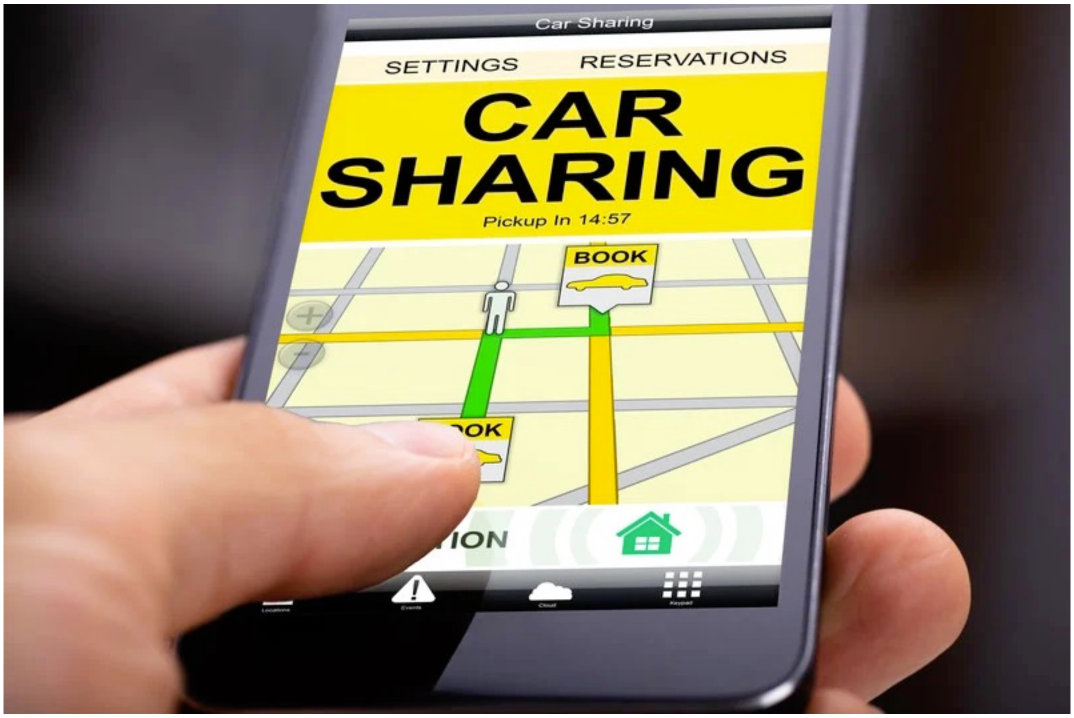
NEW CHALLENGES TO AUTO INSURANCE COMPANIES: DIFFICULTIES IN DETERMINING THE INSURANCE COVERAGE AND LACK OF POLICIES TO SUPERVISE BROUGHT BY THE RIDE-SHARE APPS



Source: Photo by Robin Olson

Prepared By

Shu Dong (David)

Department of Mathematics

University of Waterloo

Waterloo, Ontario

SUMMARY

The main pupose of this project was to focus on the specific changes in the policies regarding the supervision and classification of insurance coverage for transportation network companies and what reasons caused these changes to happen.

In order to be able to know the actual operating insurance fees the companies should afford to their drivers and customers and do the analysis efficiently, I decided to propose the survey for the 1000 drivers who work for the transportation network companies to check their insurance status before working for the transportation network companies and evaluate their personal driving habits.

By using and analyzing the gathered data regarding the driver's driving background and the amount of losses they caused, I realized that the insurance fees and the degree of supervision on the insurance coverage would lower if the drivers' for the transportation network companies had better driving habits and their losses for the accidents caused mostly perform in a low degree. Otherwise, it would increase the insurance fees and drivers would face more tightly regulations. Therefore, it meant that if there existed well insurance regulations and clear insurance coverage classifications to let the drivers maintain their good driving habits, it would minimize people's loss and their life could be protected to the most extent.

TABLE OF CONTENTS

LIST OF TABLE AND FIGURES

TABLES

FIGURES

GLOSSARY

Term	**Definition**
RStudio (R)	Rstudio is a development environment for the language R and is used for statistical computing and graphics.
X-axis	X-axis is the horizontal axis under the plane rectangular coordinate system.
Y-axis	Y-axis is the vertical axis under the plane rectangular coordinate system.
Two-way factorial design	Two-way factorial design is the statistics experiment method that designs an experiment that involves two factors with a set of treatments.
ANOVA table	ANOVA table is the statistical computing technique that can separate the components of variation between the treatments and error.
F-value	F-value is the value that can be obtained from the calculation in the regression analysis if there exist significantly different means between two populations.
Level of significance	The level of significance is the statistical method that can use to measure the statistical significance.
Critical value	The critical value is a line that can separate the graph into sections.

1.0 INTRODUCTION

The emergence of the new electronic platform for the ride-share brings new problems related to auto insurance which challenge the traditional auto insurance industry. This project mainly focuses on the specific changes in the policies which can remedy the defects regarding the original supervision and the new determination of insurance coverage for transportation network companies and what reasons caused these changes to happen.

In order to make a reasonable insurance coverage classification, all the essential information, such as driving background and original insurance coverage status regarding the drivers who work for the transportation network companies, should be collected. By using the data sets taken from the proposed survey of 1000 drivers who use the ride-share apps, the number of people who experienced each specific risk category is gathered (Table 1) and then put these data in the risk analyzing model. Compared with the historical losses caused by the risk variable data, the risk analyzing model can generate a trend graph for these data which can show how risky each driving habit can be and use this result to determine the specific insurance coverage for each individual driver.

