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.

Work Log: Using the template provided, each entry should include three data points - the date, approximate start/stop times (time will compute automatically), and a brief summary/list of what you accomplished.

Modeling Process Summary/Detailed Project Completion Plan: Read/annotate 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 that you can follow to finish your project efficiently (also review the rubric when creating your list).

Process Summary/Completion Plan (10 pts)

Due at Meeting: February 23, 2024, 3:00pm

Meetings: Schedule a meeting with your professor during the weeks of February 19, March 4, March 25, and April 8 to discuss your progress and your plan for next steps. Submit your work log in advance of the meeting. Come prepared to give an update on your progress and have at least two questions you have about the project, your project planning, or problems you are or anticipate having. This meeting will a graded activity and your grade will be based on your preparation for the meeting, your completion of work, and your participation in the discussion (the more ownership of this meeting you take, the better your grade will be).

Meeting Completed (20 pts) by: February 23, 2024, 3:00pm

Meeting Completed (20 pts) by: March 8, 2024, 3:00pm

Meeting Completed (20 pts) by: March 29, 2024, 3:00pm

Meeting Completed (20 pts) by: April 12, 2024, 3:00pm

R Initialization: Create an R project to contain all your work and save your environment/history and an R Markdown file to contain your code and the outputs. Be intentional and thorough with organization/structure/documentation of your code.

Using section labels (#, ##, etc.) in R Markdown, create an R code section and a short text description/formulas for each thing you think you will need to compute (preferably in the order that you will need to compute it) based on your process summary/completion plan.

Impose structure/hierarchy to make it easier to follow and search through your code (use the Outline feature). Work on learning to format the text portions so that it knits attractively (Latex is helpful here). You will be expected to submit both your R code and an HTML knitted file. Easy navigation of the knitted file through a table of contents is appreciated. (Resource on Using RMarkdown)

Mathematical Evidence:
Your RMarkdown/knitted files will provide the 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. It should be a well-documented summary of all the analysis you did, the decisions and conclusions you made and why they were justified. Summary statements after each code chunks should be included. I should be able to follow your file and understand your analysis and the conclusion you drew from each (even if it doesn't get included in the final report).

You will submit a draft with as much of the analysis as you can get done by the first deadline for feedback as well as a final draft. The same grading rubric will be used for both submissions.

Mathematical Evidence Draft (25 pts)

Due online: April 4, 2024, 11:59pm

Mathematical Evidence Final (150 pts)

Due online: April 18, 2024, 11:59pm

Report to Decision Maker:
Draft a typed, attractive (but professional) report that summarizes your good models (can have more than one) with all of the discussion that a non-technical audience needs to understand the model and how it can be used. 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 does not spend their time doing statistical analysis. Appendices are permitted as appropriate/necessary.

Multiple revisions of the final document are expected (include a note at the end of the report indicating your total number of revisions). 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 and make decisions without adding anything unnecessary.

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

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)

Draft (25 pts)	Due Online: April 11, 2024, 11:59pm
Final Report (150 pts)	Due Online: April 24, 2024, 11:59pm

Specialty Statistical Tools:
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

Potential Online Resources:

- 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

MA 405 PROBABILITY & STATISTICS RUBRIC NAME:

MEETINGS ~ 20 POINTS

_____/ 50% Work log submitted, appropriate time invested and progress made

_____/ 25% Prepared for the meeting with updates of progress and questions

_____/ 25% Managed the meeting well

_____/ 20 pts

MA 405 PROBABILITY & STATISTICS RUBRIC NAME:

MEETINGS ~ 20 POINTS

_____/ 50% Work log submitted, appropriate time invested and progress made

_____/ 25% Prepared for the meeting with updates of progress and questions

_____/ 25% Managed the meeting well

_____/ 20 pts

Ma 405 Mathematical Evidence Submission $\sim$ 25 or 150 points

_____/ 20% Presentation of mathematical work to an external audience (aka me):

Quality of organization/clarity of presentation - use of formatting/chunks/etc, logical presentation of content

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

Quality of commentary - purpose and conclusions for each chunk/subchunk of code is clearly expressed in a logical and concise manner

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 80% Met expectations for the mathematical components - mathematical work is comprehensive, compelling, and complete (made sufficient progress for the draft)

- 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; addressed MLE is better 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 <u>robust</u> summary of the data from the simulation (consider usual/best/worst cases)

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 0% Penalty: poor presentation, failure to follow directions, etc.

_____ / 25 or 150 points

Ma 405 Draft ~ 25 points

_____/ 20% Effective use of visuals and clear presentation of data for an external audience (aka management/decision maker)

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 40% Sufficient information (both technical and non-technical) was presented (models were all well-defined, audience of various technical skill has all they need to understand the model and how it can be used)

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 40% Clear/compelling narrative that explains and justifies the models recommended, easy to read and understand, all necessary information was presented and in the order needed to understand the argument

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 0% Penalty: poor presentation, failure to follow directions, etc.

_____ / 25 points

The final draft rubric will contain all of the above, as well as include two additional pieces:

Y/N were appropriate revisions from rough draft feedback done, and

Y/N were sufficient revisions of the written document completed (you will be asked to estimate the number of times you revised the paper (you may include both rough draft and final draft revisions) - there is not a hard number I am looking for here, but I want to see that you put effort in critiquing your own work and improved your communication based on your own critique, as opposed to just responding to my own feedback).

MA 405 FINAL REPORT ~ 150 POINTS

_____/ 8% Revisions:

Sufficient revision based on feedback? NO ——— YES

Final draft should be notably improved based on the feedback you received on the first draft.

Sufficient number of personal revisions? _____ NO ——— YES

Evaluation is based on quality of the paper and effort invested in the revision process.

_____/ 16% Effective use of visuals and clear presentation of data for an external audience (aka management/decision maker)

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 38% Sufficient information (both technical and non-technical) was presented (models were all well-defined, audience of various technical skill has all they need to understand the model and how it can be used)

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 38% Clear/compelling narrative that explains and justifies the models recommended, easy to read and understand, all necessary information was presented and in the order needed to understand the argument

F — D^- — D — D^+ — C^- — C — C^+ — B^- — B — B^+ — A^- — A — A^+

_____/ 0% Penalty: poor presentation, failure to follow directions, etc.

_____ / 150 points