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.

Research Journal

Best practices in research include a systematic method of *journaling* – keeping a record of research-related activity including observations, conjectures, results (both positive and negative), insights, and reflections. Journals should be kept by each member of the team, and your contribution to work will be evaluated using the journal. Your journal should be digital, be able to be submitted as a Word document or a PDF, and should demonstrate the following characteristics.

1. **Reflection:** The student will succinctly reflect on research activities including work attempted, insights and results gained, and further paths of inquiry to be pursued.
2. **Organization:** The student will create a single document (useful to self and others) that will serve as a quick reference of their work .
3. **Documentation:** The student will document when (dates) and for how long (approximate elapsed time) course-engagement occurred.

Guidelines for Journaling

1. **Style:** As long as all other elements are addressed, the student may use personal style in developing their research journal, although your professor reserves the right to direct changes to improve effectiveness. Journal entries are not meant to be formal pieces of research. Entries are to be useful rather than polished. Informal writing, abbreviation, and bulleted phrases are all acceptable as long as they can be easily deciphered by another mathematician. Daily entries should take less than five minutes.
2. **Frequency:** The student should journal on each day that they engage in research activities.
3. **Date/Time:** Each entry should start with the date, an approximation of total elapsed time for that day, and a running total of time spent on research.
4. **Entry:** The entry for a day should be a brief synopsis of and reflection on the research activity. Some prompts that could be addressed are ...

   (a) What was the question you focused on?
   (b) What attempts were made? with what outcomes?
   (c) What impact might these outcomes have on other avenues of inquiry?
   (d) What questions arose? Do they need immediate attention, future attention, or setting aside?
   (e) What challenges appear to be hampering progress? How might these difficulties be resolved?
   (f) What should you work on next time?

5. **Submission:** The Research Journal is to be submitted to Canvas (or elsewhere if directed) with each project submission. Submissions should include your name in the file name as well as on the first page of the file. Failure to submit the journal will result in a 0 on that portion of the submission (your professor will not go searching for it).

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 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 online: January 20, 2026, 11:59pm

Research Meetings

Schedule a meeting with your professor between January 21 and January 26, weeks of February 9, and February 23 to discuss your progress and your plan for next steps. See the rubric for the expectations.

You should plan to show your professor your R code during your first meeting and get feedback on your organization. See the R Initialization section for expectations.

Meeting Completed (10 pts) by: 1/26/26, 2/13/26, 2/27/26, 3:00pm

Population Distribution Modeling

R Initialization

Create an R project to contain all your work and save your environment/history (so you don't have to keep rerunning everything every time you sit down to work) 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)

** Files in R do not autosave, so save both your R Markdown and R Project files often.

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 or graphical 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 before each code chunks should be included. One should be able to follow your file and understand your analysis and the conclusion you drew from each piece (even if it doesn't get included in the final report).

Submission: 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. Your submission should include your journal, your RMarkdown file, and a knitted HTML file of your RMarkdown file.

Mathematical Evidence Draft (25 pts)	Due online: February 10, 2026, 11:59pm
Mathematical Evidence Final (150 pts)	Due online: February 26, 2026, 11:59pm

Report to Decision Maker

As you draw conclusions, 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. The average person should be able to read your report and understand what the results were, and the statistician should be able to read your report and be convinced based on the statistics that your conclusions are accurate. Appendices are permitted as appropriate/necessary.

This document should be drafted in parallel with your analysis run in R. 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.

Expected mathematical components for final report:

- 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 (using APA formatting) 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)

Submission: 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. Your submission should include your journal and your report.

Draft (25 pts) — Due Online: February 13, 2026, 11:59pm
Final (150 pts) — Due Online: March 6, 2026, 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

Meetings ∼ 10 points, rubrics will be scored in Canvas

_____/ 35% Project Process: Journal maintained, appropriate time invested and progress made

Good progress: A/B Journal received by the meeting, time invested was adequate, progress made is sufficient to not be behind

Acceptable: C/D Journal received by the meeting, time invested was less than expected, progress was made but the work is behind schedule

_____/ 35% Journal Quality: Journal objectives, style , frequency (including date/times), entries quality

Good Document: A/B Journal objectives are met, style is effective, frequency is complete (including date/times), entries are substantive (addressing relevant prompts; all prompts should be addressed, but not every day)

Adequate: C/D Some journal objectives are met, style is adequate but can be improved, frequency is incomplete and/or does not include date/times, entries do not address all relevant prompts and/or some prompts are never addressed

_____/ 15% Meeting Preparedness: Prepared for the meeting with updates of progress and questions

Well Prepared: A/B Gave a clear update, had specific questions that were relevant and important to the work, had a plan for future work

Adequate: C/D Gave an update, had at least a question but was somewhat superficial/could have been answered by using the text, had some idea for what was next

_____/ 15% Meeting Management: Managed the meeting well

Well Handled: A/B Took initiative in the discussion, was comfortable discussing the work accomplished, asked questions clearly, did not need prompting, interaction was comfortable and conversational in nature while still being on task

Developing: C/D Participated in the discussion, was somewhat uncertain in places about the discussion, needed some prompting, interaction did not seem comfortable and conversational in nature or did not stay on task

_____/ 10 pts

Ma 442 Mathematical Evidence Submission ∼ 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 442 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 442 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