www.mednet.gr/archives
ARCHIVES OF HELLENIC MEDICINE: ISSN 11-05-3992

SPECIAL ARTICLE
ΕΙΔΙΚΟ ΑΡΘΡΟ

ARCHIVES OF HELLENIC MEDICINE 2026, 43(5):682–690
ΑΡΧΕΙΑ ΕΛΛΗΝΙΚΗΣ ΙΑΤΡΙΚΗΣ 2026, 43(5):682–690

Ordinal logistic regression as a tool to estimate the risk of ranked outcomes

C. Gnardellis,[1]
V. Notara,[2]
G. Tzamalouka,[3]
M. Papadakaki,[3]
J. Chliaoutakis[3]

[1]Department of Fisheries and Aquaculture, School of Agricultural Sciences, University of Patras, Messolonghi
[2]Department of Public and Community Health, School of Public Health, University of West Attica, Athens
[3]Department of Social Work, School of Health Sciences, Hellenic Mediterranean University, Heraklion, Crete, Greece

The use of logistic regression models in data analysis and machine learning has expanded in recent years and has become the preferred choice of researchers in risk assessment studies across various scientific fields. These models are widely applied, from identifying the etiological factors of chronic diseases to assessing the risk of traffic accidents or evaluating credit risk in financial processes. All logistic regression models are natural extensions of the simple binary model, and their interpretation is based on it. Using data from a cross-sectional study on drowsiness while driving and its association with traffic accident risk, this study presents a detailed examination of the two main extensions of logistic regression techniques: multinomial and ordinal logistic regression. Special emphasis is placed on ordinal regression, as the outcome variable in the collision data is measured on an ordinal scale, reflecting an underlying continuous latent variable.

Η διατακτική λογαριθμική παλινδρόμηση ως μέθοδος για την εκτίμηση του κινδύνου σε διαβαθμιζόμενα αποτελέσματα

Περίληψη στο τέλος του άρθρου

Key words

Car accident data
Collision risk
Multinomial logistic regression
Ordinal logistic regression
Partial proportional odds model (PPOM)
Proportional odds model (POM)

Submitted 20.3.2025
Accepted 26.4.2025

1. INTRODUCTION

In empirical research, it is often necessary to estimate the risk associated with ranked outcomes defined by an ordinal variable. While the binary logistic regression model is suitable for a binomial event, it is inadequate for estimating a multinomial outcome with an internal ordering.[1–4] Typically, the analysis of such data –i.e., a response variable with multiple outcomes and a set of predictor variables– is performed using the multinomial logistic regression model. The latter extends the simple logistic model, with its functional definition expanding upon the properties of binary logistic regression.[5–7]

When the response variable in a model is binary, the logistic regression equation (logit) that applies to the odds of success is:

$$ln\left(\frac{p_{event}}{1\text{-}p_{event}}\right) = b_0 + b_1x_1 + b_2x_2 + ...+ b_kx_k$$

where $x_1, x_2, \ldots, x_k$ are the values of the predictor variables $X_1, X_2, ..., X_k$.

When the response variable is categorical with $m > 2$ categories, m-1 non-redundant logits can be defined. One of the dependent variable's categories is designated as the reference category (baseline category=j), and based on this, m-1 different logits are estimated using the general form:

$$ln\left(\frac{p_{category\ i}}{p_{category\ j}}\right) = b_{i0} + b_{i1}x_1 + b_{i2}x_2 + ... + b_{ik}x_k$$

or equivalent

$$\frac{p_{category\ i}}{p_{category\ j}} = e^{(b_{i0} + b_{i1}x_1 + b_{i2}x_2 + ... + b_{ik}x_k)}$$

For each of these logits, a set of coefficients is derived for the k predictor variables in the multinomial model. Each coefficient estimates the odds of occurrence of category i relative to the reference category j.

For example, if the response variable in a model has three levels, two sets of non-zero coefficients are generated, each comparing two levels of the variable against the reference category. For the reference category, the coefficients in the corresponding logit are all zero. By changing the reference category, all possible logits of a multinomial logistic model can be estimated. Multinomial logistic regression is nearly equivalent to running a series of binary logistic regressions on the data. However, the key difference is that the equations are fitted simultaneously rather than separately,[6] imposing an additional constraint; namely, that the predicted probabilities of all possible outcomes must sum to 1.

