

## Modelo de Odds Proporcionales Parciales Aplicado al Estudio de la Gravedad de Lesiones por Accidentes de Tránsito

Ronald Quispe Flores <sup>1</sup>, Ysela Agüero Palacios <sup>2</sup>

**Resumen:** Determinar los factores asociados con la gravedad de lesiones por accidentes de tránsito. Estudio descriptivo correlacional, de diseño observacional, con una muestra aleatoria de 1514 víctimas de accidentes de tránsito en la región de Madre de Dios, entre los años 2010 y 2014, registrados por DGE-MINSA. Se consideraron factores referidos al individuo, al vehículo y al entorno de ocurrencia del accidente. Se ajustó un modelo estadístico de odds proporcionales parciales ordinal. Se identificaron como factores significativamente asociados con la mayor gravedad de las lesiones por accidentes de tránsito ( $p < 0,05$ ), la edad de la víctima mayor de 65 años, accidentes involucrando vehículos motorizados grandes y el transporte de heridos al hospital por la policía o serenazgo.

**Palabras clave:** Accidentes de tránsito; Gravedad de lesiones; Modelo de odds proporcionales parciales.

## Partial Proportional Odds Model Applied to Study of Severity Injuries by Traffic Accidents

**Abstract:** To determine the risk factors associated with severity of traffic accidents injuries. Descriptive correlational, observational design, with a random sample of 1514 victims of traffic accidents in the Madre de Dios region, from January 2010 through December 2014, retrieved from DGE-MINSA. They are considered factors referred to the persons, the vehicle and the environment of the accident. Partial proportional odds models for ordinal response variable was fit. Were identified as significant risk factors for high levels of severity of injuries caused by traffic accidents ( $p < 0,05$ ), the victim higher to 65 years old, accidents involving big motorized vehicles and the transport of wounded to hospital by police or security guard.

**Keywords:** Road traffic injuries; Severity of injuries; Partial proportional odds model.

*Recibido:* 16/12/2016. *Aceptado:* 24/03/2017. *Publicado online:* 01/09/2017

<sup>1</sup>UNAMAD, Departamento Académico de Ciencias Básicas, e-mail: [rquispe@unamad.edu.pe](mailto:rquispe@unamad.edu.pe)

<sup>2</sup>UNMSM, Facultad de Ciencias Matemáticas, e-mail: [yaguerop@unmsm.edu.pe](mailto:yaguerop@unmsm.edu.pe)

## 1. Introducción

Los accidentes de tránsito constituyen una importante causa de morbimortalidad de la población. Según la Organización Mundial de la Salud [23], todos los años fallecen más de 1,2 millones de personas en las vías de tránsito del mundo, y entre 20 y 50 millones sufren traumatismos no mortales. Se prevé que para el año 2030, los traumatismos por accidentes de tránsito aumentarán para pasar a ser la quinta causa principal de mortalidad. En Perú, según informe de la Policía Nacional [15], el número de accidentes de tránsito registrado desde el año 2003 hasta el 2012, mantuvo una tendencia creciente. Respecto a los costos sociales, también se observa una tendencia creciente, pues, en el año 2010 se registraron 52572 víctimas de los cuales 94,6 % sufrieron lesiones y 5,4 % fallecieron [16], mientras que en el año 2012, se registraron 57693 víctimas con 5,6 % fallecidos. En la región de Madre de Dios, en el año 2012, el número de accidentes de tránsito se incrementó en 85,0 % respecto al año 2011. Entre los años 2007 y 2012, el 93 % de los involucrados en accidentes de tránsito presentaban lesiones graves o potencialmente graves, siendo uno de los más altos a nivel país [15]. Por otra parte, según el análisis de vulnerabilidad territorial basado en seis indicadores (tasa de accidentes de tránsito por 100000 habitantes, porcentaje de lesiones graves, porcentaje de conductores ebrios, porcentaje de lesionados trasladados por ambulancia del sector salud, tasa de mortalidad por accidentes de tránsito por 100000 habitantes y tasa de letalidad), la región Madre de Dios, junto con la región Puno, ocupan los primeros lugares en el grupo de departamentos con mayor vulnerabilidad en relación con los accidentes de tránsito [19].

## 2. Metodología

### 2.1. Modelo de regresión logística ordinal

El análisis de los datos, mediante la regresión logística multinomial, omitiría información importante correspondiente a la naturaleza ordinal de la variable respuesta, lo cual puede conllevar a la estimación de razones de ventajas, más conocidos como OR (“odds ratio”), que no responden a los objetivos del estudio. Ante esta necesidad surge el modelo de regresión logística ordinal (“*Ordered Regression Model*”) [20], introducido por McKelvey y Zavoina [13] en términos de una variable subyacente con categorías observadas ordenadas. Casi al mismo tiempo el modelo fue desarrollado en Bioestadística [14], donde es conocido como: modelo de odds proporcionales (“*Proportional Odds Model*”), modelo de regresión paralela (“*Parallel Regression Model*”) o modelo continuo agrupado (“*Grouped Continuous Model*”).

El modelo de regresión logística ordinal puede ser derivado de un modelo en el cual, la variable respuesta latente  $Y^*$  con valores en el intervalo  $(-\infty, \infty)$ , es operacionalizado de modo que la variable respuesta observada  $Y$ , toma  $k$  valores de acuerdo a la siguiente relación

$$Y_i = m, \quad \text{si } \tau_{m-1} \leq Y_i^* < \tau_m; \quad m = 1, \dots, k, \quad (1)$$

donde,  $\tau_0, \tau_1, \dots, \tau_k$  se denominan umbrales o puntos de corte (“*Thresholds or Cutpoints*”). El modelo estructural para la variable latente en la ecuación (1), se define como:

$$Y_i^* = X_i\beta + \varepsilon_i, \quad (2)$$

donde  $X_i$  es el vector de observaciones para las variables explicativas con el primer elemento igual a 1 para el intercepto,  $\beta$  es el vector de coeficientes estructurales correspondientes a las variables explicativas y  $\varepsilon_i$  es el error, que para el caso del modelo logístico ordinal, se asume que tiene distribución logística con media 0 y varianza  $\frac{\pi^2}{3}$ , con función de distribución de probabilidad acumulada dada por:

$$F(\varepsilon) = \frac{\exp(\varepsilon)}{1 + \exp(\varepsilon)}. \quad (3)$$

La probabilidad de  $Y_i = m$  para una observación, a partir de la relación (1) es:

$$P(Y_i = m/X_i) = P(\tau_{m-1} \leq Y_i^* < \tau_m/X_i). \quad (4)$$

La misma que, utilizando la ecuación estructural dada en (2) y restando la expresión  $X_i\beta$  en la correspondiente desigualdad, equivale a:

$$P(Y_i = m/X_i) = P(\tau_{m-1} - X_i\beta \leq \varepsilon < \tau_m - X_i\beta).$$

