Association Between Time and Alcohol Offenders on Automobile-related Fatalities and Injuries.

Oluwatomi Hassan

2023-11-19

Executive Summary

Negative binomial regression is used to examine the relationship between time, day and number of alcoholic offenders on the number of fatalities or injuries automobile crashes in New Zealand in 2009. Result suggests that there is association between these variables and number of alcohol offenders depend on morning time on automobile-related fatalities and injuries.

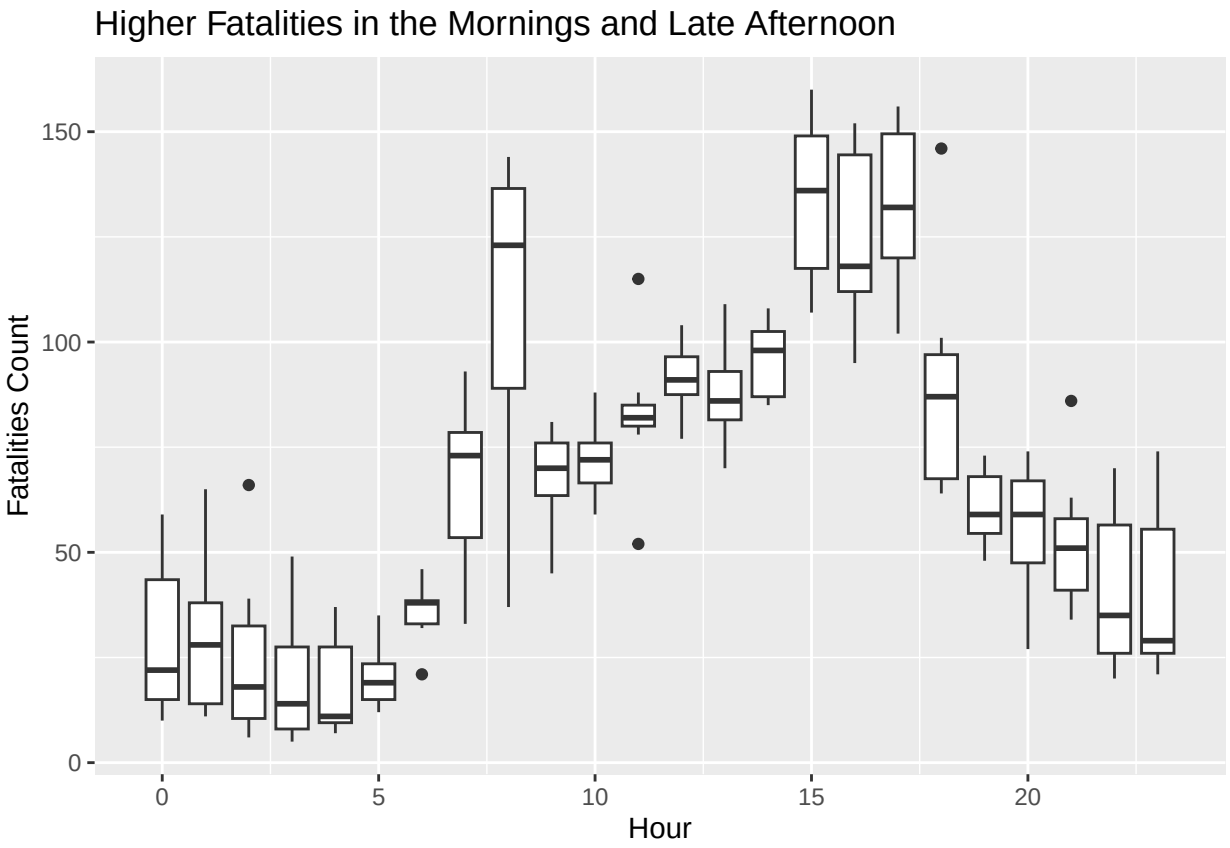
Introduction

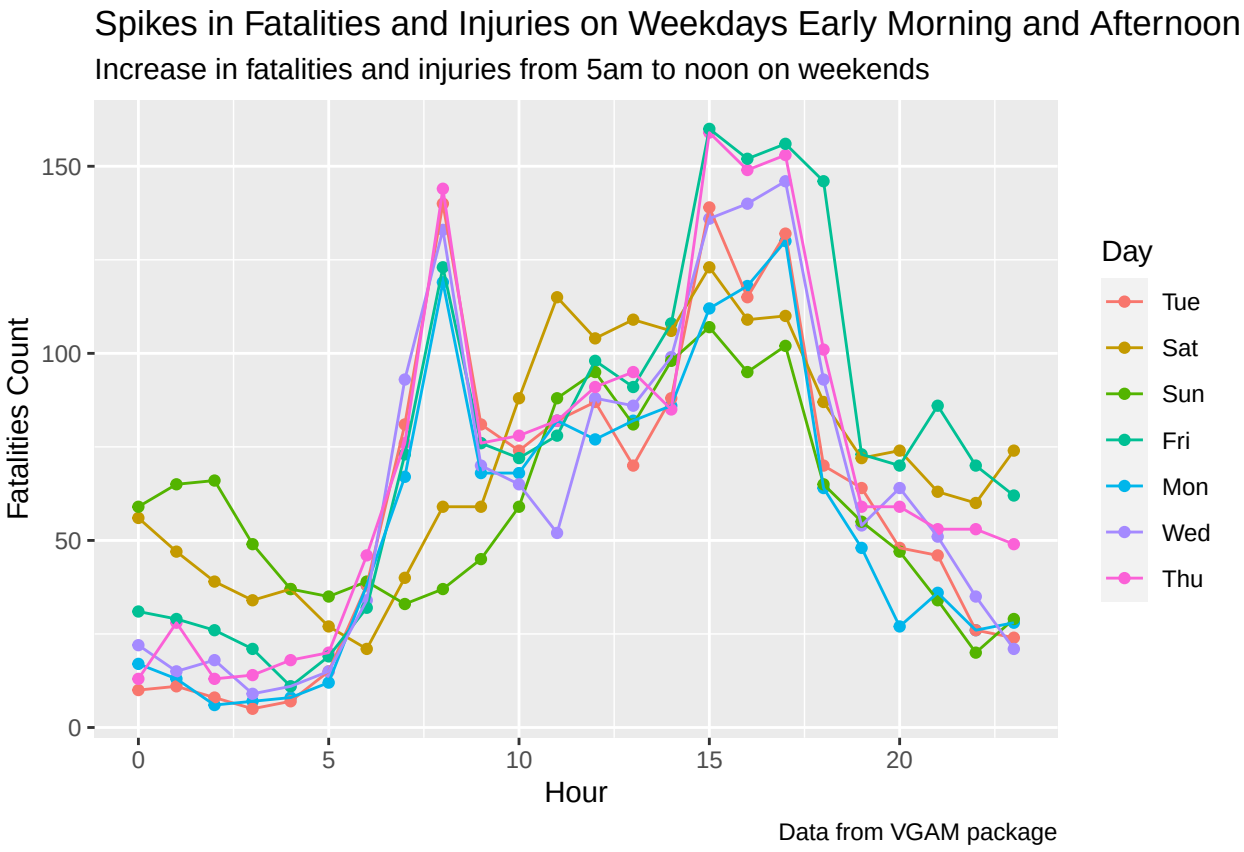
Alcohol-related crashes can lead to fatalities for involved individuals with law enforcement and public policies aimed to reduce the number of events. A combination of the crashi, crashf and alcoff dataset in the VGAM package constitutes the data used in the analyses of the relationship between fatalities and injuries in automobile crashes and alcohol offenders on a given day in New Zealand. The crashi, crashf and alcoff dataset consists of reported number of injuries, number of fatalities and number of alcohol offenders from breathalyzed drivers involved in crashes on New Zealand roads in 2009, respectively. The rows of the initial data represents the hourly times on a 24-hour clock starting at midnight and the columns represent the seven days of the week. Consider the following questions of interest: Are the number of fatalities and injures by automobile crashes related to the number of alcoholic offenders on a given day or time? Are there interactions between the time or day and the number of alcoholic offenders on the number of fatalities or injuries by car?