The multinomial logistic regression model can be modified and simplified when the outcome is ordinal. In such cases, incorporating the ordering of the response variable into the analysis can enhance interpretability and increase the statistical power of the model. Ordinal logistic regression is a specialized type of multinomial regression and is particularly useful when the response variable is ordinal.[6,8–12]

Starting again from the simple logistic model, where the occurrence of a binary event is the response variable:

$$ln\left(\frac{p_{event}}{1\text{-}p_{event}}\right) = b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k$$

We modify it to accommodate cases where the event of interest is not a single discrete outcome (e.g., yes/no), but a set of possible outcomes coded according to an ordinal variable. For example, in vehicle accident data, the response variable might represent the number of crashes coded as follows: 0=no crash, 1=up to two crashes, 2=three or more crashes. For each possible value of the crashes variable Y, we define the following odds:

$$\theta_1 = \frac{p_{(y=0)}}{p_{(y>0)}} \quad (1)$$

$$\theta_2 = \frac{p_{(y\leq1)}}{p_{(y>1)}} \quad (2)$$

For the last category of the crashes variable, the odds (for $y\leq2$) is equal to unity.

The prior odds (1) and (2) can be restated as:

$$\theta_j = \frac{p_{(y\leq j)}}{p_{(y>j)}}$$

or equivalently,

$$\theta_j = \frac{p_{(y\leq j)}}{1\text{-}p_{(y\leq j)}}$$

In the general case, where the dependent variable is ordered with m>2 categories, m-1 non-redundant logits can be estimated linearly. In the simple case of a univariate model with a single independent variable X, the logistic equation is slightly modified as:

$$ln(\theta_j) = a_j - bX, j=1,2\ldots m\text{-}1$$

This model has a negative slope to align its interpretation with that of linear models.[13] In linear models, the response variable covaries positively with the predictors. In ordinal regression, positive coefficients indicate that an increase in a predictor variable corresponds to increased odds of the event defined by $Y\leq j$. If the model had a positive slope, it would suggest that increasing X is associated with decreasing Y. To maintain interpretability consistent with other linear models, a negative sign is used. Alternatively, the ordinal model could be expressed with a positive slope:

$$ln\left(\frac{p_{(y>j)}}{1\text{-}p_{(y>j)}}\right) = a_j + bX, j = 1,2\ldots m\text{-}1$$

Here, the positive slope refers to the odds of $y>j$. Some statistical software packages use this formulation, where the exponential transformation of b corresponds to the odds of $y>j$. In such cases, we can compute its reciprocal to obtain the probability of $y\leq j$.[13,14]

The logits of the ordinal model are estimated by m-1 linear equations, differing only in their constant term a_j, while the coefficients of the independent variables remain the same across all logits. This means that the effect of each independent variable is assumed to be consistent across all levels of the response variable, a crucial assumption that must be tested when using ordinal regression. Due to this property, the ordinal regression model is also known as the proportional odds model (POM). The constant term of each logit, known as the threshold, is typically not of particular interest when interpreting results. Its role is analogous to the intercept in linear regression.

The ordinal regression coefficients are estimated using the maximum likelihood method, as in other logistic models. The method estimates regression coefficients and tests them using Wald's statistic. Additionally, the coefficients and overall model fit can be evaluated via the likelihood ratio test. To interpret the coefficients in terms of odds, they must be exponentiated. Some software, such as Statistical Package for Social Sciences (SPSS), provides regression coefficients without this transformation, requiring the user to compute e^{bk} and the corresponding confidence intervals.

According to the model definition, each coefficient e^{bk} represents the change in cumulative odds for $Y \leq j$ for each unit increase in X_k. For categorical independent variables, e^{bk} defines the cumulative odds ratio (CumOR) for each category relative to the reference category.

2. ORDINAL LOGISTIC REGRESSION VERSUS MULTINOMIAL

2.1. Car crashes data

The ordinal logistic model has greater power than the multinomial model by providing a single estimation of the odds for an ordinal variable without requiring multiple comparisons of individual logits.[6,15–18] We will compare it with multinomial logistic regression using field data on the effects of sleepiness, fatigue, and sleep disorders on car crashes.[19]

The study involved 1,366 drivers who were surveyed about incidents of falling asleep and experiencing fatigue while driving during the year of the field survey. They were also asked about potential sleep disturbances and the frequency of daytime sleepiness, as measured by the Epworth Sleepiness Scale.[20] Participants were given a self-administered questionnaire and asked to provide informed consent before participating.

The questionnaire included socio-demographic variables, such as gender, age, and educational level, as well as driving-related characteristics (e.g., years of driving license possession, total distance driven since obtaining a license). The number of crashes each participant had caused since obtaining their license was recorded through an open-ended question.