A su vez, en función de las probabilidades acumuladas se puede expresar como:

$$P(Y_i = m/X_i) = F(\tau_m - X_i\beta) - F(\tau_{m-1} - X_i\beta); \quad m = 1, \dots, k. \quad (5)$$

Por otra parte, la probabilidad que la variable observada, tome valores menores o iguales a  $m$ , es:

$$P(Y_i \leq m/X_i) = \sum_{j=1}^m P(Y_i = j/X_i); \quad m = 1, \dots, (k-1). \quad (6)$$

De acuerdo con la definición de probabilidad para la variable observada  $Y_i$  dada en la ecuación (5) y después de realizar algunas operaciones, la probabilidad acumulada anterior es equivalente a:

$$P(Y_i \leq m/X_i) = -F(\tau_0 - X_i\beta) + F(\tau_m - X_i\beta). \quad (7)$$

De acuerdo a la definición dada en (1),  $\tau_0 = -\infty$ , por lo que, según propiedad de distribución de probabilidad acumulada [24]  $F(\tau_0 - X_i\beta) = F(-\infty - X_i\beta) = 0$ . Por lo tanto, la probabilidad acumulada dada en (7) se reduce a

$$P(Y_i \leq m/X_i) = F(\tau_m - X_i\beta); \quad m = 1, \dots, (k-1). \quad (8)$$

La chance (“odds”) que la variable observada  $Y_i$  sea menor o igual a  $m$  se define como

$$\text{odds}(Y_i \leq m) = \frac{P(Y_i \leq m/X_i)}{1 - P(Y_i \leq m/X_i)}; \quad m = 1, \dots, (k-1).$$

De acuerdo a las definiciones de probabilidad acumulada dadas en (8) y (3) (para la variable observada y error aleatorio respectivamente), esta definición se puede escribir como:

$$\frac{P(Y_i \leq m/X_i)}{1 - P(Y_i \leq m/X_i)} = \exp(\tau_m - X_i\beta). \quad (9)$$

Tomando el logaritmo en esta última ecuación, se define el modelo logístico ordinal, como

$$\ln \left[ \frac{P(Y_i \leq m/X_i)}{1 - P(Y_i \leq m/X_i)} \right] = \tau_m - X_i\beta; \quad m = 1, \dots, (k - 1). \quad (10)$$

De manera similar, a partir de la definición dada en (9), el modelo logístico ordinal para odds de  $Y_i$  mayores a  $m$ , se expresa como,

$$\ln \left[ \frac{P(Y_i > m/X_i)}{P(Y_i \leq m/X_i)} \right] = -\tau_m + X_i\beta; \quad m = 1, \dots, (k - 1). \quad (11)$$

### 2.1.1. Modelo de razón de continuidad

Propuesto por Fienberg en 1980 [3]. Este modelo permite comparar la probabilidad que la respuesta sea igual a la categoría  $m$  con la probabilidad de que la respuesta sea mayor a dicha categoría; razón por la cual, resulta ser el modelo más apropiado cuando hay un interés intrínseco en una categoría específica de la variable respuesta. El modelo se define como

$$\ln \left[ \frac{P(Y_i = m/X_i)}{P(Y_i > m/X_i)} \right] = \alpha_m + (\beta_{m1}x_{i1} + \dots + \beta_{mp}x_{ip}); \quad m = 1, \dots, (k - 1), \quad (12)$$

donde,  $\alpha_m$  es el intercepto y  $\beta_{m1}, \dots, \beta_{mp}$ , son los coeficientes de las variables explicativas  $X_{m1}, \dots, X_{mp}$ , respectivamente

### 2.1.2. Modelo estereotipo

Propuesto por Anderson [2], este modelo es el más flexible para el análisis de respuestas ordinales y puede ser considerado una extensión del modelo de regresión multinomial. Compara cada categoría de la variable respuesta con una categoría de referencia, normalmente la primera o la última.

El modelo puede ser escrito en la siguiente forma

$$\ln \left[ \frac{P(Y_i = m/X_i)}{P(Y_i = k/X_i)} \right] = \alpha_m + \phi_m(\beta_1x_{i1} + \dots + \beta_px_{ip}); \quad m = 1, \dots, (k - 1). \quad (13)$$

En este caso, la última categoría  $k$  se considera como referencia,  $\alpha_m$  es el intercepto.  $\phi_m$  es la restricción usada para asegurar la identificabilidad del modelo con variable respuesta ordinal, cumpliendo la siguiente restricción( [8], [9])

$$0 = \phi_1 \geq \phi_2, \dots \geq \phi_k = 0$$

La combinación lineal que mejor discrimina entre los niveles de la variable respuesta está dada por  $\sum_{j=1}^p \beta_j x_{ij}$ . Las distancias entre los niveles de la respuesta, en términos de la combinación lineal indicada, están dadas por los parámetros  $\phi_m$  [7].

