Process Summary/Completion Plan ~ Due in class February 15 ~ 10 points
Project Updates ~ Due periodically online, end of week, 11:59 pm ~ 20 points
Mathematical Evidence ~ Tentatively Due online April 19, 11:59 pm ~ 150 points
Polished First Draft ~ Tentatively Due online April 19, 11:59 pm ~ 50 points
Final Draft ~ Tentatively Due online April 26, 11:59 pm ~ 150 points
Reflection ~ Tentatively Due online April 28, 11:59 pm ~ 20 points

The Assignment: This is your primary goal with your work; review it often (and try to think from the perspective of your client). You need to use this to keep your work on target.

Given the data set, find good models for the size and the frequency of claims.

The Deliverables: Save all work in your Teams channel. Be sure to clearly label each file. Upload all final submissions to Canvas when due.

1. Modeling Process Summary/Detailed Project Completion Plan (1-2 pages):

   Read/annote the complete project instructions before continuing.

   Read the Abstract, Introduction and the intro of the Statistical Modeling section of Modeling the Frequency and Severity of Auto Insurance Claims Using Statistical Distributions. Throughout the rest of the project, you may find that the remainder of this article is a valuable resource.

   Using the template provided, briefly summarize the modeling process in your own words ($< \frac{1}{2}$ page). Also include a detailed plan in the table provided for completion of your project for approval (revisions may be requested prior to approval).

   Use the provided course schedule to help you determine appropriate interim due dates. Assume that things will take longer than you plan, and account for this in your schedule.

   Your goal is to create a detailed list to follow to finish your project efficiently.

2. Log of your work:

   Using the template provided, each entry should include five data points - the date, approximate start/stop times (time will compute automatically), who did the work, a brief summary/list of what you worked on, and the results/conclusions of the work.

   There should be one log per team, and this log will be used to evaluate participation by both team members, faithfulness to your schedule, and overall good project management skills.

   Recommended format: Excel file to be saved in Teams (and should be updated on a daily basis), expect me to check on it periodically

3. Mathematical evidence:

   Supporting work to justify the models and parameters you used in your models (formulas and solutions), details on your goodness-of-fit tests and likelihood ratio tests, any other supporting computational evidence.

   There should be one mathematical evidence file for each team.

   Required format: a well formatted and well commented RMarkdown file, submitted as both an HTML (for grading) and a PDF file (for integrity checking), that is neatly organized and presented with content clearly explained (Resource on Using RMarkdown)

4. Report to the decision maker:

   Summarize good models (can have more than one) with all of the discussion that a non-technical audience needs to understand the model and <u>how it can be used</u>; assume the audience is well versed in the data and context (but you still need to be clear about variable definitions and terms you use), but has limited understanding of statistical analysis/tools.

   There should be one report per team. Multiple revisions of the final document are expected (be prepared to tell me how many times you revised). The primary goal is clarity to the audience while clearly justifying your conclusions and presenting all the technical details they need to use your models.

   You will submit the best rough draft you can write (after you have written and then revised it). Feedback will be given and a final draft will be submitted.

   Required format: typed, with attractive (but professional) formatting; appendices are permitted as appropriate/necessary; include a brief commentary on your revision process in your submission (include the number of times the paper was revised)

5. Reflection - Write 1-2 paragraphs for each of the following (clearly label each section, each student should submit their own reflection):

   (a) Experience:

   Briefly summarizing the work done in this project and cite several, specific illustrations of how the experience reinforced or changed your perceptions and attitudes about data modeling.

   (b) Connect to Worldview:

   Briefly discuss how the work done in this project specifically expands a biblical perspective of self, the nature of humanity, and interactions with others in relationship to creation, fall, and redemption. Connects this perspective to course content, citing multiple illustrations and relevant Scripture verses.

   (c) Connect to Discipline:

   Demonstrate the connection between this project and the content covered in Ma 404 and 405 by combining examples, facts, or theories from inside and outside your field of study or perspective.

(d) Transform Perspective:

Discuss how this project impacted your understanding of future life/career. Discuss how this project developed as many of the career competencies as you feel were developed.

The eight career competencies as defined by NACE are critical thinking/problem solving, oral/written communication, teamwork/collaboration, digital technology, leadership, professionalism/work ethic, career management, and global/intercultural fluency. Competency descriptions are available upon request.

(e) Transform Future Work:

Illustrate how you might apply the skills, abilities, theories, and/or methodologies that you developed during this project to current or future situations to solve difficult problems or explore complex issues.

Expected mathematical components:

- Empirical distributions/appropriate graphs
- Select and justify a threshold/benchmark for "large" claims if the data supports it (if not, why not)
- Parametric distribution(s) with parameter estimates (MOM, percentile matching, LSE, MLE) - at least two distribution families for each scenario should be presented

  Omari et al. claim that MLE often yields the better estimate compared to the other estimators. Determine if your results support this claim.
- All appropriate test results for each model/parameter combination, reporting both numerical evidence from tests and graphical evidence of the distribution fit
- Simulate ten runs of 1 year of claims using your best models - include a robust summary of the data from the simulation (consider usual/best/worst cases)

For "large" claims, you should be aware of the following specialty tools: extreme value distributions (beware of special cases of these tools in R that apply only to testing normal distributions) and peaks over threshold (POT) techniques for claim sizes

<u>Potential Resources</u>:

- Storytelling with Data by Knaflic
- Fitting Distributions with R
- Extreme Values in R
- Goodness-of-Fit Tool Package in R
- Anderson-Darling Test
- Chi-Square Goodness of Fit Test

- Kolmogorov-Smirnov Goodness of Fit Test
- Introduction to Simulation Using R
- The Modelling of Extreme Events
- The POT Package
- POT: Modelling Peaks Over a Threshold, see page 34
- A User's Guide to the POT Package
- Peaks Over Threshold Plot
- Application of the Peaks-Over-Threshold Method on Insurance Data
- Peaks Over Threshold (POT): A Methodology for Automatic Threshold Estimation