The variables for falling asleep and fatigue were defined as follows: (a) Falling asleep while driving in the past 12 months: Participants reported their involvement in three different scenarios; near misses, fall-asleep incidents, and veering off the road due to sleepiness. A six-point scale (0=never to 5=very often) measured these incidents. (b) Fatigue while driving in the past 12 months: This was assessed based on the frequency of nine risky driving behaviors, such as failing to remember the last few kilometers, having incoherent thoughts while driving, struggling to keep their head up, and missing road signs. A six-point scale (0=never, 1=1–2 times, 2=3–4 times, 3=5–6 times, 4=7–8 times and 5=9+ times) was used. (c) Daytime sleepiness in the past 12 months: Evaluated using the eight-item Epworth Sleepiness Scale. Participants rated their likelihood of dozing or falling asleep in various situations (e.g., while reading, watching television, sitting in a public place). A scale from 0 (never) to 5 (9+ times) was used, with total scores ranging from 0 to 40. Higher scores indicated a greater tendency toward daytime sleepiness. (d) Sleep disorder symptoms in the past month: Seven sleep-related issues were assessed, such as difficulty falling asleep, waking up during the night or early morning, and feeling tired upon waking. A four-point scale (0=never, 1=one time per week, 2=2–3 times per week, and 3=4+ times per week) measured these symptoms.

The fatigue scale values (from 0=never to 5=9+ times) were summed to create a cumulative score ranging from 0 to 45, reflecting the overall frequency of fatigue incidents. The same summation was applied to the Epworth Sleepiness Scale (c), producing a total score from 0 to 40. The seven sleep disorder items were also summed, yielding a cumulative variable ranging from 0 to 21.

Due to the highly skewed distribution of fall-asleep incidents while driving (with very few values different from 0), this variable was recoded into a binary format: 0=no incident and 1=at least one incident in the past 12 months.

2.2. Comparison of ordinal and multinomial logistic models

Data analysis was initially conducted using an ordinal logistic regression model, with the number of crashes as the response variable. Predictor variables included drivers' sex, years of license possession, age, alcohol consumption (glasses per week), number of fall-asleep incidents (binary: yes/no), and the three cumulative scales for fatigue, daytime sleepiness, and sleep disorders.

Given the highly skewed distribution of crash counts (90% of values fell within the 0–3 range), the variable was categorized into an ordinal scale: 0=no crashes, 1=up to two crashes, 2=three or more crashes. To compare the ordinal model with a multinomial logistic regression model, we reanalyzed the data using the same predictor variables. All statistical analyses were performed using SPSS, version 28.0.

At first, sleep-related variables were examined graphically to visualize their effects on the number of crashes through cumulative frequency plots. For instance, figure 1 presents the cumulative plot for the variable "falling asleep while driving". The plot indicates that at lower crash rates, the cumulative frequency of drivers with no sleep incidents is consistently higher than that of drivers who had at least

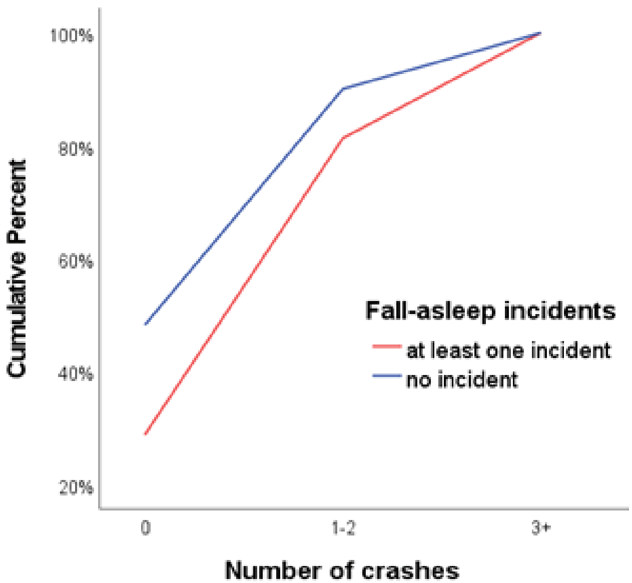

Figure 1. Plot of fall-asleep variable.

