**Team Name**: MASK

**Team Members**:

1. Anusuya Manickavasagam (A0163300Y)
2. Chinnasubbareddygari Mohan Reddy (A0163433L)
3. Sridharan Kesavan (A0163207M)
4. **Title:** **Poisson Regression Model for Count Data**

**Description:** The study investigates the factors affecting the female crabs having males which are called as satellites. The female crab’s color(C), Spine condition (S),weight (Wt) and carapace width (W) were made as the explanatory variables and the number of satellites were made as the response variable.

**Paper Link and Paper Date**

<https://www.researchgate.net/publication/229588257_Satellite_Male_Groups_in_Horseshoe_Crabs_Limulus_polyphemus>

The paper is used in below Poisson model

<https://onlinecourses.science.psu.edu/stat504/node/223>

Date of publication: April 2010

**Solution:** Poisson regression model was fitted against the data and the following steps were followed to improve the model fit.

1) adjusting for Over dispersion by scale using Pearson.

2) Including 'color' as a qualitive and numerical predictor

3)Fitting the Poisson model.

The width (W) was selected as the best predictor variable and the Poisson model was fitted. The value of β > 0, giving the indication the wider the female crab the greater expected number of male satellites on the multiplicative order as exp(0.1640) = 1.18. More specifically for one unit increase in the width, the number of Sa will increase and it will be multiplied by 1.18.

**Datafile link:**

<https://onlinecourses.science.psu.edu/stat504/sites/onlinecourses.science.psu.edu.stat504/files/lesson07/crab.txt>

**2) Title:** **Negative Binominal Regression**

**Description:**

The paper analyses the survey of 1308 people, how many homicide victims they knew within the past 12 months. The response variable is the number of victims the respondent knows, and the predictor variable is the race of the respondent (black or white). First, Poisson regression was applied. On analyzing the results it was found that the variability of counts was higher for both races(predictor variable). Hence, Negative binomial model was applied.

**Paper Link:** <http://data.library.virginia.edu/getting-started-with-negative-binomial-regression-modeling/>

**Paper Date:** May 5,2015

**Solution:** Negative Binomial Model was fitted against the data and the following steps were followed.

1)Exponentiate the race coefficient to get a ratio of sample means and make predictions to get original sample means.

2)Observing the value of Theta(dispersion parameter) and use it to get the estimated variance for counts.

3)Use model parameters to simulate data from the negative binomial distribution.

When we plot the model using rootogram we see that it looks better than the Poisson Model. Even, the simulated data is very similar to the observed data giving us the confidence that choosing Negative Binomial Model is better than the Poisson Model.

**Datafile link:** The original data was collected from

<http://gss.norc.org/documents/codebook/GSS_Codebook.pdf>

The cleaned data was used in the book, Categorical Data Analysis, by Alan Agresti (2002) <https://mathdept.iut.ac.ir/sites/mathdept.iut.ac.ir/files/AGRESTI.PDF>

**3) Title:** **Conjoint Analysis**

**Description:** In this paper, conjoint analysis is used to identify the various components of a treatment in survey experiments. There are two empirical examples that were used in the paper. One is the US citizens preference across presidential candidates which were conducted in 2012 which has attributes like religion, college education, profession, annual income, racial/ethnic background, age, military service and gender and the other is conjoint experiment on attitude towards immigrants which has variables like gender, education level, employment plans, job experience, profession, language skills, country of origin, reasons for applying and prior trips to the United States.

**Paper Link:** <http://web.mit.edu/teppei/www/research/conjoint.pdf>

**Paper Date:** 19 December,2013

**Solution:** Conjoint Analysis was done on both the examples and the results were observed.

1)The candidate experiment- Eight attributes of would-be presidential candidates were considered and respondents were given a survey with multiple questions (which were used as dependent variable) where they were asked to choose between two candidates.

2)The immigrant experiment- In this experiment, Knowledge Networks were asked to act as immigration officials and decide which pair of immigrants they would choose for admission. They were given a pair of profiles with a set of attributes and then asked to rate each profile on a scale of one to seven.

The following assumptions were used while conducting Conjoint Analysis experiments:

1)Stability and No carryover effects

2)No-profile-order-effects

3)Randomization of profiles

4)Conditionally independent Randomization

5)Completely independent Randomization

**Datafile link:** <https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/22603>

**4) Title:** **Multinomial Regression Ordinal Response**

**Description:** Respondents were asked to evaluate the statement that a working mother can establish warm and secure relationship with a child as a mother who does not work. Responses were coded as 1 to 4 with 1 having strongly agree and 4 having strongly disagree. The ordered logit model was followed and the results were noted.

**Paper Link:**

1. <https://www3.nd.edu/~rwilliam/gologit2/MSS2010-Handout.pdf>

**Paper Date:** December 2014

**Solution:** The following Proportional Odds & Partial Proportional Odds/ Parallel Lines & Non-Parallel Lines models were followed and the responses were noted.

1. Model 0: Perfect Proportional Odds/ Parallel Lines- Proportional Odds works perfectly in this model and the Betas are all the same.
2. Model 1: Partial Proportional Odds I-Gender has greatest effect at the lowest levels of attitudes, i.e women are much less likely to strongly disagree than men are. The effect of gender is consistently positive.
3. Model 2: Partial Proportional Odds II-Gender has its greatest -and only -effect at the lowest levels of attitudes, i.e. women are much less likely to strongly disagree than men are. The ologit estimate underestimates the effect of gender on the lower levels of attitudes and overestimates its effect at the higher levels.
4. Model 3: Partial Proportional Odds III-The effect of gender varies in both sign and magnitude across the range of attitudes. Use of ologit in this case would be highly misleading.

**Datafile link:** <http://www.indiana.edu/~jslsoc/stata/spex_data/ordwarm2.dta>

1. **Title:** **Multinomial Regression with nominal response**

**Description:** Vehicle following behavior in mixed traffic conditions. Following vehicle cannot strictly follow the lead vehicle due to weak lane discipline. The follower behavior can be 1) Car following, 2) staggered following and 3) following between 2 vehicles (dependent variables), we predict using independent variables such as size of follower and leader, speed of follower and leader, longitudinal gap between follower and leader, types of follower and leader. As the predictor is as multi-level categorical variable multi nominal regression should be used for this problem.

**Paper Link:** http://www.sciencedirect.com/science/article/pii/S235214651730844X

**Paper Date:** 10-15 July 2016

**Solution:** The multi nominal regression output gives leader size has higher impact for the following vehicles behavior, distance between follower and leader has minimal effect, If leader speed is higher than follower has more freedom to drive, car following heavy vehicles makes staggered car behavior.

**Datafile link:** http://toledo.net.technion.ac.il/mixed-traffic-trajectory-data/