MKTG/STAT 4760/7760
Spring 2026

Homework #2A

1. The HW2A dataset has a spreadsheet containing Internet visit data for 2728 Comscore panelists, showing the number of visits each person made to one website (with the disguised name of khakichinos.com). You should ignore the covariate data (we'll discuss/use it in a few weeks). Fit a plain NBD and an NBD with spike at zero to this dataset. Briefly discuss your results.

   Note that the data are not in a histogram format: we want you to run your models using the disaggregate, raw data as presented in the spreadsheet. This might seem like a nuisance, but sometimes you have to run NBD models this way, as we'll see in a few weeks when we bring the covariates into the model.[1]

   Remember (as discussed in class): while you'll use the raw data to run the models, you should right-censor the data (in an appropriate manner) to evaluate the goodness-of-fit.

2. The HW2A dataset contains a second sheet showing the distribution of the number of albums purchased across a set of 1574 shopping trips for an online music seller.

   a. Estimate a shifted NBD and a truncated NBD. Describe the relevant features and differences between these two models.

   b. Also fit and describe a "zero-truncated, one-inflated" model.

   c. Finally, fit an sBG model to the data. What would its "story" be in this context (compared to the way we used it in Session 1)?

   Which of these four models would you select, and why?

[1] This means that you must put in the complete NBD formula for each person, so you'll need to use the gamma function. Newer versions of Excel have GAMMA as a built-in function, but older versions don't. If you're using an older version, you'll have to use log(gamma(x)) via the GAMMALN function. So to get gamma(x) you type EXP(GAMMALN(X)). (All the more reason to use the recursion formula whenever you can!)