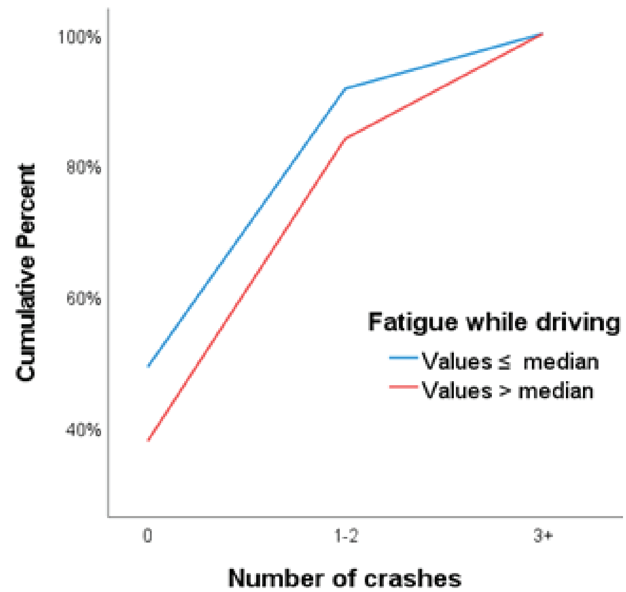

Figure 2. Plot of fatigue variable.

one incident. This difference decreases gradually until the two lines converge at 100% of observations. Because drivers who reported falling asleep while driving tended to have lower frequencies in the lower crash categories and higher frequencies in the upper crash categories, we expect the regression coefficient for this variable to be positive. Similar patterns emerged for the fatigue, sleep disorder, and daytime sleepiness variables, which were recoded as binary variables (above or below the median) for graphical representation. Figures 2 and 3 depict the cumulative frequency plots for fatigue and sleep disorders, respectively.

Cumulative frequency plots are an essential preliminary step in the analysis, particularly for the primary variables of interest. They serve as an introductory criterion for employing ordinal logistic regression by providing a descriptive validation of the proportional odds assumption. Ideally, the plots should resemble those in figures 1, 2, and 3, without any intersection of the two lines except at their final convergence at 100% of observations. Beyond this descriptive approach, it is necessary to formally test the proportional odds assumption using the parallel lines test.[21]

Table 1 evaluates the ordinal logistic model using the likelihood ratio statistic. The response variable of the model is the number of crashes, while the predictor variables include drivers' sex, years of license possession, drivers' age, alcohol consumption in glasses per week, sleepiness while driving, fatigue while driving, frequency of daytime sleepiness, and sleep disorders. Based on the likelihood ratio test, the null hypothesis that all model coefficients

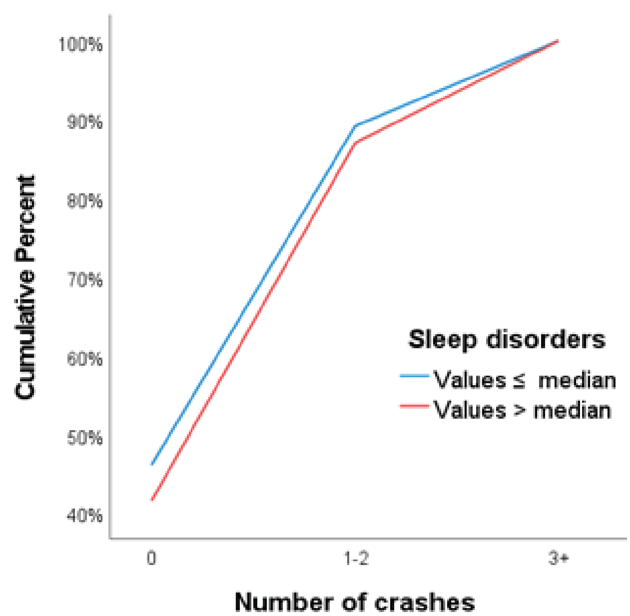

Figure 3. Plot of sleep disorders.

are equal to zero is rejected ($p<0.001$). Table 1 also evaluates the multinomial logistic model, and the same null hypothesis is rejected in this case as well.

Table 2 assesses the proportional odds assumption through the test of parallel lines. The null hypothesis, which states that the slopes (location parameters) are equal across all logits, is not rejected ($p=0.363$). Since this assumption holds, the ordinal logistic regression model is deemed ap-

Table 1. Ordinal and multinomial model fitting with the number of crashes as response variable.

Model	Model fitting information			
	-2LL	Chi-square	df	p value
Ordinal				
Intercept only	2,615.6	107.9	8	<0.001
Final model	2,508.5			
Multinomial				
Intercept only	2,615.6	115.1	16	<0.001
Final model	2,500.6			

Table 2. Parallel lines test for the ordinal model.

	Test of parallel lines			
	-2LL	Chi-square	df	p value
Null hypothesis	2,508.5	8.8	8	0.363
General	2,499.8			

Table 3. Cumulative odds ratios derived from the ordinal logistic regression with the number of crashes as response variable.

