Piotr Dzięcioł-Kryciński
github.com/PethertheGreat

# APPLICATION OF AN ORDERED LOGIT MODEL TO ANALYZE THE IMPACT OF FACTORS ON TRAFFIC ACCIDENT FATALITIES

## Introduction

About million people dies every year according to the WHO. This frightening statistic needs to be studied more closely to identify the most important factors determining the fatality rate. This report attempts to analyze traffic accident mortality (killed, injured or no casualties) in Montgomery County, USA from various factors based on an ordered logit model. The estimated model can help identify to what extent the various factors influence accident fatality.

Attempts to conduct a similar analysis have been made before. Researchers from Korea and the United States have used a logit model to test the effect of factors such as gender, driver age, built environment and road conditions on traffic accident fatalities using data from countries such as Saudi Arabia, the United States and Poland (Lee et al., 2023). An article by researchers at the University of Texas using an ordered probit model examines the degree of driver damage depending on the type of collision (Kockelman & Kweon, 2002), and an article by researchers at the University of Wyoming using an ordered logit model examines the determining factors on the occurrence of serious damage in car accidents (Rezapour et al., 2019)

## Data analysis

The dataset comes from the official US government website (data.gov). It contains detailed information on all traffic collisions occurring in Montgomery County, collected through the Automated Crash Reporting System (ACRS) of the Maryland State Police and reported by local police. The entry was last edited on February 9, 2024, attesting to the timeliness of the dataset.

Link to the dataset: Crash Reporting

After the dataset was downloaded, it was initially reduced by redundant categories using Excel. Rows containing incomplete or ambiguous data were also removed, thus reaching about 90,000 records for analysis. The cleaned dataset was then imported into STATA.

The first 5 rows of the cleaned dataset are presented as follows:

| ReportType | Weather | Light | DriverSubstanceAbuse | DriverdDistracted | SpeedLimit |
|---|---|---|---|---|---|
| Injury Crash | CLEAR | 0 | NONE DETECTED | 0 | 35 |
| Injury Crash | CLEAR | 0 | NONE DETECTED | 0 | 35 |
| Property Damage Crash | CLEAR | 2 | NONE DETECTED | 0 | 40 |
| Property Damage Crash | RAINING | 2 | NONE DETECTED | 0 | 20 |
| Injury Crash | CLEAR | 0 | NONE DETECTED | 0 | 35 |

*ReportType:* explanatory, categorical variable having 3 values:

- Fatal Crash - accident with fatalities
- Injury Crash - accident with injured persons
- Property Damage Crash - accident without human casualties (property damage)

*Weather:* a categorical variable having 4 values:

- CLEAR - cloudless weather
- RAINING - rainfall
- CLOUDY/FOGGY - overcast or foggy weather
- - ICE/SNOW - ice or snow

*Light:* a categorical variable having 3 categories:

- 0 - in case the accident occurred during the daytime
- 1 - in case the accident occurred at dawn/dusk
- 2 - in case the accident happened at night

*DriverSubstanceAbuse:* a categorical variable having 4 values:

- ALCOHOL - presence of alcohol in the driving person
- ALCOHOL & DRUGS - presence of alcohol and drugs in the driving person
- DRUGS - presence of drugs in the driving person
- NONE DETECTED - no presence of drugs and alcohol in the driving person

*DriverDistracted:* a categorical variable having 2 values:

- 0 - when the driver was not distracted during the accident
- 1 - in a situation when the driver causing the accident was distracted (using the phone, eating, smoking cigarettes, etc.).

*SpeedLimit:* a discrete variable expressed in miles indicating the speed limit in effect at the scene of the accident having values from 5 to 75.

In STATA, the text variables ReportType, Weather, DriverSubstanceAbuse were transformed to numeric form, creating the variables ReportType_n, Weather_n, DriverSubstanceAbuse_n, respectively.