### 2.1.3. Modelo de odds proporcionales

La extensión del modelo de regresión logística para variables respuesta binarias, hacia variables repuesta ordinales, implica el modelamiento de logits acumulativos; es decir, cuando la variable respuesta  $Y_i$  es ordinal con valores  $1, \dots, k$ , las probabilidades acumuladas se pueden definir como

$$P(Y_i \leq m/X_i) = P(Y_1 = 1/X_i) + \dots + P(Y_m = m/X_i); \quad m = 1, \dots, k.$$

Permitiendo definir el modelo logit acumulativo (*cumulative logit model*) dada por la ecuación (10), que en forma extendida se puede escribir como

$$\ln \left[ \frac{P(Y_i \leq m/X_i)}{1 - P(Y_i \leq m/X_i)} \right] = \tau_m - (\beta_1 x_{i1} + \dots + \beta_p x_{ip}); \quad m = 1, \dots, (k - 1). \quad (14)$$

Es preciso indicar que este modelo consta de  $k - 1$  logit acumulativos como consecuencia del uso de  $k - 1$  puntos de corte para dicotomizar la variable respuesta  $Y_i$ . El coeficiente de regresión  $\beta_j$ ;  $j = 1, \dots, p$  para la variable explicativa  $X_j$ , es el logaritmo del *OR*. Donde el *OR* es una medida de asociación entre la variable respuesta dicotomizada  $Y_j$  y  $X_j$ , controlando las restantes variables explicativas. Puesto que el coeficiente  $\beta_j$  no depende de  $m$ , el modelo asume que la relación entre la variable explicativa  $X_j$  y la variable respuesta dicotomizada  $Y_i$ , no depende del punto  $m$  donde se hizo el corte. McCullagh [14] lo denomina supuesto de proporcionalidad de odds, de allí que el modelo dado en (14) se denomina **modelo de odds proporcionales**.

A fin de interpretar de manera usual los signos de los coeficiente de las variables explicativas, otra forma de expresar el modelo (14), es como sigue [1]

$$\ln \left[ \frac{P(Y_i > m/X_i)}{P(Y_i \leq /X_i)} \right] = \tau_m + (\beta_1 x_{i1} + \dots + \beta_p x_{ip}); \quad m = 1, \dots, (k - 1). \quad (15)$$

### 2.1.4. Modelo de odds proporcionales parciales

El supuesto de proporcionalidad de odds no necesariamente se cumple para todas las variables explicativas, algunas pueden cumplir, mientras que otras no necesariamente por estar más asociadas con la probabilidad de ocurrencia de alguna o algunas categorías de la variable respuesta. Es así que surge el modelo de odds proporcionales parciales [17], el cual se puede escribir como

$$\ln \left[ \frac{P(Y_i > m/X_i)}{P(Y_i \leq /X_i)} \right] = \tau_m + (\beta_1 + \gamma_{m1})x_{i1} + \dots + (\beta_q + \gamma_{mq})x_{iq} + \beta_{q+1}x_{iq+1} + \dots + \beta_p x_{ip}, \quad (16)$$

donde,  $m = 1, \dots, (k - 1)$ .

Además, se puede observar que las variables explicativas  $X_1, \dots, X_q$  ( $q \leq p$ ) no cumplen el supuesto de proporcionalidad de odds, porque adicionalmente a los coeficientes  $\beta$ , poseen los coeficientes  $\gamma$  que varían según la categoría  $m$  elegida para dicotomizar la variable repuesta. En cambio, las variables  $X_{q+1}, \dots, X_p$ , cumplen el supuesto, puesto que, solamente presentan los coeficientes  $\beta$  que no dependen de la categoría  $m$ .

Si todos los coeficientes adicionales fueran iguales a cero ( $\gamma_{m1} = \dots = \gamma_{mq} = 0$ ), entonces el modelo de odds propocionales parciales (16) sería equivalente al modelo de odds proporcionales (15). Es así que, se realiza una prueba simultanea de la suposición de proporcionalidad de odds, para las  $X_1, \dots, X_q$  variables explicativas, formulando la hipótesis nula como

$$H_0 : \gamma_{m1} = \dots = \gamma_{mq} = 0; \quad m = 2, \dots, (k - 1).$$

Se supone que  $\gamma_{11} = \dots = \gamma_{1q} = 0$ .

Por otra parte, si ninguna variable explicativa cumpliera el supuesto de proporcionalidad de odds ( $q = p$ ), entonces el modelo de odds proporcionales parciales dado en la relación (16) sería equivalente al modelo logístico ordinal generalizado [25]. Por estas razones, el modelo de odds proporcionales parciales, se puede considerar como un modelo intermedio entre el modelo de odds proporcionales y el modelo logístico ordinal generalizado.

En el modelo de odds proporcionales parciales dado en (16), una variable explicativa  $X_j$  ( $j = 1, \dots, q$ ) que no cumple el supuesto de proporcionalidad de odds, requiere  $k - 2$  parámetros adicionales  $\gamma_{mj}$  para representar sus coeficientes, compara que ésta cumpliera el supuesto. Sin embargo, si existiera una relación lineal entre el logaritmo de los  $OR$  acumulativos condicionales y los puntos de corte  $m$  para dicotomizar la variable respuesta, se requeriría sólo un parámetro adicional. Además, si esta condición fuera válida para las  $q$  variables que no cumplen el supuesto de proporcionalidad de odds, el modelo dado en (16) se escribiría como

$$\ln \left[ \frac{P(Y_i > m/X_i)}{P(Y_i \leq /X_i)} \right] = \tau_m + (\beta_1 + \alpha_m \gamma_1)x_{i1} + \dots + (\beta_q + \alpha_m \gamma_q)x_{iq} + \beta_{q+1}x_{iq+1} + \dots + \beta_p x_{ip}; \quad m = 1, \dots, (k - 1). \quad (17)$$

Así, el modelo dado en la ecuación (17), se denomina **modelo de odds proporcionales parciales restringido**, donde los escalares  $\alpha_m$ , con  $\alpha_1 = 0$ , son pre especificados y se les denomina restricciones.

En forma particular, si  $\alpha_m = m - 1$ , el logaritmo del  $OR$  en la primera ecuación del modelo es  $\beta_j$ , en el segundo  $\beta_j + \gamma_j$ , en el tercero  $\beta_j + 2\gamma_j$ , así sucesivamente hasta  $\beta_j + (k - 1)\gamma_j$  en la última ecuación.