	Ordinal logistic model			
	B	CumOR	95% CI	p value
Threshold				
Crashes: 0	0.219	1.24	0.74 – 2.09	0.409
Crashes: 1–2	2.608	13.57	7.88 – 23.36	<0.001
Location				
Age	-0.006	0.99	0.97 – 1.01	0.494
Sex (men versus women)	-0.006	0.99	0.79 – 1.25	0.958
Years of license	0.035	1.04	0.97 – 1.61	0.004
Alcohol consumption glasses/week	0.011	1.01	0.99 – 1.03	0.141
Fatigue Scale	0.065	1.07	1.04 – 1.09	<0.001
Day sleepiness	-0.038	0.96	0.94 – 0.98	<0.001
Sleep disorders	0.021	1.02	0.99 – 1.05	0.138
Fall-asleep incidents (yes versus no)	0.548	1.73	1.32 – 2.28	<0.001

95% CI: 95% confidence interval

propriate for the data. If the assumption had been violated, a multinomial logistic regression or an alternative ordinal model would be considered.

Table 3 presents the coefficients for the ordinal logistic model, where the logits differ only in threshold values. The estimated regression coefficients (location parameters) remain the same across thresholds, as indicated by the parallel lines assumption. The model's parameters suggest that sleep-related variables significantly affect crash likelihood. Specifically, fatigue and falling asleep while driving are associated with a higher probability of crashes, whereas short daytime naps appear to have a protective effect.

From the exponential transformation of the model parameters, drivers who reported incidents of sleepiness had, on average, a 73% higher likelihood of experiencing more crashes compared to those without such incidents (CumOR=1.73, 95% confidence interval [CI]: 1.32–2.28, p<0.001). Additionally, each unit increase in the Cumulative Fatigue Scale was associated with a 7% increase in the odds of more crashes (CumOR=1.07, 95% CI: 1.04–1.09, p<0.001). Conversely, each unit increase in the Daytime Sleepiness Scale corresponded to a 4% (1–0.96) reduction in the odds of more crashes (CumOR=0.96, 95% CI: 0.94–0.98, p<0.001).

The odds ratios obtained from the ordinal model align with those from a multinomial logistic model. However, ordinal regression provides estimates across all logits under the proportional odds assumption. To compare the two approaches, we reanalyzed the data using a multinomial logistic model.

Table 4 presents the parameters of the multinomial logistic model. The first section of the table refers to the odds of experiencing at most two crashes (1–2 crashes) versus none. The results indicate that increased years of driving, fatigue, and sleep-related variables significantly contribute to crash risk. Each unit increase in the Fatigue Scale was associated with a 5% increase in collision odds (OR=1.05, 95% CI: 1.02–1.09, p<0.001). Furthermore, drivers who experienced at least one fall-asleep incident in the past year had a 77% higher crash risk (OR=1.77, 95% CI: 1.29–2.45, p<0.001). An increase in sleep disorders also increased the likelihood of collisions by 4% (OR=1.04, 95% CI: 1.01–1.07, p=0.019), whereas frequent short naps were associated with a 5% reduction in crash odds (OR=0.95, 95% CI: 0.93–0.98, p<0.001).

The second section of table 4 presents odds ratios for a higher number of crashes (≥3 crashes) versus none. The main risk factors remain fatigue and sleep incidents. A one-unit increase in the Fatigue Scale was associated with an 11% increase in the odds of collisions (OR=1.11, 95% CI: 1.06–1.15, p<0.001). Additionally, drivers who experienced at least one fall-asleep episode in the past year had double the crash risk compared to those without such incidents (OR=2.08, 95% CI: 1.33–3.26, p<0.001). In contrast, sleep

Table 4. Odds ratios (OR) derived from the multinomial logistic regression with the number of crashes as response variable (0 crashes=reference category).

	Multinomial logistic model			
Number of crashes	**B**	**OR**	**95% CI**	**p value**
1–2 crashes				
Age	-0.011	0.99	0.97 – 1.01	0.305
Sex (men versus women)	-0.002	0.99	0.77 – 1.29	0.988
Years of license	0.030	1.03	1.00 – 1.06	0.025
Alcohol consumption glasses/week	0.011	1.01	0.99 – 1.03	0.241
Fatigue Scale	0.052	1.05	1.02 – 1.09	<0.001
Day sleepiness	-0.048	0.95	0.93 – 0.98	<0.001
Sleep disorders	0.038	1.04	1.01 – 1.07	0.019
Fall-asleep incidents (yes versus no)	0.573	1.77	1.29 – 2.45	<0.001
3+ crashes				
Age	-0.006	0.99	0.96 – 1.03	0.729
Sex (men versus women)	-0.031	0.97	0.64 – 1.47	0.882
Years of license	0.051	1.05	1.01 – 1.10	0.014
Alcohol consumption glasses/week	0.021	1.02	0.99 – 1.05	0.081
Fatigue Scale	0.100	1.11	1.06 – 1.15	<0.001
Day sleepiness	-0.036	0.97	0.93 – 0.99	0.035
Sleep disorders	0.009	1.01	0.96 – 1.06	0.712
Fall-asleep incidents (yes versus no)	0.733	2.08	1.33 – 3.26	<0.001

