Research on applications of Generalised Linear modelling techniques

***Submitted By,***

***Muni Ranjan (A0163382E)***

***Pradeep Kumar (A0163453H)***

***Sambit Kumar Panigrahi (A0163285B)***

# Poisson-Based Regression Analysis of Aggregate Crime Rates

**Summary:**

This paper applies regression model based on Poisson distribution to solve problems arising from analyzing aggregate crime rates used for small populations and low base-rates. The units of the sample are aggregations of individuals, such as neighborhoods, cities, and schools. Crimes rate is often reported as crimes per 100,000 population.

As an illustration of the study, aggregate crime rates with an analysis of rates of juvenile violence in Florida, Georgia, South Carolina, and Nebraska, having total populations ranging from 560 to 98,000 is considered. The measure of offending is the number of juveniles (ages 11 through 17) arrested for robberies in each county, pooled over the period of 1989 through 1993. The explanatory variables are:

* Residential instability: proportion of households occupied by persons who had moved from another dwelling in the previous 5 years.
* Ethnic heterogeneity: proportion of households occupied by white versus nonwhite persons.
* Family disruption, indexed by female-headed households: proportion of all households with children.
* Poverty: proportion of persons living below the poverty level.
* Unemployment rate (coded as proportion of the workforce).
* Proximity to metropolitan counties, 1 being adjacent to a metropolitan and 0 being nonadjacent.

Note: As the population base increases, the crime rates decreases, even though the range of crime counts increases. Therefore, the basic Poisson regression model is altered by adding natural logarithm of the size of the population at risk so that it provides an analysis of per capita crime rates rather than counts of crimes.



Ships\_damage\_Data: <http://data.princeton.edu/wws509/datasets/#ship>

# Robust Inference in the Negative Binomial Regression Model with an Application to Falls Data

**Summary:**

This paper aims at analyzing and reducing the occurrence of falls in elderly patients, or patients with Parkinson’s disease. The main outcome is the number of falls reported during a specified period of time, and covariates includes - an intervention effect, mobility-related information such as gait speed, and a history of the individual’s falls.

The study was conducted between 2008 and 2012 in the metropolitan area of Sydney, Australia, and aimed at evaluating the effectiveness of an exercise intervention at reducing the number of falls among elder people suffering from Parkinson’s disease, from which follows the name of the study: Parkinson’s disease Frailty Intervention Trial (PD-FIT). This article uses two approaches for building robust M-estimators of the regression by exploring different choices for the bounding functions. The response variable is the reported number of falls during the intervention period. Covariates are: pastf, the reported number of falls during the past 12 months; and severity, a score indicating the severity of the disease measured at baseline with the Unified Parkinson’s Disease Rating Scale.

Through the analysis of falls data from a randomized control trial, the usefulness of these robust estimators for safe inference and as diagnostic tools is highlighted.



# A multinomial logit approach to exchange rate policy classification with an application to growth

**Summary:**

In this paper, a country’s exchange rate policy is modeled using multinomial logit response conditioned on various factors. The category with the highest predictive probability implied by the logit regression serves as de facto exchange rate policy. However, many observers noted that the de jure currency management for some countries seemed unsatisfactory for assessing the role of exchange rate stability motivating researchers to propose classifications that are based on the behavior of the parallel market exchange rates on the grounds that they better reflect underlying market and monetary conditions than the country's official exchange rates. The two explanatory variables are volatility of a bilateral nominal exchange rate against an anchor currency and the volatility of international reserves.

The task was to classify country exchange rate policy according to the six de jure categories arranged in order of increasing stability as follows:

1) Independently Floating

2) Managed Floating

3) Adjusted According to a Set of Indicators

4) Cooperative Arrangements

5) Limited Flexibility

6) Currency Peg

This analysis selects the most favorable de jure category and accordingly helps in effective formation of national policy towards exchange rate management.



# Multilevel ordinal models for examination grades

**Summary:**

Schools use either point score by subject or summative scores to provide a total points for each candidate in assessing student achievement and they are used in this form by the Government in the production of ‘performance indicators’. One of the drawbacks to the use of scores is that information is lost about the actual distribution of grades in particular subjects when inferences are made at the level of the school.

This paper develops explanatory models for the actual grades and compares these with the standard point scoring system. The aim is to gain additional insights from using the former as opposed to the latter models. This is done using multilevel models that recognize the essentially hierarchical nature of examination data with students nested within schools. For the point scoring system standard normal theory models are applied, while for the grades less well-known ordered categorical response models are used. A technical advantage of the latter models lies in the fact that they do not require strong scaling assumptions, but merely the existence of an ordering. The six categories of response A–F are multilevel ordinal models for examination grades denoted by integer labels 1, 2, 3, 4, 5, 6 using glm with cumulative probabilities associated with responses as dependent. A monotonic ‘link’ transformation of a set of cumulative probabilities on the [0, 1] scale to the real line facilitates effects to operate in a linear and additive fashion.



Dataset Link: <http://www.ats.ucla.edu/stat/data/ologit.dta>

# The use of conjoint analysis in the design of bicycle related clothing

**Summary:**

This study uses conjoint analysis in bicycle-related clothing design to provide designers with an efficient approach to identify customer preferences. To satisfy the physiological functions of the human body, comfort, functionality, good appearance, and materials, the care and management of clothing and consumer lifestyles have to be taken into account for the design of professional clothing for bicycling-related sports. ‘AIO’ measurement has been used to study the segmentation of consumer lifestyles, in which the letter ‘‘A’’ stands for ‘‘Activities’’, ‘‘I’’ for ‘‘Interests’’, and ‘‘O’’ for ‘‘Opinions’’. The intention is to identify the set of preference factors regarding products’ popularity in the market that provide designers with a frame of reference as they design new textile products that match the preferences of specific consumer groups. The questionnaire used in stage one adopts the AIO measurement chart to gauge attitudes in conjunction with a 9-level Likert Scale that can be applied to the development of functional clothing and other innovative products. In stage two, factor analysis is used to categorize customer requirements. The five clusters formed are described as:

Cluster 1 is of the adjustable type; cluster 2, comfortable and noticeable; cluster 3, comfortable and healthy; cluster 4, adjustable and style and brand name; cluster 5, special function and noticeable. Finally, conjoint analysis is used in this study to evaluate and measure customer preference attitudes toward functional textile products to assemble an optimal design solution.



Dataset Link: www.dobney.com/Conjoint/ExcelBasedConjointDemo.xls