Como se puede observar, el modelo (17), es apropiado si las variables explicativas requieren restricciones idénticas; sin embargo, se pueden establecer restricciones no necesariamente idénticas [17].

### Estimación

La estimación de parámetros para el modelo de odds proporcionales parciales dado en (16) y el modelo de odds proporcionales parciales restringido, se realiza mediante el método de máxima verosimilitud. Para este propósito el logaritmo de la función de verosimilitud del modelo de odds proporcionales parciales dado en (16), es:

$$L = \sum_{i=1}^n \sum_{m=1}^k I_{im} \log P(Y = m/X_i) = \sum_{i=1}^n \sum_{m=1}^k I_{im} \log P_{im}, \quad (18)$$

donde,  $I_{im}$  es una variable indicadora para la observación  $i$ , tal que,  $I_{im} = 1$ , si  $Y_i = m$  y  $I_{im} = 0$ , si  $Y_i \neq m$ ,  $m = 1, \dots, k$ , además  $P_{im}$  es:

$$\begin{aligned} P_{i1} &= 1 - C_{i1}, & \text{si } Y_i = 1 \\ P_{im} &= C_{i,m-1} - C_{i,m}, & \text{si } 1 < Y_i < k \\ P_{ik} &= C_{i,k-1}, & \text{si } Y_i = k \end{aligned}$$

donde,  $C_{im}$  es la probabilidad que la variable respuesta  $Y$  sea mayor a la categoría  $m$  en la  $i$ -ésima observación; es decir:

$$C_{im} = P(Y > m/X_i) = \frac{1}{1 + \exp[-\tau_m - (\beta_1 + \gamma_{m1})x_{i1} - \dots - (\beta_q + \gamma_{mq})x_{iq} - \beta_{q+1}x_{iq+1} - \dots - \beta_p x_{ip}]}$$

La estimación de los parámetros que maximizan la función de verosimilitud, se obtienen resolviendo mediante métodos numéricos, el sistema de ecuaciones que resulta de igualar a cero cada gradiente del logaritmo de la función de verosimilitud. Las estimaciones resultantes, son asintóticamente eficientes, consistentes y normales; es decir:

$$\widehat{\beta} \sim N(\beta, Var(\widehat{\beta})).$$

Para un parámetro individual

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j)),$$

donde, la varianza del vector de estimadores máximo verosímil  $\widehat{\beta}$ , se define como la inversa de la matriz de información [20]; es decir:

$$Var(\widehat{\beta}) = -E[H(\beta)]^{-1},$$

siendo  $H(\beta)$ , la matriz Hessiana del logaritmo de la función de verosimilitud.

Es preciso indicar que,  $Var(\widehat{\beta}_j)$ , es el  $j$ -ésimo elemento de la diagonal principal de  $Var(\widehat{\beta})$ .

### Interpretación

Una variable explicativa  $X_j$  que cumple el supuesto de proporcionalidad de odds, tiene efecto uniforme para todos los niveles de la variable respuesta [21], el parámetro correspondiente  $\beta_j$  tiene la interpretación usual del logaritmo de  $OR$  acumulado, en cada uno de los  $k - 1$  *logit* acumulados [1]. Así por ejemplo, para un incremento de  $X_j$  en  $\delta$ , la razón de odds para  $X_j$  y  $X_j + \delta$ , obtenidas a partir de la ecuación (17), viene a ser:

$$OR = \frac{odds(X, X_j)}{odds(X, X_j + \delta)} = exp(\delta\beta_j). \quad (19)$$

Es decir, para un incremento en  $X_j$ , el odds o chance que la variable respuesta sea mayor a la categoría  $m$ , es cambiada por el factor  $exp(\delta\beta_j)$ , permaneciendo constantes todas las restantes variables explicativas.

El efecto de una variable explicativa que no cumple el supuesto de proporcionalidad de odds, es explicado mediante el efecto marginal, dado que, el signo de su coeficiente para categorías intermedias de la variable respuesta, no siempre determina la dirección de su efecto [21].

El efecto marginal o cambio parcial, es la pendiente de la curva que relaciona  $X_j$  a  $P(Y_i = m/X_i)$ , permaneciendo constante las demás variables explicativas [20]. Es así que, el efecto marginal para  $X_j$ , a partir de la ecuación (5) es:

$$\frac{\partial P(Y_i = m/X)}{\partial X_j} = \beta_j [f(\tau_{m-1} - X\beta) - f(\tau_m - X\beta)]. \quad (20)$$

Sin embargo, la interpretación usando el efecto marginal puede resultar erróneo cuando la curva de probabilidad cambia rápidamente o cuando la variable explicativa es artificial (dummy). Para el modelo regresión ordinal, la medida de cambio discreto es más informativo [20].

En efecto, el cambio discreto, es el cambio en la probabilidad predicha, para un cambio en  $X_j$  del

valor inicial  $x_s$  al valor final  $x_e$  (por ejemplo de 0 a 1), expresado como:

$$\frac{\Delta P(Y_i = m/X)}{\Delta X_j} = P(Y_i = m/X, X_j = x_e) - P(Y_i = m/X, X_j = x_s), \quad (21)$$

donde  $P(Y_i = m/X = x_j)$ , es la probabilidad que  $Y = m$ , dado  $X$ , observando un valor específico de  $x_j$ . El cambio dado en (21) se interpreta de la siguiente forma: cuando  $X_j$ , cambia del valor inicial  $x_s$  al valor final  $x_e$ , la probabilidad predicha de  $Y = m$ , cambia en  $\Delta P(Y_i = m/X)/\Delta X_j$ , permaneciendo constante las restantes variables explicativas.

Dado que, el modelo no es lineal, el valor del cambio discreto, depende de tres factores: (a) nivel de las variables explicativas que “no están cambiando”; (b) valor inicial del cambio de  $X_j$ ; (c) aumento en el cambio de  $X_j$ . El valor inicial de  $X_j$  y el aumento, dependen del propósito de análisis.