95% CI: 95% confidence interval

disorders did not significantly impact the risk of multiple crashes, whereas short naps exhibited a protective effect (OR=0.97, 95% CI: 0.93–0.99, p=0.035).

3. DISCUSSION

Ordinal logistic regression, assuming the proportional odds assumption holds, estimates the cumulative odds for ordinal outcomes without assessing individual logits. In the example presented, the odds for the variable of collisions were estimated with a single value for each predictor variable in the ordinal model. Numerically, each regression coefficient represents the average change in log-odds for increasing levels of collisions. This applies to each unit increase in a predictor variable, assuming all other variables are held constant.

Furthermore, compared to the multiple logits estimated by multinomial regression, the coefficients of the ordinal model summarize changes in collision risk in a manner that is both concise and consistent. For example, the average change across all levels of crashes for a one-point increase in the Fatigue Scale is approximately 7% according to the ordinal model. This is comparable to the average change in the odds of the two logits estimated by multinomial regression: 5% for the logit comparing 1–2 crashes versus none and 11% for the logit comparing three or more crashes. The same pattern applies to the coefficients of other predictor variables in both models, except for the variable of sleep disorders. In the multinomial model, significant differences were found for sleep disorders in the logit comparing 1–2 crashes (p=0.019), but not in the logit for three or more crashes (p=0.712). In contrast, the ordinal model, which summarizes the average change across all levels of the collision variable, did not detect significant differences (p=0.138).

In our example, ordinal and multinomial regression yielded similar results, including comparable predicted probabilities for the number of collisions. Both models are reasonable choices when dealing with an ordinal dependent variable. The advantage of using ordinal regression over multinomial regression is that it increases the power of the analysis by incorporating the ordering of the dependent variable into the model. In this way, a single slope/odds ratio is estimated for all categories of the ordinal variable instead of multiple logits for smaller subsets of the data.[6] However, a limitation of ordinal logistic regression is that it estimates an average effect of the predictor variables across all outcome levels, meaning that if the effects vary across levels, the model will fail to capture this variation. The validity of this approach depends on satisfying the proportional odds assumption; otherwise, the use of the ordinal model may lead to incorrect results.

When the proportional odds assumption does not hold, multinomial logistic regression serves as an alternative. However, it requires estimating a larger number of parameters than the ordinal model, which increases the number of degrees of freedom needed for model fitting and can place excessive demands on the dataset. Additionally, interpreting multinomial regression results is more complex and less concise than interpreting ordinal regression results.[9,17]

Another limitation of multinomial regression is that it does not account for the ordinal nature of the response variable. As a result, its statistical power in detecting associations with explanatory variables is suboptimal, particularly when those associations follow a linear trend (either positive or negative).[17] In extreme cases, an explanatory variable may be linearly associated with the ordinal outcome, with small but consistent changes in odds across outcome levels. In such cases, individual logits in the multinomial model may fail to detect significant differences, whereas cumulative odds in the ordinal model may show significance. Unlike multinomial regression, ordinal regression is specifically designed for response variables with an ordinal structure, as it works with cumulative distributions and estimates parameters that represent the overall trend across ordinal values.

In cases where the response variable is ordinal, but the proportional odds assumption does not hold for all predictor variables, either multinomial regression or the partial proportional odds model (PPOM) can be used.[22–26] The PPOM allows some predictor variables to satisfy the proportional odds assumption while introducing special parameters for those that do not, which vary across different outcome categories. For example, if explanatory variables X_i and X_j satisfy the proportional odds assumption, their coefficients (b_i and b_j) remain the same across all levels of the response variable. In contrast, if variable X_z does not satisfy the proportional odds assumption, its coefficient (b_{zj}) can differ across levels (j) of the response variable.[26,27] The interpretation of PPOM coefficients is similar to that of multinomial and proportional odds regression coefficients.