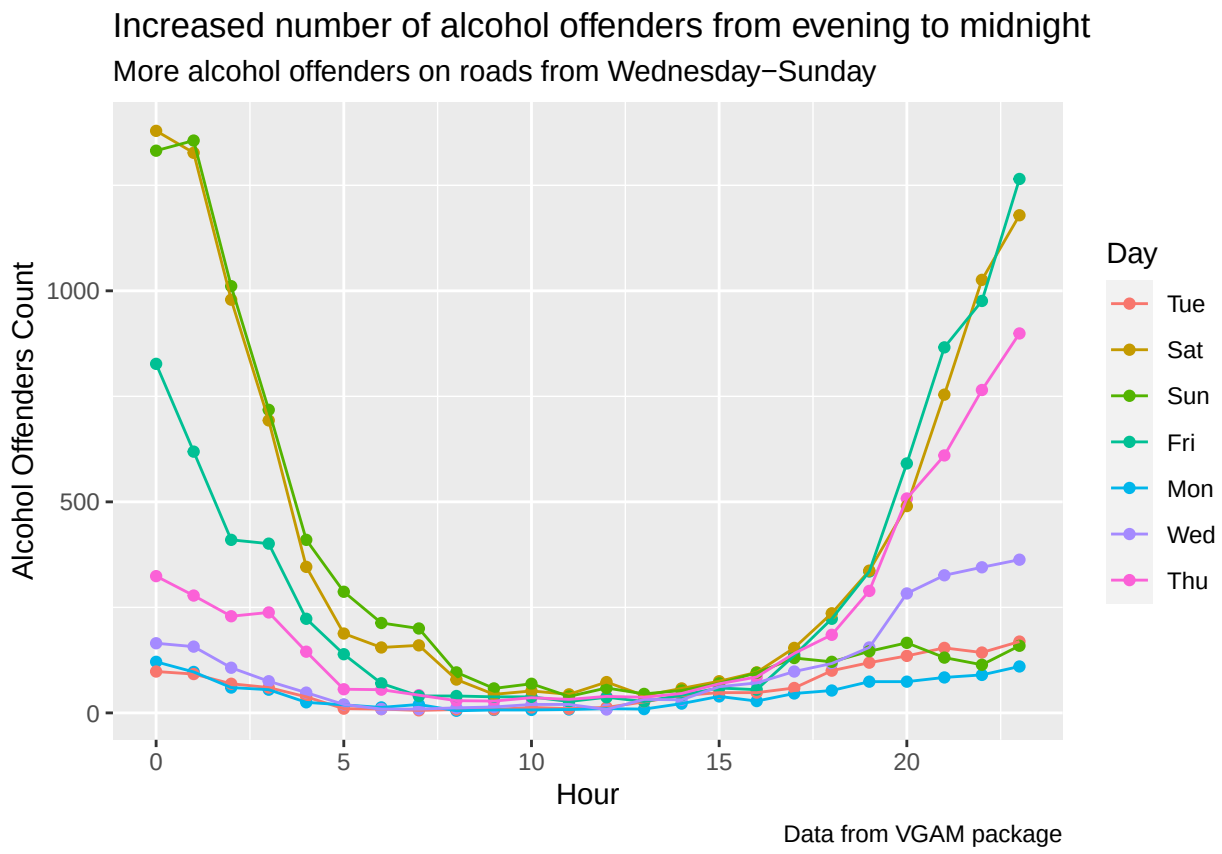
Methods and Models

In this analysis, the response variable is the combined number of fatalities and injuries by automobile crashes so poisson regression is initially fitted to the data. Since the response variable is composed of unbounded non-zero counts with evidence of over-dispersion unaccounted by poisson model, negative binomial regression model is a more appropriate method of analysis. Exploration of the data suggest that the number of fatalities and crashes on weekdays differs from weekends in certain period of time. Time is grouped into early morning, morning, afternoon and evening with weekdays, Friday, Saturday and Sunday. To answer the questions of interest, negative binomial regression model is utilized to model the number of fatalities and injuries with explanatory variables time, day and number of alcoholic offenders. A series of drop in deviance test is performed to analysis the relationship between day, time, and the number of alcoholic offenders and their interactions(in appendix). ANOVA function is used to compare the full model with interactions to the reduced model without interactions for the poisson and negative binomial models. The LRstats function in the R VcdExtra package to compare the poisson and negative binomial models fit to the dataset. The data description and design suggests that that the assumption of independence is met. The Pearson and Deviance residual shows that the model is an appropriate fit for the data.