Precisamente, respecto al factor (a), existen diferentes formas de controlar las variables explicativas que “no están cambiando”. Una forma es considerando los promedios, para obtener los “efectos marginales en el promedio” (MEMs); una segunda forma es considerar los datos tal como se observaron para obtener los “efectos marginales promedio” (AMEs) y otra forma es considerar valores representativos, para obtener los “efectos marginales en valores representativos” (MERS) [26].

### Prueba del supuesto de proporcionalidad de odds o paralelismo de regresión

Existen dos métodos para la prueba del supuesto de paralelismo de regresión: Test Score (*Score test*) y el test de Wald (*Wald test*) [20].

**Test Score**, para entender como ésta prueba se utiliza en la evaluación del supuesto de paralelismo, considerar el modelo de regresión ordinal como un conjunto de  $k - 1$  logits binarios de la forma

$$P(Y \leq m/X) = F(\tau_m - X\beta_m); \quad m = 1, \dots, (k - 1), \quad (22)$$

donde, se impone la restricción que los vectores de coeficientes asociados con las variables explicativas para cada uno de los logits binarios, son iguales; es decir:

$$\beta_1 = \beta_2 = \dots = \beta_{k-1} = \beta. \quad (23)$$

Estimando el modelo dado en (22), la prueba Score evalúa cómo el logaritmo de verosimilitud del modelo de regresión ordinal cambia si la restricción (23) se suprime.

El estadístico de prueba resultante tiene distribución *chi-cuadrado* con  $p(k - 2)$  grados de libertad.

**Test de Wald**, permite probar que los vectores de coeficientes,  $\beta_m$ , son iguales para todos los  $k - 1$  logits acumulados, así mismo, permite probar la igualdad de coeficientes para variables individuales. Es así que, para la prueba de igualdad de los coeficientes, la hipótesis nula es

$$H_0 : \beta_{-1} = \dots = \beta_{-k-1}.$$

La notación con sub-guión para los vectores de coeficientes, indica que sólo se consideran los coeficientes de las variables explicativas para cada logit, más no, los interceptos. El estadístico de prueba tiene distribución *chi-cuadrado* con  $p(k - 2)$  grados de libertad.

La hipótesis nula para la prueba de igualdad de coeficientes de una variable individual es:

$$H_0 : \beta_{j1} = \dots = \beta_{j,k-1}.$$

El estadístico de prueba tiene distribución *chi – cuadrado* con  $(k - 2)$  grados de libertad.

Además de estas pruebas, la opción “*autofit*” de STATA [25], primero ajusta el modelo sin restricción de proporcionalidad de odds, seguidamente, mediante un proceso de ajuste iterativo y utilizando la prueba de Wald para cada variable, determina las variables que cumplen el supuesto de proporcionalidad de odds y las que no cumplen. Finalmente, realiza un prueba global de Wald del modelo final con restricciones versus el modelo inicial sin restricciones, un valor de la prueba estadísticamente no significativos, indica que el modelo final cumple el supuesto de paralelismo o proporcionalidad de odds.

### Prueba de bondad de ajuste

Dado que la estimación de parámetros se realiza mediante el método de máxima verosimilitud, la prueba de bondad de ajuste del modelo se realiza mediante la prueba de razón de verosimilitud “*likelihood ratio*”(LR). La misma que, se define de la siguiente forma. El modelo restringido  $M_C$  con parámetros  $\beta_c$  es anidado en el modelo no restringido  $M_U$  con parámetros  $\beta_U$ . La hipótesis nula es que la restricción impuesta para crear  $M_C$ , es verdadera. Sea  $L(M_U)$  el valor de la función de verosimilitud evaluado en la estimación de máxima verosimilitud del modelo no restringido, y sea  $L(M_C)$  el valor de la función de verosimilitud evaluado en la estimación de máxima verosimilitud del modelo restringido. EL estadístico de la prueba de razón de verosimilitud es [20]

$$G^2(M_C/M_U) = 2\ln L(M_U) - 2\ln L(M_C). \quad (24)$$

Bajo condiciones muy generales, si la hipótesis nula es verdadera,  $G^2$  tiene distribución chi-cuadrado asintótica, con grados de libertad igual al número de restricciones independientes. Esta prueba puede ser utilizado para comparar cualquier par de modelos anidados, pero, frecuentemente se utiliza para comparar un modelo dado con un modelo restringido en la que todos los coeficientes de las variables explicativas son iguales a cero. Esta prueba es conocida como la prueba de razón de verosimilitud chi-cuadrado (“*LR chi-square*”). Para definir el estadístico de prueba, consideremos el modelo  $M_\beta$  no restringido que incluye el intercepto, las pendientes y otros parámetros del modelo, y sea,  $M_\alpha$  el modelo restringido que no incluye las pendientes del modelo. De esta forma, para probar la hipótesis que todos los coeficientes de las variables explicativas o pendientes son iguales a cero, el estadístico de prueba es:

$$G^2(M_\alpha/M_\beta) = 2\ln L(M_\beta) - 2\ln L(M_\alpha). \quad (25)$$

El mismo, bajo la hipótesis nula (los coeficientes de todas las variables explicativas son iguales a cero) verdadera, tiene distribución chi-cuadrada con grados de libertad igual al número de variables explicativas.

## 2.2. Tipo y diseño de investigación

El tipo de investigación fue descriptivo correlacional y el diseño fue observacional con obtención de datos de fuente secundaria.

### 2.3. Unidad de análisis y población de estudio

La unidad de análisis fue un individuo lesionado en un accidente de tránsito ocurrido en la región Madre de Dios, entre enero del 2010 y noviembre del 2014, el mismo, fue atendido en un establecimiento del Ministerio de Salud (MINSA) de esta misma región. La población de estudio estuvo conformado por 1954 individuos.

