MAS61004 - Assessed Project

Allan Cousins, Matthew Knowles, Oliver Hewitt, Andrew Smith

2/23/2023

Executive Summary

The main purpose of this investigation is to describe the relationship primarily between accident severity and speed limit, and other factors included in the data provided by the UK Government's Department for Transport.

Our focus is to assess the risk of a fatal or severe injury, given that one has already occurred - and so the data in concern is road safety data exclusively from the year 2020. Analysis in the report will be carried out in R using Generalised Linear Models (GLMs), as it would be incorrect to assume the data is normally distributed.

(Outline conclusions here)

Method

Our primary analysis, as outlined in the summary, will be Generalised Linear Models in R. The data provided are public domain, and include: "Road Safety Data - Accidents 2020", "Road Safety Data - Vehicles 2020" and "Road Safety Data - Casualties 2020" merged by an accident index present in all three data sets.

As mentioned previously, we wouldn't assume the data is normally distributed nor would we expect the outcome or our residuals to be normally distributed - i.e. we assume that a large demographic of the UK were not involved in a traffic accident in 2020, and an even smaller demographic were involved once or multiple times. In addition, not all response variables considered in the analysis are continuous - hence GLM would be an appropriate method to use.

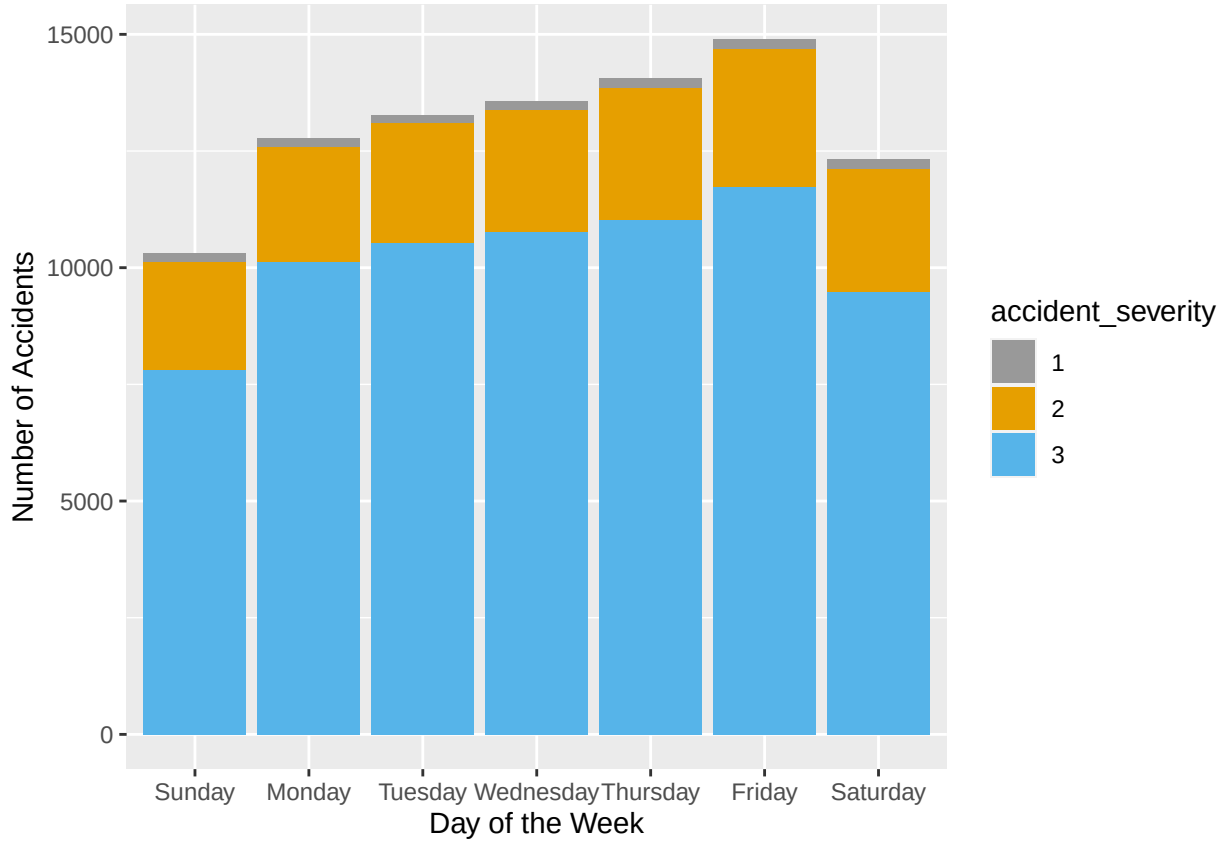
With this established, we determined which GLM to use using ggplot (see Appendix, A1). There doesn't appear to be an immediate pattern or linearity in terms of the effect of speed limit on accident count, and so we may choose a standard Poisson Regression model, which is best used for modelling events where the outcomes are counts. Additionally, the Poisson Distribution is most commonly used to find the probability of events occurring within a given time interval. [1]

(Talk about the 3 components of GLM here)

Results

Desciptive plots:

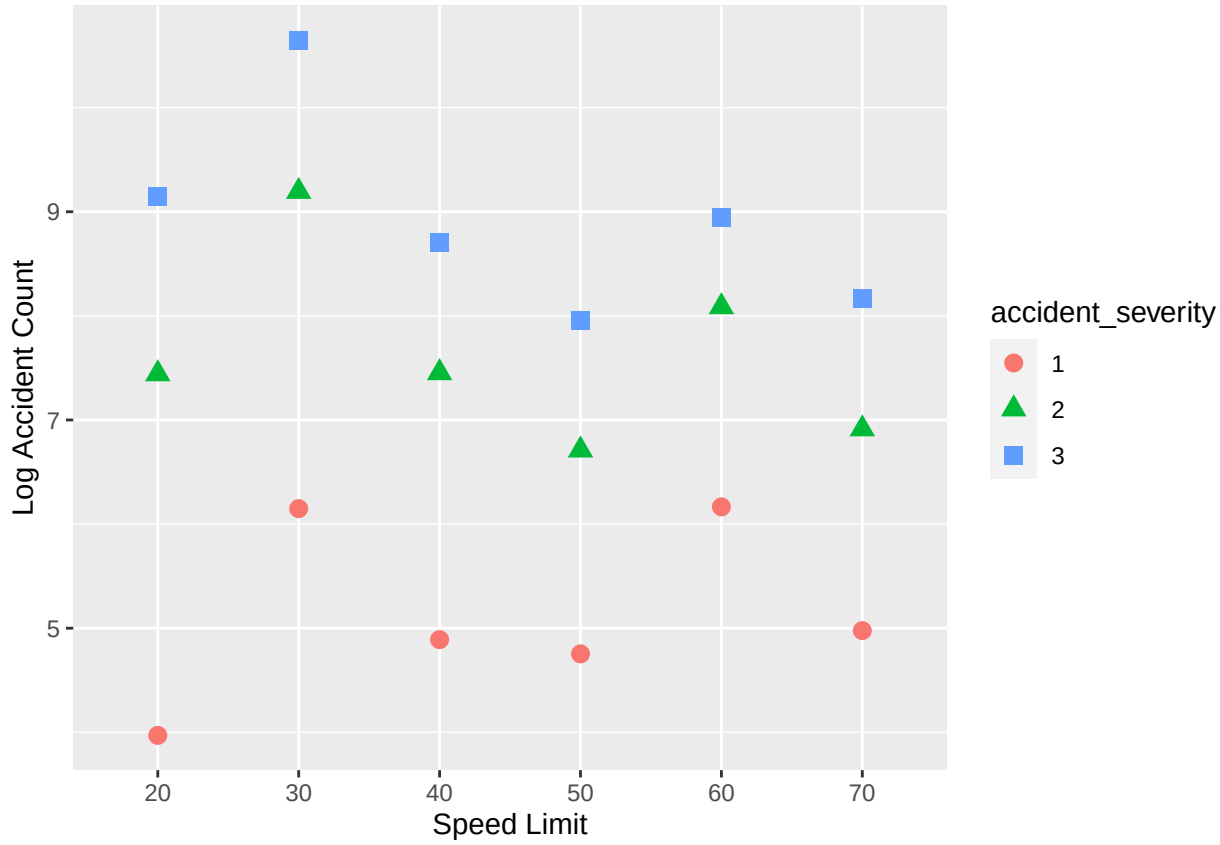
Figure 1: Plot of Number of Accidents by Day.



GLM:

Appendix

A1: Log Accident Count by Speed Limit



A2: *glm* function in R using the *poisson* argument

```
use_in_formula <- c(
    9, 10, 13, 18, 20, 21, 22, 24, 28, 29, 30
)
accidents_key <- accidents[, use_in_formula]
has.neg <- apply(accidents_key, 1, function (row) any(row < 0))
accidents_data <- accidents_key[-which(has.neg), ]
model <- glm(data = accidents_data,
             formula = accident_severity ~ .,
             poisson
             )
```

A3: Summary of A2

```
##
## Call:
## glm(formula = accident_severity ~ ., family = poisson, data = accidents_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.42308   0.06348   0.11865   0.14871   0.29355
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               1.058e+00  1.438e-02  73.569  < 2e-16 ***
## number_of_vehicles        1.834e-02  2.956e-03   6.205 5.46e-10 ***
## day_of_week              -1.013e-05  1.033e-03  -0.010 0.992176
## first_road_class         -1.467e-03  1.508e-03  -0.973 0.330715
## road_type                -4.353e-03  1.244e-03  -3.500 0.000466 ***
## speed_limit              -1.444e-03  1.616e-04  -8.932  < 2e-16 ***
## junction_detail           4.604e-04  1.624e-04   2.834 0.004593 **
## second_road_class         9.300e-04  7.907e-04   1.176 0.239556
## light_conditions         -3.200e-03  1.171e-03  -2.733 0.006271 **
## weather_conditions        2.449e-03  1.171e-03   2.091 0.036487 *
## road_surface_conditions   3.949e-03  2.338e-03   1.689 0.091241 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 7743.6  on 90759  degrees of freedom
## Residual deviance: 7559.3  on 90749  degrees of freedom
## AIC: 270827
##
## Number of Fisher Scoring iterations: 4
```

References

[1] "Tutorial: Poisson Regression in R" - dataquest, created February 27th 2019. https://www.dataquest.io/blog/tutorial-poisson-regression-in-r/