Result

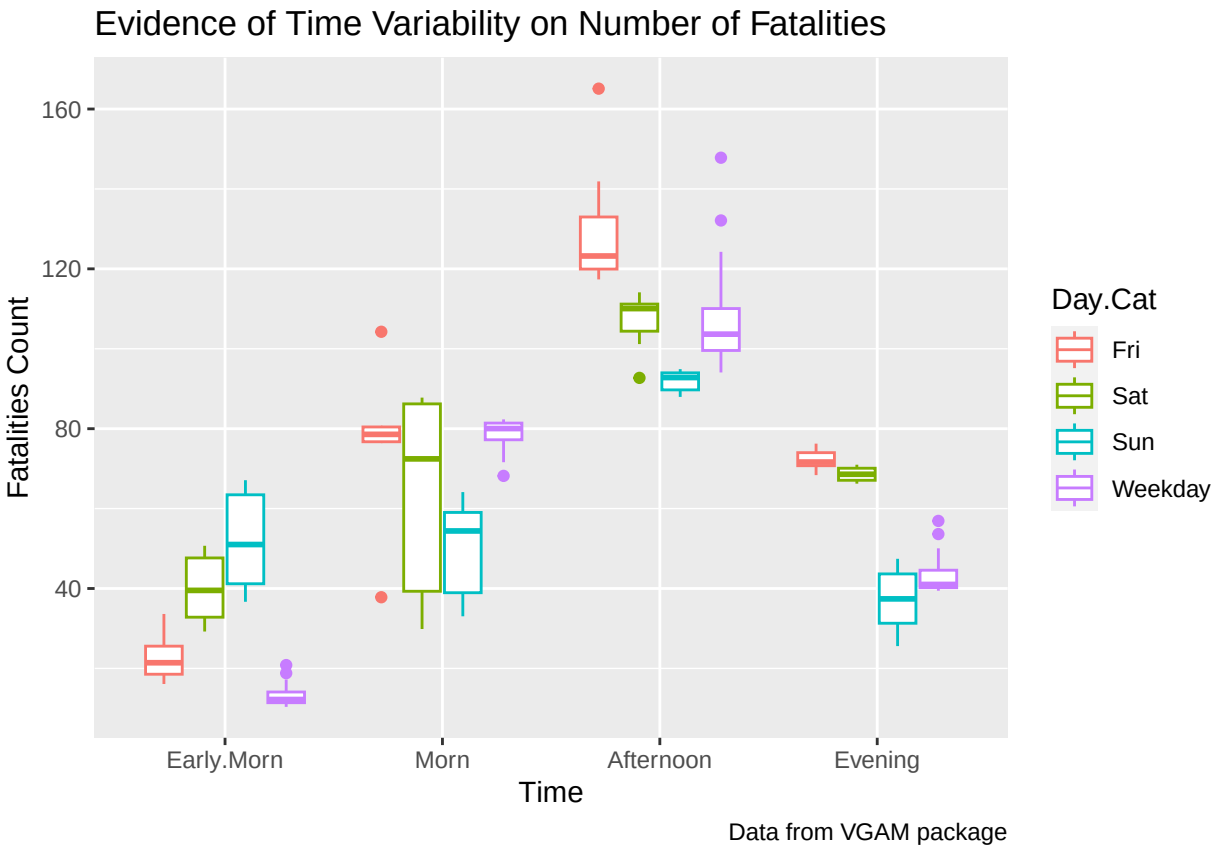

Higher Fatalities in the Mornings and Late Afternoon
Fatalities Count
150
100
50
0
0
5
10
15
20
Hour


Spikes in Fatalities and Injuries on Weekdays Early Morning and Afternoon
Increase in fatalities and injuries from 5am to noon on weekends
Fatalities Count
150
100
50
0
0
5
10
15
20
Hour
Data from VGAM package
Day
Tue
Sat
Sun
Fri
Mon
Wed
Thu



Comparison of the rich and reduced model without interactions suggests that the rich model is a more appropriate fit(Drop in Deviance Test, p-value = 7.206e-11). Residual plots does not indicate any violations from assumption of independence and linearity(See appendix). The result suggest that all three variable are associated with the number of fatalities and injuries involved in automobile crashes(table below).There is convincing evidence that the number of fatalities and injuries by automobile crashes is positively associated with time of crash on New Zealand roads in 2009(table below). There is moderate evidence that Sunday is positively associated with the number of fatalities and injuries in car-related crashes. The number of alcohol offenders is somewhat positively associated with the number of fatalities and injuries in car-related crashes in New Zealand in 2009(table below). To understand if association between the number of fatalities and injuries and time or day of crash depends on the number of alcohol offenders, two way interactions between time, day and alcohol offenders variables is examined. There is convincing evidence that the number of alcohol offenders is negatively correlated with morning time on the number of fatalities and injuries in automobile crashes(p-value = 0.00271).

	Estimate	P-value
Sunday	0.809	0.030
Morning	2.653	3.26e-09
Afternoon	2.088	2.18e-11
Evening	1.743	2.14e-05
Alcohol offenders	0.001	0.044
Morning:Alcohol offenders	-0.025	0.003



Conclusion

This study aims to find the relationship between time, day and number of alcoholic offenders on the number of fatalities and injuries in automobile crashes and determine if there is any interaction between these variables. Using negative binomial model, there is evidence of positive association between these variables and the number of fatalities and injuries. Although there is strong evidence of association between the number of alcohol offenders and automobile fatalities and injuries, it is a relatively small positive relationship. The number of fatalities and injuries in car-related crashes depend on both morning time and number of alcohol offenders. Since the distributions of car-related fatalities and injuries show the same trend, further studies can determine if grouping Friday, Saturday and Sunday together leads to different results.