**Model selection**

Due to the fact that the explanatory variable has more than two values and can be ordered in a certain way (a better situation is when no one dies in an accident, and an even better situation is when no one is injured), it was decided to choose an ordered logit model for the analysis.

**Results**

After creating an ordered logit model in STATA, the following results were obtained:

```
. olog ReportType_n i.Weather_n i.Light i.SubstanceAbuse_n i.DriverdDistracted i.SpeedLimit

Iteration 0:   log likelihood = -63646.947
Iteration 1:   log likelihood = -63190.558
Iteration 2:   log likelihood = -63188.992
Iteration 3:   log likelihood =  -63188.87
Iteration 4:   log likelihood = -63188.844
Iteration 5:   log likelihood = -63188.838
Iteration 6:   log likelihood = -63188.837
Iteration 7:   log likelihood = -63188.837
Iteration 8:   log likelihood = -63188.837

Ordered logistic regression                     Number of obs = 93,752
                                                LR chi2(23)   = 916.22
                                                Prob > chi2   = 0.0000
Log likelihood = -63188.837                     Pseudo R2     = 0.0072
```

The value of ,,Prob > chi2" equal to 0 indicates the correctness of the performed model.

```
       /cut1 |  -7.76132   .1767289                     -8.107702   -7.414938
       /cut2 | -1.236527   .1427126                     -1.516239   -.9568154
```

The values of cut1 and cut2 at levels of -7.76 and -1.24, respectively, means that when the base model is set (clear sky, full light, presence of alcohol in the driver, speed limit = 0), we get 3 compartments for classification. The value for the base model is 0, and this is modified if other coefficients are selected.

If the sum of the coefficients of the variables of a given combination of variables is less than -7.76 the item would be classified as ,,Fatal Crash".

In the case the sum of the coefficients was between -7.76 and -1.24 the item would be classified as ,,Injury Crash".

In the case the sum of the coefficients was greater than -1.24 the item would be classified as , "Property Damage Crash".

Based on the table below, the interpretation of each variable is as follows:

| ReportType_n | Coefficient | Std. err. | z | P>\|z\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| Weather_n | | | | | | |
| FOGGY/CLOUDY | -.0428352 | .0208167 | -2.06 | 0.040 | -.0836351 | -.0020353 |
| ICE/SNOW | .1287012 | .0581423 | 2.21 | 0.027 | .0147444 | .242658 |
| RAINING | -.0414859 | .0198971 | -2.09 | 0.037 | -.0804835 | -.0024883 |
| Light | | | | | | |
| 1 | .0940932 | .0332958 | 2.83 | 0.005 | .0288346 | .1593519 |
| 2 | .0666405 | .0161826 | 4.12 | 0.000 | .0349231 | .0983579 |
| SubstanceAb~n | | | | | | |
| ALCOHOL & .. | -.5123736 | .2762946 | -1.85 | 0.064 | -1.053901 | .0291539 |
| DRUGS | -.741105 | .1588279 | -4.67 | 0.000 | -1.052402 | -.4298081 |
| NONE DETEC.. | -.4667064 | .0456366 | -10.23 | 0.000 | -.5561526 | -.3772603 |
| 1.DriverdDi~y | -.0814784 | .0253304 | -3.22 | 0.001 | -.1311251 | -.0318317 |
| SpeedLimit | | | | | | |
| 10 | .1366263 | .1812596 | 0.75 | 0.451 | -.218636 | .4918885 |
| 15 | .2196722 | .1519981 | 1.45 | 0.148 | -.0782386 | .5175831 |
| 20 | .1840663 | .1742256 | 1.06 | 0.291 | -.1574095 | .5255422 |
| 25 | -.0467188 | .1364275 | -0.34 | 0.732 | -.3141118 | .2206743 |
| 30 | -.2841661 | .1362326 | -2.09 | 0.037 | -.5511771 | -.0171551 |
| 35 | -.3953908 | .1356348 | -2.92 | 0.004 | -.6612302 | -.1295514 |
| 40 | -.5224866 | .135866 | -3.85 | 0.000 | -.7887791 | -.2561941 |
| 45 | -.6105998 | .1370564 | -4.46 | 0.000 | -.8792255 | -.3419741 |
| 50 | -.5643992 | .140398 | -4.02 | 0.000 | -.8395742 | -.2892241 |
| 55 | -.3576379 | .141286 | -2.53 | 0.011 | -.6345535 | -.0807224 |
| 60 | -.3455472 | .3297524 | -1.05 | 0.295 | -.99185 | .3007557 |
| 65 | -.0776146 | .3932004 | -0.20 | 0.844 | -.8482732 | .6930441 |
| 70 | 12.87355 | 536.3358 | 0.02 | 0.981 | -1038.325 | 1064.072 |
| 75 | 12.83218 | 929.1962 | 0.01 | 0.989 | -1808.359 | 1834.023 |