The PPOM is a generalized form of ordinal regression and can be implemented in SAS using the LOGISTIC procedure[28] and in Stata using the GOLOGIT2[29] module. However, SPSS does not currently support this model.

4. CONCLUSION

The extensive use of logistic models in data analysis requires their application to be carefully justified based on the measurement level of the response variable and the assumptions that must hold. Binary logistic regression serves as the foundation for these models but is inadequate for response variables with multiple categories, especially when those categories are ordered. In such cases, as illustrated by the example of car crashes, ordinal logistic regression is recommended, provided the proportional odds assumption holds. Otherwise, multinomial regression or the PPOM should be considered.

The PPOM extends the POM to handle predictor variables that violate the proportional odds assumption. In these models, variables that meet the assumption are treated within the framework of ordinal regression, while those that do not are analyzed using a multinomial approach.

ΠΕΡΙΛΗΨΗ

Η διατακτική λογαριθμική παλινδρόμηση ως μέθοδος για την εκτίμηση του κινδύνου σε διαβαθμιζόμενα αποτελέσματα

Χ. ΓΝΑΡΔΕΛΛΗΣ,[1] Β. ΝΟΤΑΡΑ,[2] Γ. ΤΖΑΜΑΛΟΥΚΑ,[3] Μ. ΠΑΠΑΔΑΚΑΚΗ,[3] Ι. ΧΛΙΑΟΥΤΑΚΗΣ[3]

[1]Τμήμα Αλιείας και Υδατοκαλλιεργειών, Σχολή Γεωπονικών Επιστημών, Πανεπιστήμιο Πατρών, Μεσολόγγι, [2]Τμήμα Δημόσιας και Κοινοτικής Υγείας, Σχολή Δημόσιας Υγείας, Πανεπιστήμιο Δυτικής Αττικής, Αθήνα, [3]Τμήμα Κοινωνικής Εργασίας, Σχολή Επαγγελμάτων Υγείας, Ελληνικό Μεσογειακό Πανεπιστήμιο, Ηράκλειο, Κρήτη

Αρχεία Ελληνικής Ιατρικής 2026, 43(5):682–690

Η χρήση των υποδειγμάτων λογαριθμικής παλινδρόμησης στην ανάλυση δεδομένων και στη μηχανική μάθηση έχει διευρυνθεί σημαντικά τα τελευταία έτη, αποτελώντας την κύρια προτίμηση των ερευνητών για την εκτίμηση του κινδύνου σε ένα μεγάλο εύρος επιστημονικών πεδίων: Από τον προσδιορισμό των αιτιολογικών παραγόντων χρόνιων νοσημάτων μέχρι την εκτίμηση του κινδύνου σε τροχαία ατυχήματα ή τον υπολογισμό του πιστωτικού κινδύνου σε χρηματοπιστωτικές διαδικασίες. Όλα τα λογαριθμικά υποδείγματα είναι φυσικές προεκτάσεις του δίτιμου λογαριθμικού υποδείγματος και η ερμηνεία των αποτελεσμάτων τους βασίζεται σε αυτό. Χρησιμοποιώντας τα δεδομένα μιας συγχρονικής μελέτης με θέμα την υπνηλία κατά την οδήγηση και τον κίνδυνο τροχαίων ατυχημάτων, τα δύο βασικά διευρυμένα υποδείγματα των λογαριθμικών μεθόδων, η πολυωνυμική (πολύτομη) και η διατακτική λογαριθμική παλινδρόμηση, παρουσιάζονται λεπτομερώς. Η έμφαση δίνεται στη διατακτική λογαριθμική παλινδρόμηση, εφόσον η μεταβλητή έκβασης των δεδομένων των τροχαίων ατυχημάτων της μελέτης εκφράζεται ως διατεταγμένο μέγεθος, ορίζοντας μια λανθάνουσα συνεχή κλίμακα.

Λέξεις ευρετηρίου: Δεδομένα τροχαίων ατυχημάτων, Διατακτική λογαριθμική παλινδρόμηση, Ελλιπές μοντέλο αναλογικών πιθανοτήτων, Κίνδυνος συγκρούσεων, Μοντέλο αναλογικών πιθανοτήτων, Πολυωνυμική λογαριθμική παλινδρόμηση

