**BUAN 6383**

**Modelling with Business Analytics**

**PART - I**

1. A) Poisson model

Graphical user interface, text, application

Description automatically generated

B) The NBD model

Shape parameter(n) = 0.9977

Alpha value = 0.2500

Log- likelihood = -1140.237

C) Zero-inflated nbd model

Pi value = 0.1131,

Shape parameter = 1.5039

Apha parameter = 0.3342

Log – likelihood = -1136.165

D) finite mixture model

2 – segment model

Pi1 = 0.2991

Pi2 = 0.701

Lamda1 = 9.121

Lamda2 = 1.802

Log – likelihood = -1188.833

3 – segment model

Pi1 = 0.543

Pi2 = 0.2767

Pi3 = 0.1799

Lamda1 = 3.483

Lamda2 = 0.2905

Lamda3 = 11.215

Log-likelihood = -1132.043

4 – segment model

Pi1 = 0.151

Pi2 = 0.2442

Pi3 = 0.101

Pi4 = 0.5027

Lamda1 = 7.419

Lamda2 = 0.2047

Lamda3 = 12.8737

Lamda4 = 3.022

Log-likelihood = -1130.0705

2) Based on log-likelihood we can see that 4- segment mixture model has the best log likelihood value followed by 3 segment mixture model with log-likelihood of -1132.043. But efficiently we can see that increasing one parameter that is increasing one segment is not increasing the log-likelihood to a considerable amount.

For calculations of AIC and BIC please check code.

In an overall comparison, we can see that poisson model significantly performed bad when compared to the other models. This indicates that the poisson model is not a good fit for the setup.

When comparing both AIC and BIC we could see that 3 segment mixture model is the best model since it has the least AIC and BIC values.

The addition of more variables might increase the log likelihood values but there might not be any explanatory power for these variables.

Both AIC and BIC penalizes the model for more number of parameters which is evident from the above values.

3) Based on the 2, 3, and 4-segment finite mixture models, how many packs are the following customers likely to

purchase over the next 8 weeks?

1. a customer who purchased 5 packs in the past week

2 – segment model

Expected purchases in 8 weeks: 42.780077179153444

3 -segment model

Expected purchases in 8 weeks: 30.82452037507946

4 -segment model

Expected purchases in 8 weeks: 33.674460564811405

1. a customer who purchased 9 packs in the past week.

2 – segment model

Expected purchases in 8 weeks: 72.8732122974245

3 -segment model

Expected purchases in 8 weeks: 80.06410100715745

4 -segment model

Expected purchases in 8 weeks: 69.32021752407665

**Part – II:**  
Q1. Estimate all relevant parameters for Poisson regression using MLE. Report your code, the estimated parameters and the maximum value of the log-likelihood. What are the managerial takeaways — which customer characteristics seem to be important?

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Married, Prestige and menpubs independent variables are the most important customer characteristic as they have positive effect on Beta. Kids and female variables have negative effect on Beta.

2.Estimate all relevant parameters for NBD Regression using MLE. Report your code, the estimated parameters and the maximum value of the log-likelihood. What are the managerial takeaways — which customer characteristics seem to be important?

Graphical user interface, text, application

Description automatically generated

Married, Prestige and menpubs independent variable are the most important customer characteristic as they have a greater effect with the positive value of beta.

3.Explain the logic used in developing the model in detail.

Report your code, the estimated parameters and the maximum value of the log-likelihood. What are the managerial takeaways — which customer characteristics seem to be important?

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Married independent variable is the most important customer characteristic as they have a greater effect with the highest value of beta.

4. Evaluate the models developed; explain which of them is best, and why. Are there any significant differences among the results from these models? If so, what exactly are these differences? Discuss what you believe could be causing the differences.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | LL value | AIC | BIC |
| Poisson Regression | -1651.05 | 3314.112 | 3343.025 |
| NBD Regression | -1560.95 | 3135.916 | 3169.649 |
| Zero Inflated NBD Regression | -1560.95 | 3137.916 | 3176.46 |

On comparing the LL, AIC and BIC values, we can observe that the NBD Regression has the highest LL value and lowest AIC and BIC values.

Overall, NBD regression performs the best for the given dataset.

Yes there are significant differences in results among these models. NBD Regression performs better than other Regression models. Also, the other reason might be due to heterogeneity. There is not a significant difference between NBD regression and Zero Inflated NBD regression.

Loglikelihood value assesses the fitness of models. If more variables are added to the model, the Loglikelihood value will increase but we can’t trust the variables to have significant explanatory power

AIC penalizes the model with more parameters by taking into consideration both the loglikelihood values and the no of parameters. The model with the lowest AIC is the Zero Inflated NBD Regression

BIC penalizes the model with more parameters and a higher number of observations. The NBD Regression has the lowest BIC.