The University of Texas at Dallas
School of Management

Syam Menon   **Advanced**
MIS 6334   **Business Intelligence**

Project 01

Objectives

- Integrate knowledge learned in the second half of the semester with SAS (Learn to build customized models)

- Hands-on experience for BI implementation using base SAS

Instructions

- **Due Date:** See Syllabus.

- Submit one report per group via eLearning.
  eLearning will stop accepting submissions after the due date, and late submissions will not be accepted.

- Report should be a single Microsoft Word document with your group number and all group member names clearly identified.

- Be as clear as possible. Vague answers — however long — will not receive full credit.

- A professional quality report is expected — messy or hard-to-read reports will be penalized.

- If a question requires you to write SAS code, **include the SAS code you created in the submission file (clearly identify which question the code is for). We should be able to run the codes you submit.**

- This project counts for 200 points.

Additional Files

- **Data Sets**

  - `billboard.sas7bdat`
    Billboard Exposures data set (needed for **Part I** of the project).
  - `kc.sas7bdat`
    Khaki Chinos data set (needed for **Part I** of the project).
  - `books.txt`
    This is the dataset to be analyzed in **Part II** of the project.

- **Sample Code**

  - `usefulSAScodesProject01.txt`
    This file contains some codes for **Part I** of the project. You can adapt them to complete **Part II**.

There are two parts to this project.

1. **Part I** familiarizes you with the use of SAS to solve the count models considered in class. You will use the codes provided in the file `usefulSAScodesProject01.txt` to learn how the NBD model, the Poisson regression model, and the NBD Regression Model can be implemented, for the **Billboard Exposures** and **Khaki Chinos** examples from class. `usefulSAScodesProject01.txt` contains SAS codes for all but the NBD model; you should understand and run them to jump-start your project. At the end of **Part I**, you should have good sample codes that can be modified and used for the analysis needed in **Part II**. **Part I** is worth 50 points.

2. **Part II** requires you to analyze a real data set using the advanced models from class, and present managerial insights gained from your analysis. The file `books.txt` provided with this project contains the data needed for **Part II**. **Part II** is worth 150 points.

Part I: Examples Integrating SAS and Advanced Modeling

1. **The NBD Model**
   Consider the *billboard exposures* example from class. Write SAS code and conduct maximum likelihood estimation (MLE) for the NBD Model; estimate r and α. Report your code and the estimated values. **When reporting MLE results, please provide the optimized LL value, all the estimated parameter values, and the corresponding p-values**. Other statistics are optional — you need report them only if you want to comment on them in some way. In addition, please add comments to your SAS code to make your code easy to understand.

2. **The Poisson Regression Model**
   Consider the *khakichinos.com* example from class. Write SAS code to estimate parameters (λ_0 and the vector β) using MLE for the Poisson Regression Model. Report your code and the estimated values. What are some managerial takeaways?

3. **The NBD Regression Model**
   Consider the *khakichinos.com* example again. Write SAS code to estimate parameters (r, α and the vector β) using MLE for NBD Regression Model. Report your code and the estimated values. What are some managerial takeaways? Explain the difference in results between the NBD and the Poisson Regression Model.

Part II: Analysis of New Real Data

In this part of the project, you will adapt the models you used in Part I and apply them to the dataset **books.txt**. The dataset records customer purchases at two competitors, *Amazon.com* and *BARNES & NOBLE* (B&N) in 2007. Some customer demographic variables — education, household size (hhsz), income, and race — are also in the dataset.

Suppose you are working for B&N and would like to understand the factors that affect customer purchasing behavior there. In particular, you are interested in the following questions:

1. How do you build a customized model to fit the existing data, and make predictions at the population-level?

2. What are the consumer characteristics that differentiate who purchase many books (say, ¿10 a year) from those who buy just a few?

3. Why do certain customers prefer Amazon over B&N?

Your objective is to leverage the modeling skills you learned from this class to answer these business questions. Specific steps are below.

1. Write a SAS program that reads the data in **books.txt** and generates a count dataset (similar to that used in the *khaki chinos* example). That is, for each customer count the number of books purchased from B&N in 2007, while keeping the demographic variables. Print the first 10 records of this dataset.

2. Build an NBD model, ignoring the demographic variables. Report your results. (Hint: you will need to create a data set similar to that used in the *billboard exposures* example.)

3. Calculate the values of (i) Reach, (ii) Average Frequency, and (iii) Gross Ratings Points (GRPs) based on the NBD Model. Show your work.

4. Build a Poisson regression model using the demographic information (customer characteristics) provided. Report your results. What are the managerial takeaways — which customer characteristics seem to be important?
   Optional: You have flexibility in choosing the variables to include — if you wish to do so, you can choose to eliminate some (via feature selection, for example) or create new ones (from the variables you have available — for example, fraction of weekend purchases). This is optional for this project, but if you do anything along these lines, please provide your justification.

5. Next, we start the setup for developing an NBD regression model. What is the formula for the log-liklihood expression, $\mathcal{LL}$?

6. Build a NBD regression model using the demographic information provided. Report your results. What are the managerial takeaways — which customer characteristics seem to be important?
   Optional: As with the Poisson regression, you have flexibility in choosing the variables to include — if you wish to do so, you can choose to eliminate some (via feature selection, for example) or create new ones (from the variables you have available — for example, fraction of weekend purchases). This is optional for this project, but if you do anything along these lines, please provide your justification.

7. Are there any significant differences between the results from the Poisson and NBD regressions? If so, what exactly is the difference? Discuss what you believe about the cause(s) of the difference.

8. Briefly summarize what you learned from this project. This is an open-ended question, so please include anything you found worthwhile — relating to the modeling tool (SAS), the modeling process, insights from the modeling, any managerial takeaways that were insightful to you, and so on.