### 2.4. Tamaño de muestra

El tamaño de muestra se determinó en 1514, para un tasa de incidencia de accidente de tránsito de 0.5, error relativo 0.02, nivel de confianza del 95 % y tasa de no respuesta (registros incompletos) de 28.8 %

### 2.5. Variables

De acuerdo a la teoría epidemiológica de los accidentes de tránsito y la matriz de Haddon [6]), las variables explicativas estudiadas corresponden al huésped o víctima, agente o energía transferida y entorno o lugar de ocurrencia del accidente. Estas variables se encuentran operacionalizadas en la ficha de vigilancia epidemiológica de lesiones por accidentes de tránsito del MINSA [18]. La variable respuesta, (“*gravedad de lesión*”), definida a partir de la condición de egreso del establecimiento de salud, del lesionado, es ordinal con las siguientes categorías: (i) No requiere rehabilitación, (ii) Requiere rehabilitación y (iii) Fallecido.

### 2.6. Datos

La Dirección General de Epidemiología (DGE) del MINSA en coordinación con las direcciones regionales de salud (DIRESA), realiza la vigilancia epidemiológica de lesiones por accidentes de tránsito a nivel nacional y en particular para la región de Madre de Dios, por lo que se solicitó la base de datos digital para la mencionada región a la DGE - MINSA y a la Oficina Regional de Epidemiología de Madre de Dios, para luego obtener los datos de la muestra. El procesamiento de datos se realizó utilizando el software estadístico STATA versión 13.

## 3. Resultados

Se ajustaron tres modelos logísticos ordinales relacionando la gravedad de lesión con cada uno de los factores (Individuo, vehículo y entorno). En estos ajustes, se observó que la edad del lesionado; la condición del vehículo (particular/oficial), el tipo de vehículo; la vía de ocurrencia del accidente, el día de ocurrencia y el medio de traslado de los lesionados que corresponden a las características de la víctima, del vehículo y del entorno respectivamente, explican significativamente la variabilidad de la gravedad de lesión.

**Tabla 1. Gravedad de lesiones en accidentes de tránsito según características individuales, del vehículo y del entorno - Modelo de odds proporcionales parciales.**

Variables	P. de paralelismo	Panel I		Panel II	
	$p^{\mathfrak{S}}$	Coef.	$p^{\mathfrak{N}}$	Coef.	$p^{\mathfrak{N}}$
<b>Edad (años)</b>					
De 15 a 29	0,517	0,618	0,008	0,618	0,008
De 30 a 65	0,991	0,723	0,002	0,723	0,002
Mayor de 65	0,014	1,145	0,031	2,651	0,000
<b>Condición del vehículo</b>					
Particular	0,456	0,826	0,000	0,826	0,000
<b>Tipo de vehículo</b>					
Automóvil	0,532	0,296	0,156	0,296	0,156
V. M. mayor	0,044	0,341	0,211	1,319	0,007
<b>Vía</b>					
Carretera	0,293	0,633	0,000	0,633	0,000
<b>Día</b>					
De sábado a lunes	0,922	0,229	0,122	0,229	0,122
<b>Traslado del lesionado</b>					
Policía o serenazgo	0,024	0,688	0,000	1,610	0,000
Ambulancia	0,803	1,050	0,000	1,050	0,000
Umbral		-2,350	0,000	-6,242	0,000

Prueba de razón de verosimilitud del modelo: Chi-cuadrado = 111,61; g.l.=13; p=0.000.

Categorías de referencia: Menor de 15; De empresa/institución; V. M. menor; Calle, jirón o avenida;

De martes a viernes; Involucrado o tercero.

$\mathfrak{S}$ : Prueba Wald significativo al 0,05.

$\mathfrak{N}$ : Prueba Z significativo al 0,05.

Luego, el modelo general, ajustado con las variables indicadas en el párrafo anterior, es significativo (prueba de razón de verosimilitud chi-cuadrado:  $p < 0,05$ ), indicando que, al menos una de las características de la víctima, del vehículo o del entorno, contribuye de manera significativa la explicación de la variabilidad en la gravedad de lesión. Precisamente, estas variables son la edad, la condición del vehículo, el tipo de vehículo, la vía de ocurrencia y medio de traslado del lesionado (prueba Z:  $p < 0,05$ ). A su vez, para estas variables se observa que las categorías edad mayor de 65 años, vehículo motorizado mayor y traslado del lesionado por policía o serenazgo, no cumplen el supuesto de proporcionalidad de odds o paralelismo (prueba Wald:  $p < 0,05$ ), en cambio las categorías de edad, de 15 a 29 años, de 30 a 65 años, condición de vehículo particular, vía de ocurrencia carretera y traslado de lesionado en ambulancia, si cumplen (Tabla 1)

### Estimación de la razón de odds (OR)

Para cuantificar e interpretar la asociación de la gravedad de lesión con variables explicativas que cumplen el supuesto de proporcionalidad de odds, tal como se indica en la sección (2.1.4) (interpretación), se estimaron los *OR*, cuyos resultados se presentan en la tabla 2.

**Tabla 2.** “Odds ratios” ajustados para categorías que cumplen el supuesto de proporcionalidad de odds.

Variables	Panel I			Panel II		
	$\widehat{OR}$	IC (95 %)		$\widehat{OR}$	IC (95 %)	
<b>Edad (años)</b>						
De 15 a 29	1,86	1,18	2,92	1,86	1,18	2,92
De 30 a 65	2,06	1,30	3,27	2,06	1,30	3,27
<b>Condición del vehículo</b>						
Particular	2,29	1,48	3,54	2,29	1,48	3,54
<b>Vía</b>						
Carretera	1,88	1,37	2,58	1,88	1,37	2,58
<b>Traslado del lesionado</b>						
Ambulancia	2,86	1,68	4,86	2,86	1,68	4,86

Categorías de referencia: Menor de 15; De empresa/institución; Calle, jirón o avenida; Involucrado o tercero.

Es así que respecto a la edad, considerando como referencia a los menores de 15 años, los lesionados de 15 a 29 años tienen 1,86 veces más posibilidades de sufrir accidentes de mayor gravedad ( $IC(95\%) : 1,18 - 2,92$ ). Mientras que, para los del grupo de edad de 30 a 65 años el riesgo se incrementa pues, la chance de una mayor gravedad del accidente es 2,06 veces mayor ( $IC(95\%) : 1,30 - 3,27$ ).