Appendix

```
# Data Pre-processing

## combining and cleaning crashi and crashf data
hour <- rownames(crashf) ## grab the hours
crashi2 <- stack(crashi)
crashf2 <- stack(crashf)
alcoff2 <-stack(alcoff)
resp <- crashi2["values"] + crashf2["values"]
crashf2["alc_offenders"] = alcoff2["values"]
names(crashf2) <- c("fatal_count","Day", "alc_offenders")
```

```r
crashf2["fatal_injured"] = resp
crashf2$Day <- factor(crashf2$Day,levels(crashf2$Day)[c(2,6,7,5,1,3,4)])  # make sure the days are orde
crashf2$Hour <- as.numeric(rep(hour, ncol(crashf)))

## Create a weekday variable with Mon, Tue, Wed, Thu and Fri
## categorize the times into Early Morning, Morning, Afternoon, and Evening
crashf2 %<>% mutate(.,Day.Cat = ifelse(( Day !="Fri" & Day != "Sat" & Day != "Sun"),"Weekday",as.charac
    breaks = c(-1, 5.5, 11.5, 18.5, 25),
    labels = c("Early.Morn", "Morn", "Afternoon", "Evening")))

# EDA

ggplot(data = crashf2, mapping = aes(x = Hour, y = fatal_injured)) + geom_boxplot(mapping = aes(group =



## Scatter plot shows the number of fatal and injured vs the hour separated by the Day
ggplot(data = crashf2, aes(x = Hour, y = fatal_injured, group = Day)) +
    geom_point(aes(colour = Day)) +
    geom_line(aes(colour = Day))+ facet_wrap(~Day) + labs(title="Daily Distribution of Automobile Victi


## scatter plot of fatal_injured vs hours split by day
ggplot(data = crashf2, aes(x = Hour, y = alc_offenders, group = Day)) +
    geom_point(aes(colour = Day)) +
  geom_line(aes(colour = Day)) + facet_wrap(~Day) + labs(title="Daily Distribution of Alcohol offenders

## Scatter plot shows the overall distribution of the number of fatal and injured vs the hours
ggplot(data = crashf2, aes(x = Hour, y = fatal_injured, group = Day)) +
    geom_point(aes(colour = Day)) +
    geom_line(aes(colour = Day)) + labs(title=paste("Spikes in Fatalities and Injuries on Weekdays Earl

# Model Fitting

## fitting a rich poisson model with interactions
## Evidence of overdispersion(763.05/144 > 1)
rich_pos <- glm(fatal_injured ~ Day.Cat * Time.Cat * alc_offenders, data = crashf2,
                family = "poisson")
summary(rich_pos)

## fitting the reduced poisson model with no interactions
pos_mod <- glm(fatal_injured ~ Day.Cat + Time.Cat + alc_offenders, data = crashf2,
    family = "poisson")
summary(pos_mod)

## fitting full negative binomial regression with interactions
rich_nb_mod <- glm.nb(fatal_injured ~ Day.Cat * Time.Cat * alc_offenders, data = crashf2)
summary(rich_nb_mod)

## fitting negative binomial regression without interactions
nb_mod <-  glm.nb(fatal_injured ~ Day.Cat + Time.Cat + alc_offenders, data = crashf2)
summary(nb_mod)
```

```r
## Comparing rich negative binomial, negative binomial, rich poisson and poission model
LRstats(rich_nb_mod,nb_mod,rich_pos,pos_mod)

## Drop in deviance test for negative binomial model with and without interactions.
anova(nb_mod,rich_nb_mod,test="Chisq")

## fitting nested negative binomial regression with only Day.Cat and Time.Cat variables
nb_red_mod2 <-  glm.nb(fatal_injured ~ Day.Cat + Time.Cat, data = crashf2)
summary(nb_red_mod2)

## fitting nested negative binomial regression with only Day.Cat variable
nb_red_mod3 <-  glm.nb(fatal_injured ~ Day.Cat, data = crashf2)
summary(nb_red_mod3)

# Box plot of fitted response variables vs Time.Cat explanatory model
crashf2$fits <- predict.glm(rich_nb_mod,crashf2,type="response")
ggplot(data = crashf2, aes(x = Time.Cat, y = fatal_injured)) + geom_boxplot(aes(Time.Cat,fits, color = I

# Residuals

## Drop in deviance test for nested model and full model without interactions.
anova(nb_red_mod2,nb_mod,test="Chisq")

## Drop in deviance test for nested model and full model without interactions.
anova(nb_red_mod3,rich_nb_mod,test="Chisq")

## residual plots using full negative binomial model
crashf2$residuals_deviance <- residuals(rich_nb_mod)
crashf2$residuals_pearson <- residuals(rich_nb_mod, type = "pearson")
ggplot(data = crashf2, aes(Hour,residuals_deviance)) + geom_point()

## person residual plots using full negative binomial model
ggplot(data = crashf2, aes(Hour,residuals_pearson)) + geom_point()
ggplot(data = crashf2, aes(residuals_deviance,residuals_pearson)) + geom_point()
```