*Weather_n*

Compared to the base category (CLEAR), rainy or overcast weather increases the chance of a more fatal accident. Winter weather (SNOW/ICE) decreases this chance.

*Light*

Compared to the base category(0 - full light), both twilight and night reduce the chance of those injured or dead in an accident.

*DriverDistracted*

The fact that the driver is distracted according to the model increases the chance of an accident fatality.

DriverSubstanceAbuse

According to the model, any situation other than the presence of alcohol in the driver (alcohol along with drugs, the presence of drugs alone or the absence of drugs) increases the chance of an accident fatality.

*SpeedLimit*

Based on the model, we can conclude that speed limits of 10, 15 and 20 miles per hour reduce accident fatality compared to a speed limit of 5 miles per hour. Speed limits of 25 to 65 miles per hour increase the risk of fatality compared to the base category, and limits of 70 and 75 miles per hour increase the chance of no fatalities significantly. The phenomenon may be due to the low number of observed fatalities at higher speed limits or the low representation of data containing high speed limits.

**Conclusions**

The paper uses data obtained from a U.S. government website, which went there through the local government. Due to the lack of confidence in the accuracy of data reporting, it is difficult to assess the reliability of the results obtained. Assuming the authenticity of the data, an attempt was made to analyze the dependence of traffic accident fatalities (killed, injured or no casualties) on various factors based on the econometric model built. In order to determine the magnitude of the influence of the explanatory factors of the exam result, the results obtained from the estimated ordered logit model were used.

Interpretation of the model's parameter evaluations leads to the conclusion that, with respect to the base model (cloudless skies, full illumination, no driver distraction, driver blood alcohol found, speed limit of 5 miles per hour), factors such as rainfall, cloud cover, the fact of driver distraction or a speed limit between 25 and 65 miles per hour increase the probability of accident fatality. However it is worth noting that due to lack of input data cohesion some results of the analysis may be not intuitive. That implies further research on the given topic.

**STATA code**

```
import excel "C:\Users\Piotr\Downloads\CRASH_STATA.xlsx", sheet("Crashes")
firstrow

encode ReportType, generate(ReportType_n)

encode Weather, generate(Weather_n)

encode DriverSubstanceAbuse, generate(SubstanceAbuse_n)

olog ReportType_n i.Weather_n i.Light i.SubstanceAbuse_n i.DriverdDistracted
i.SpeedLimit
```

**List of bibliographic items used in the work.**


- Kockelman, K. M., & Kweon, Y. J. (2002). Driver injury severity: an application of ordered probit models. Accident Analysis & Prevention, 34(3), 313-321.
- Lee, D., Guldmann, J.-M., & von Rabenau, B. (2023). Impact of driver’s age and gender, built environment, and road conditions on crash severity: A logit

modeling approach. International Journal of Environmental Research and Public Health, 20(3), 2338. doi:10.3390/ijerph20032338

- Rezapour, M., Moomen, M., & Ksaibati, K. (2019). Ordered logistic models of influencing factors on crash injury severity of single and multiple-vehicle downgrade crashes: A case study in Wyoming. Journal of safety research, 68, 107-118.