References

1. HOSMER DW Jr, LEMESHOW S, STURDIVANT RX. *Applied logistic regression.* 3rd ed. John Wiley & Sons, New York, 2013:269–311
2. POSTON DL Jr, CONDE E, FIELD LM. *Applied regression models in the social sciences.* Cambridge University Press, Cambridge, 2023:339–388
3. PAMPEL FC. *Logistic regression: A primer.* 2nd ed. Sage Publications, Thousand Oaks, CA, 2021:81–108
4. AGRESTI A. *Categorical data analysis.* 3rd ed. John Wiley & Sons, Hoboken, New Jersey, 2013:293–313
5. KWAK C, CLAYTON-MATTHEWS A. Multinomial logistic regression. *Nurs Res* 2002, 51:404–410
6. SAINANI KL. Multinomial and ordinal logistic regression. *PM R* 2021, 13:1050–1055
7. PARK KH, KERR PM. Determinants of academic performance: A multinomial logit approach. *J Econ Educ* 1990, 21:101–111
8. HIJOL NBJ, SALMA N, SARKER I. Assessing the mental health of slum dwellers: An ordinal logistic approach. *J Health Popul Nutr* 2024, 43:76
9. FULLERTON AS, ANDERSON KF. Ordered regression models: A tutorial. *Prev Sci* 2023, 24:431–443
10. LI L, RYSAVY MA, BOBASHEV G, DAS A. Comparing methods for risk prediction of multicategory outcomes: Dichotomized logistic regression versus multinomial logit regression. *BMC Med Res Methodol* 2024, 24:261
11. GEHRINGER CK, MARTIN GP, VAN CALSTER B, HYRICH KL, VERSTAPPEN SMM, SERGEANT JC. How to develop, validate, and update clinical prediction models using multinomial logistic regression. *J Clin Epidemiol* 2024, 174:111481
12. ANDERSON JA. Regression and ordered categorical variables. *J R Statist Soc B* 1984, 46:1–30
13. ANDERSON JA, PHILIPS PR. Regression, discrimination and measurement models for ordered categorical variables. *Appl Statist* 1981, 30:22–31
14. CAMPBELL MK, DONNER A. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression. *J Am Stat Assoc* 1989, 84:587–591
15. BENDER R, GROUVEN U. Ordinal logistic regression in medical research. *J R Coll Physicians Lond* 1997, 31:546–551
16. FRENCH B, SHOTWELL MS. Regression models for ordinal outcomes. *JAMA* 2022, 328:772–773
17. WARNER P. Ordinal logistic regression. *J Fam Plann Reprod Health Care* 2008, 34:169–170
18. KOLETSI D, PANDIS N. Ordinal logistic regression. *Am J Orthod Dentofacial Orthop* 2018, 153:157–158
19. GNARDELLIS C, TZAMALOUKA G, PAPADAKAKI M, CHLIAOUTAKIS JE. An investigation of the effect of sleepiness, drowsy driving, and lifestyle on vehicle crashes. *Transp Res Part F: Traffic Psychol Behav* 2008, 11:270–281
20. JOHNS MW. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep* 1991, 14:540–545

21. BRANT R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 1990, 46:1171–1178
22. PETERSON B, HARRELL FE Jr. Partial proportional odds models for ordinal response variables. *Appl Statist* 1990, 39:205–217
23. SINGH V, DWIVEDI SN, DEO SVS. Ordinal logistic regression model describing factors associated with extent of nodal involvement in oral cancer patients and its prospective validation. *BMC Med Res Methodol* 2020, 20:95
24. LELISHO ME, WOGI AA, TAREKE SA. Ordinal logistic regression analysis in determining factors associated with socioeconomic status of household in Tepi Town, Southwest Ethiopia. *ScientificWorldJournal* 2022, 2022:2415692
25. MOORADIAN J, IVAN JN, RAVISHANKER N, HU S. Analysis of driver and passenger crash injury severity using partial proportional odds models. *Accid Anal Prev* 2013, 58:53–58
26. SASIDHARANA L, MENÉNDEZ M. Partial proportional odds model – an alternate choice for analyzing pedestrian crash injury severities. *Accid Anal Prev* 2014, 72:330–340
27. KURNIAWATI Y, KURNIA A, SADIK K. A comparison of polytomous model with proportional odds and non-proportional odds model on birth size case in Indonesia. *Media Statistika* 2021, 14:79–88
28. FLOM P. Multinomial and ordinal logistic regression using PROC LOGISTIC. 23rd Annual Conference of the North East SAS Users Group (NESUG), Baltimore, Maryland, 2010:1–12
29. MEHMETOGLU M, JACOBSEN TG. *Applying statistics using Stata: A guide for the social sciences.* Sage Publications, London, 2017:161–192

Corresponding author:

C. Gnardellis, 91–93 Markou Mousourou street, 116 36 Athens, Greece
e-mail: hgnardellis@upatras.gr