Respecto a la condición del vehículo, los lesionados en accidentes de tránsito ocasionados por vehículos particulares, tienen 2,29 ( $IC(95\%) : 1,48 - 3,54$ ) veces más posibilidades de sufrir accidentes de mayor gravedad, en comparación a los lesionados en accidentes ocasionados por vehículos de empresas o instituciones.

Con respecto al lugar de ocurrencia del accidente, los lesionados en accidentes de tránsito ocurridos en carreteras, tienen 1,88 ( $IC(95\%) : 1,37 - 2,58$ ) veces más posibilidades de sufrir accidentes de mayor gravedad, en comparación a los lesionados en accidentes de tránsito suscitados en calles, avenidas o jirones.

El modo de traslado del lesionado para ser atendido en un establecimiento de salud es de vital importancia. En el análisis de datos se observa que los lesionados trasladados en ambulancia, tienen 2,86 ( $IC(95\%) : 1,68 - 4,86$ ) veces mayor la posibilidad de haber sufrido una lesión más grave, comparado con los lesionados que se trasladaron por sí mismos, por el ocasionante o por terceras personas.

**Estimación de efectos marginales (E.M.)**

Para las variables explicativas que no cumplieron el supuesto de proporcionalidad de los odds, como se indica en la sección (2.1.4) (interpretación), se calcularon los cambios discreto promedio (AMEs), cuyos resultados, bajo el encabezado de E.M. se presentan en la tabla 3.

**Tabla 3. Efectos marginales de las categorías que no cumplen el supuesto de proporcionalidad de odds.**

Variables	Gravedad de lesión					
	No rehabilitación		Rehabilitación		Fallecido	
	E. M.	$p^\phi$	E. M.	$p^\phi$	E. M.	$p^\phi$
<b>Edad (años)</b>						
Mayor de 65	-0,249	0,035	0,095	0,387	0,154	0,034
<b>Tipo de vehículo</b>						
V. M. mayor	-0,076	0,217	0,017	0,980	0,059	0,051
<b>Traslado del lesionado</b>						
Policía o serenazgo	-0,157	0,000	0,102	0,011	0,055	0,002

$\phi$ : Prueba Z significativo al 0,05.

E.M.: Efecto Marginal (Cambio discreto promedio).

Según estos resultados, las personas lesionadas en accidentes de tránsito con edades mayores de 65 años, tienen una mayor probabilidad estimada de fallecer ( $E.M = 0,154, p < 0,05$ ), comparado con los lesionados menores de 15 años. Las personas lesionadas en accidentes de tránsito ocasionados por vehículos motorizados mayores, tienen mayor probabilidad estimada de fallecer ( $E.M. = 0,059, p = 0,051$ ), comparado con los lesionados en accidentes de tránsito ocasionados por motocicletas o motocar. Los lesionados trasladados al establecimiento de salud por la policía o serenazgo, tienen mayor probabilidad estimada de requerir rehabilitación ( $E.M. = 0,102, p < 0,05$ ); o fallecer ( $E.M. = 0,055, p < 0,05$ ), comparado con los lesionados trasladados por terceras personas o involucradas.

## 4. Discusión

Singleton y otros [22], encontraron que la edad es uno de los factores de riesgo para altos niveles de gravedad de lesión, así mismo, López [10] y Hanrahan y otros [5] indican que el aumento de la edad se asocia fuertemente con el riesgo de sufrir graves lesiones o fallecer. También, en otros estudios ([21], [28]), se determinó que la probabilidad de ocurrencia de una fatalidad o lesión grave, es alta para personas mayores. Este mismo fenómeno, se observa en el presente estudio dado que, las personas mayores de 65 años presentan mayor probabilidad de fallecer, lo cual confirmaría que la edad avanzada es un factor de riesgo de graves lesiones en accidentes de tránsito.

En éste estudio, se encontró asociación entre la gravedad de lesiones y el tipo de vehículo particular. Resultados similares o contrarios no se han podido encontrar. De alguna forma, la peculiaridad del resultado podría obedecer a dos aspectos, al tipo de vehículo y a la responsabilidad del conductor. Puesto que, en el ámbito geográfico de estudio, los vehículos de uso frecuente son las motocicletas y motocar conducidos por personas que en su mayoría incumplen el reglamento de tránsito.

La posibilidad de ocurrencia de lesiones graves o fatales son altas en accidentes de tránsito que involucran vehículos pesados [21]. Este fenómeno también se observa en la presente investigación, puesto que los lesionados en accidentes de tránsito con vehículos motorizados mayores presentaron mayor probabilidad de fallecer. Estos resultados pueden obedecer a la enorme potencia destructiva de una gran masa unida a la fuerza de impacto a gran velocidad [11].

Con respecto al lugar de ocurrencia del accidente, en un estudio realizado [21] se determinó que la probabilidad de fallecer o sufrir lesión grave, es baja en zonas urbanas . En el presente estudio, la posibilidad de sufrir mayor gravedad de lesión se encuentra en carreteras en comparación con calles, avenidas o jirones que se encuentran en la zona urbana, lo cual concuerda con el estudio previo. Estos resultados posiblemente obedezcan a la oportuna asistencia que de alguna forma tienen las víctimas de accidentes de tránsito en zonas urbanas.

En el presente estudio, los lesionados trasladados al establecimiento de salud por la policía o el serenazgo presentan mayor probabilidad de requerir rehabilitación o fallecer, lo cual podría indicar la falta de una atención especializada de parte de las instituciones de salud, hacia las víctimas de accidentes de tránsito, puesto que, según Gutierrez y otros [4], muchas acciones pueden realizarse para minimizar el impacto de las lesiones producidas; siendo esencial la disponibilidad y eficiencia del cuidado prehospitalario y hospitalario.

