**Analysis Project 2**

One of the core objectives of this class is that you can conceptualize, implement, interpret, and write-up statistical analyses. The first analysis project will evaluate your ability to carry out a mixed model. Projects will be distributed and collected via Canvas; you are responsible for confirming that uploaded files are complete. Late projects will be penalized at a rate of 5 points per day. **Your assignment responses are expected to be entirely your own work. Projects cannot be discussed with anyone *except* the instructor and TA**; you may use your notes, class handouts, assigned readings, and statistical software package. Use of the internet is not permitted with the exception of the class resources on Canvas. If you turn in your project before the deadline, you may resubmit incorrect portions for up to ½ of the points lost. This assignment is due **11:59 pm March 31th**.

---

It is being proposed[1] that the state should make a change to vehicle registration laws to remove all cars more than 10 years old from the roads, particularly highways, because of the risk of death associated with high-speed crashes in older vehicles. Advocacy for this change is being supported by a data set, "Accident2022.csv", which reports information about serious car accidents where at least one vehicle was towed. In these data, whether a person in the front seat died is related to the age of the car at the time of accident, even after controlling for the estimated car speed at the time of the crash. The variables in the data set are as follows:

| | |
|---|---|
| died | Whether or not a person died in the accident |
| speedCat | Ordered factor consisting of estimated car speed at time of impact in mph |
| seatbelt | Whether the person was wearing a seatbelt ("belted") or not ("none") |
| frontImpact | Whether the accident involved hitting the front of the car ("FrontalImpact") or not ("NonFrontalImpact") |
| ageOFocc | Age of occupant in years |
| occRole | Whether the occupant was in the driver ("driver") or passenger ("pass") seat |
| yearVeh | Year the vehicle was manufactured |
| vehicleAge | Age of the vehicle at the time of the accident (in years) |
| Airbag | Whether the care was equipped with airbags ("Airbag"), or had no airbags ("NoAirbag") |
| Deploy | Whether there was airbag deployment ("AirbagDeployed") or note ("NoAirbagDeployed"). All cars without airbags will have a value of "NoAirbagDeployed". |

**Requirements:**

**Goal:** Your primary goal is to address whether or not there are differences in the odds of dying in a serious car accident based on the age of the vehicle at the time of the crash. Naturally, this question is more complicated than just regressing the death variable on the age of the vehicle, as there are many other factors that may affect the odds of death in a car accident. The data set includes additional variables related to the car, accident conditions, and individual in the accident. Begin by confirming whether the estimated speed of the car (speedCat) and age of the vehicle (vehicleAge) are significant predictors of death (died), and discuss how the odds of death change based on the age of the vehicle. Subsequently, based on your best understanding (or best guess) as to factors that relate to surviving car accidents, develop an analysis plan that considers alternative explanations between the age of the car and accident survival. Aim to provide a detailed discussion of whether or not these data support the need to remove older cars from the roads because of the danger posed for the vehicle occupants.

---

[1] This scenario, to my knowledge, is fictional. I have no knowledge that Utah is actually considering this change to vehicle registration laws.

**Write-up:** The write up should consist of 4 sections: 1) an introduction to your question(s) of interest and hypotheses, 2) a detailed description of the analyses that were conducted and the models you chose to run, 3) a description of the results of your analyses, and 4) a short discussion summarizing the implications of your results. That is, sections corresponding to the introduction, methods, results, and discussion sections of a typical research paper.

Write-ups should be clear, concise, and professionally written. There should be no misuse/ambiguity regarding statistical terms, and clear statistical evidence should be given for all statements/claims made. There is no minimum or maximum page limit, but it is likely that at least 3-5 pages (double-spaced, type 12) will be required to address the goal well.

**Syntax/Output:** Regardless of the program used for analysis, please provide copies of all syntax and output used in the analyses. Please provide your write-up as a separate file from your syntax/output. Including comments in your code would be very helpful for Steven.

**Analysis:** The analysis will draw on skills that you have learned this semester and in your prior statistics classes. In your analyses, the following skills should be demonstrated:

- Use of the generalized linear model to fit a non-normally distributed outcome, with a link function other than the identity link
- Consideration of multiple models with varying complexity (i.e., not just models considering each of the variables in isolation). There are seven variables aside from died, vehicleAge, and speedCat. Use at least 3 of these additional 7 variables.
- Interpretation of model parameter estimates (coefficients), and consideration of the magnitude of effects (i.e., practical significance) not just p-values. Because of the size of the data set, many effects will be statistically significant, but may be of less practice significance because of the size of the effects.
- Calculations of confidence intervals, and pseudo $R^2$. While pseudo $R^2$ can't be understood as variance explained, the values and changes in values might be helpful in considering the size of effects. When interpreting the pseudo $R^2$, you can use values from ordinary regression $R^2$ to give you a sense of whether effects are small/medium/large.
- Test and interpret at least one 2-way interaction; if none of the 2-way interactions you examine are significant, still demonstrate your ability to interpret the coefficients for at least one non-significant interaction
- Creating a plot in the metric of the dependent variable (i.e., probabilities) of at least one interaction