONCLUSIONES***

De acuerdo con los resultados obtenidos, y bajo las condiciones del presente estudio, se puede concluir que el tipo de parto anterior (TPA) afecta el tamaño de camada, puesto que para las ovejas que han tenido parto múltiple anteriormente la probabilidad de parir dos o más corderos se incrementa.

De igual manera, el tipo de nacimiento de la madre (PG) influye de manera significativa en el tamaño de camada, puesto que la probabilidad de que una oveja que haya nacido en un parto múltiple tenga más de un cordero en un parto es mayor que para aquella oveja que nació en uno simple. Estas dos variables representan características propias de las borregas que no pueden ser controladas más que por medio de criterios de selección.

El intervalo entre partos no fue significativo, ya que en el análisis logístico se rechazó como variable importante en la explicación de la probabilidad, de lo cual se puede concluir que el tiempo que descansa la borrega entre parto y parto no influye en la probabilidad de parto simple o múltiple. Lo mismo que el número de parto (NP), es decir, la cantidad de partos que tenga la borrega no afecta el tamaño de la camada.

Como recomendación, se puede decir que, dados los datos observados, una buena manera de incrementar el número de corderos por oveja es elegir para el rebaño productor aquellas que hayan nacido en un parto múltiple y que también hayan tenido partos múltiples con anterioridad, es decir, eliminar del rebaño a las borregas que tengan partos simples desde el principio.

# **BIBLIOGRAFÍA**

1. Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc.
2. Agresti, A. c1986. *Statistical Methods for the Social Sciences*. Collier McMillan.
3. Berruecos V, José M; Valencia, Mario; Castillo Héctor. 1973. *Genética del Borrego Tabasco o Peligüey*. Técnica pecuaria.
4. Campos M. G. R.; Sánchez G. M. G.; Pliego C. B.; Castro G. H. 2001. *Comparación de la Productividad de Hembras de las tres Variedades de Color del Ovino Pelibuey*. Memorias del Segundo Congreso Latinoamericano de Especialistas en Pequeños Rumiantes Y Camélidos Sudamericanos. XI Congreso Nacional de Ovinocultura. Yucatán, México.
5. Cruz, L. C, Fernández-Baca, S., Álvarez L, Pérez R, H. 1994. *Variaciones estacionales en presentación de ovulación, fertilización y sobrevivencia embrionaria de ovejas Tabasco en el trópico húmedo*. *Veterinaria México*. 25(1): 23-27
6. Fayes, I., Marai, M. 1994. *Nuevas Técnicas de Producción Ovina*. Editorial Acribia.
7. Fernández B., S. *Memorias "Aspectos de Reproducción Ovina"*. Conferencia. 1980. Impresión de las memorias con fondos del Programa de Apoyo de las Divisiones de Estudios de Posgrado (PADEP) de la UNAM.
8. Fraser, A., Stamp, J.T. 1989. *Ganado Ovino, Producción y Enfermedades*. Ediciones Mundi-Prensa.

---

## BIBLIOGRAFÍA

9. García E. 1988. Modificaciones al sistema de clasificación de Köpen. Universidad Nacional Autónoma de México.
10. González G. R.; Torres H.; Becerril P. C.; Díaz R.P. 2001. Relación del Color del Pelaje y Factores Ambientales con Características Reproductivas en Ovejas Tropicales. *Agrociencia*. 35: 41-50
11. Guevara V, G., Romero M., S., Paneca A, A., Estévez A., Ceró R., A. 1993. Repetibilidad del Tamaño de la Camada en Ovejas Pelibuey. *Revista de Producción Animal, Cuba*. Vol. 7 No. 3: 177-179
12. Hosmer, D. Jr.; Lemeshow, S. 1989. *Applied Logistic Regression*. John Wiley and Sons.
13. Lanestosa Z., U. 1981. Contribución al Estudio del Borrego Pelibuey en el Estado de Tabasco. Tesis de Licenciatura. Universidad Juárez Autónoma de Tabasco.
14. Lloyd, C. 1999. *Statistical Analysis of Categorical Data*. John Wiley & Sons. First Edition.
15. Menard, S.W. 1995. *Applied Logistic Regression Analysis. Series: Quantitative Applications in Social Sciences*. Sage Publications.
16. Morales M., M. 1982. Parámetros Reproductivos y Explotación del Borrego Pelibuey. Tesis de Licenciatura. Universidad Juárez Autónoma de Tabasco. Escuela de Medicina Veterinaria y Zootecnia.
17. Plackett. 1974. *The Analysis of Categorical Data*. London: Criffin
18. Rojas R. O.; Rodríguez R. O. 1995. Factores que Modifican la Prolificidad en Ovejas Blackbelly en Clima Tropical. *Tec. Pec. Mex.* 33(3): 159-167

---

*BIBLIOGRAFÍA*

19. Romano M., J.L. 1981. Crecimiento y Composición Corporal del Borrego Pelibuey. UNAM.
20. Segura, J.C., Sarmiento, L., Rojas, O. 1996. Productivity of Pelibuey And Blackbelly ewes in Mexico under extensive management. Small Ruminant Research. 21(1): 57-62.
21. Serrato G., M.L. 1985. Comportamiento Reproductivo de los Borregos Pelibuey en el Altiplano. Tesis de Licenciatura. Universidad Autónoma del Estado de México. Facultad de Medicina Veterinaria y Zootecnia.
22. Valencia, Mario; Castillo, Héctor; Berruecos, José M. 1973. Reproducción y Manejo del Borrego Tabasco o Peligüey. Técnica Pecuaria.