Varios trabajos de investigación ([21], [27] y [28]) utilizaron el modelo de odds proporcionales parciales, para identificar los factores asociados con la mayor o menor gravedad de lesiones por accidentes de tránsito. De la misma forma, Mooradian y otros [12] indican que el modelo de odds proporcionales parciales, presenta mejor ajuste a comparación de otros modelos, en el estudio de la gravedad de lesión por accidentes de tránsito. En el presente estudio también, el uso de este modelo permitió una mejor identificación de los factores de riesgo asociados con la mayor gravedad de lesión.

## 5. Conclusiones

La edad del lesionado, el tamaño y condición del vehículo ocasionante del accidente de tránsito y el medio de traslado del lesionado, se asocian con la gravedad de lesión por accidente.

La edad mayor de 65 años y los vehículos motorizados mayores (microbuses, camiones o trailer), son factores de riesgo para el mayor nivel de gravedad de lesión; además, los lesionados trasladados al establecimiento de salud por la policía o serenazgo, tienen mayor probabilidad de requerir rehabilitación o fallecer.

El modelo logístico ordinal de odds proporcionales parciales, es una técnica de análisis de datos útil para identificar los factores asociados con la gravedad de lesión por accidente de tránsito, así mismo, permite identificar los factores de riesgo asociados con la mayor gravedad de lesión y los factores protectores.

## Referencias Bibliográficas

- [1] Agresti, A. (2010). *Analysis of ordinal categorical data*. 2a ed. USA: John Wiley and Sons.
- [2] Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46 (1), 1-30.
- [3] Fienberg, S. E. (2007). *The analysis of cross-classified categorical data*. USA: Springer Science and Business Media.
- [4] Gutierrez, C., Romaní, F., Wong-Chero, P., y Montenegro-Hidrogo, J. J. (2014). Perfil epidemiológico de la discapacidad por accidentes de tránsito en el Perú, 2012. *Revista Peruana de Medicina Experimental y Salud Pública*, 31(2), 267-273.
- [5] Hanrahan, R., Layde, P., Zhu, S., Guse, C., & Hargarten, S. W. (2009). The Association of Driver Age with Traffic Injury Severity in Wisconsin. *Traffic Injury Prevention* 10(4), 361-367.
- [6] Haddon W. (1980). Advances in the epidemiology of injuries as a basis for public policy. *Public health reports*, 95(5), 411.
- [7] Lunt, M. (2005). Prediction of ordinal outcomes when the association between predictors and outcome differs between outcome levels. *Statistics in medicine*, 24(9), 1357-1369.
- [8] Liu, X. (2014). Fitting Stereotype Logistic Regression Models for Ordinal Response Variables in Educational Research (Stata). *Journal of Modern Applied Statistical Methods*, 13(2), 31.
- [9] Long, J. (2012). *Regression models for nominal and ordinal outcomes*. Recuperado el 23 de enero del 2017 de <http://www.indiana.edu/~jslsoc>.
- [10] López, C. (2007). *Factores de riesgo de mortalidad y morbilidad en accidentes de tráfico de ciclomotores y motocicletas*. Tesis de Doctor en Medicina Preventiva y Salud Pública. Universidad de Granada, España.
- [11] Lam, L. T. (2004). Environmental factors associated with crash-related mortality and injury among taxi drivers in New South Wales, Australia. *Accident Analysis & Prevention*, 36(5), 905-908.
- [12] Mooradian, J., Ivan, J. N., Ravishanker, N., & Hu, S. (2013). Analysis of driver and passenger crash injury severity using partial proportional odds models. *Accident Analysis & Prevention*, 58, 53-58.
- [13] McKelvey, R. D., y Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of mathematical sociology*, 4(1), 103-120.
- [14] McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society Series B (Methodological)*, 42 (2) , 109-142.
- [15] Policía Nacional del Perú, Dirección Nacional de Gestión Institucional Dirección Ejecutiva de Información y Comunicaciones, Dirección de Estadística. (2014). *ANUARIO ESTADÍSTICO 2012*. PNP. Lima.

- [16] Policía Nacional del Perú, Estado Mayor General. (2011). *ANUARIO ESTADÍSTICO 2010*. PNP. Lima.
- [17] Peterson, B., Harrell E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39 (2), 205-217.
- [18] Perú, Ministerio de Salud. (2007). *Norma técnica de salud para la vigilancia epidemiológica de lesiones por accidentes de tránsito*. MINSA. Lima.
- [19] Rojas, Y., Arias, M. (2013). *Análisis epidemiológico de las lesiones causadas por accidentes de tránsito en el Perú, 2013*. Ministerio de Salud, Dirección General de Epidemiología. Lima: SINCO Editores SAC
- [20] Scott, L. (1997). *Regression models for categorical and limited dependent variables*. Advanced quantitative techniques in the social sciences 7. USA: SAGE Publications
- [21] Sasidharan, L., Menendez, M. (2014). *Partial proportional odds model-An alternate choice for analyzing pedestrian crash injury severities*. *Accident Analysis & Prevention*, 72, 330-340.
- [22] Singleton, M., Qin, H., y Luan, J. (2004). Factors Associated with Higher Levels of Injury Severity in Occupants of Motor Vehicles That Were Severely Damaged in Traffic Crashes in Kentucky, 2000-2001. *Traffic Injury Prevention*, 5(2), 144-150.
- [23] World Health Organization, Dept. of Violence and Injury Prevention and Disability. (2009). *Informe sobre la situación mundial de la seguridad vial: es hora de pasar a la acción*. Ginebra: Organización Mundial de la Salud.
- [24] Wackerly, D. D., Mendenhall, W., Scheaffer, R. L., Muñoz, J. R., y Hernández, A. (2010). *Estadística matemática con aplicaciones*. 7a ed. México: Cengage Learning Editores.
- [25] Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, 6(1), 58-82.
- [26] Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal*, 12(2), 308.
- [27] Wang, X., & Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis & Prevention*, 40(5), 1674-1682.
- [28] Wang, C., Lu, L., & Lu, J. (2015). statistical Analysis of Bicyclists' Injury Severity at Unsignalized Intersections. *Traffic Injury Prevention*, 16(5), 507-512.