Table 1: Survey for the ride-share apps' drivers' driving habits

The driving scenario which may cause risk	**Number of drivers who experienced the scenario**
Exceeded the speed limit more than 10 times a year	376
Failed the written driving test more than three times	34
Caused more than 5 accidents each year	173
Had the driving experience for less than 1 year	47
Experienced aggressive driving more than 10 times	259
Driving in poor weather more than 10 times a year	915
Experienced pedestrian accidents	21

Additionally, the loss data should be collected under both the previous policies' regulation and the new policies' supervision to compare to know which categories of the policies can have the most effective way to minimize the losses and protect both the drivers' and customers' life to the most extent. Therefore, the auto insurance industry can use the degree of losses caused by government policies’ regulations and drivers' driving habits to minimize the insurance companies’ losses and design new auto insurance products that meet the current market needs.

There are two main objectives that the project wants to achieve:

- Used the data gathered from the driving habits survey to analyze each driver who works for the transportation network companies degree of the risk and determine how strict supervision and insurance coverage they should receive.
- Compared and analyzed the performance of which government policies’ regulations can minimize the amount of losses and protect the driver’s and customers’ life to the most extent in order to design suitable auto insurance products which can meet the current market needs regarding minimizing both insurance companies and insured’s losses.

2.0 RISK ANALYSIS AND INSURANCE COVERAGE DETERMINATION FOR DRIVERS

The risk analysis is based on the data collected by the proposed survey for 1000 drivers who work for the transportation network companies, and all the graphs and trends are generated by using the R operation which is referred to in Appendix A. As shown in Figure 1, the direct graph which represent the number of drivers who experienced the corresponding scenario can be presented and I am able to know how serious for each risk catagary can be from the bar plot.

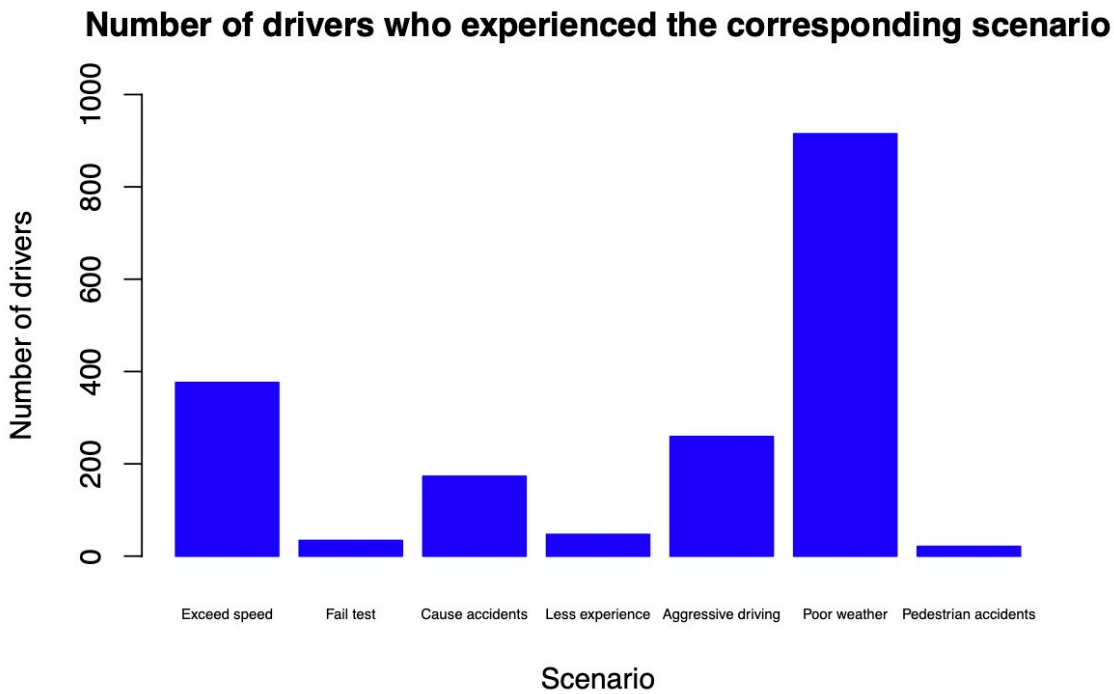


Figure 1: Bar plot for the number of drivers who experienced the corresponding scenario

Since Lehmann (2014) interpreted the risk index for the scenario described above (Table 2), I can use the collected data set for the survey to compare with the degree of the risk for each risk scenario in order to know how severe each risk scenario can be. The more severe the scenario achieves, the more losses will be caused. Therefore, it means that I can use the serious degree for each scenario to determine which kinds of insurance coverage they should receive.

Table 2: Risk Index for each corresponding scenario

The driving scenario which may cause risk	Risk Index
Exceeded the speed limit more than 10 times a year	7

Failed the written driving test more than three times	4
Caused more than 5 accidents each year	9
Had the driving experience for less than 1 year	6
Experienced aggressive driving more than 10 times	5
Driving in poor weather more than 10 times a year	3
Experienced pedestrian accidents	8

The analysis process can be broken into two sections: assumption for the relationship between the types of risky scenarios and risk index for the corresponding risky scenario and determination for the insurance coverage by calculating.

2.1 Assumption for the Relationship between the Types of Risky Scenarios and Risk Index

Figure 2 shows the scatter plot for the risk index and the number of drivers.

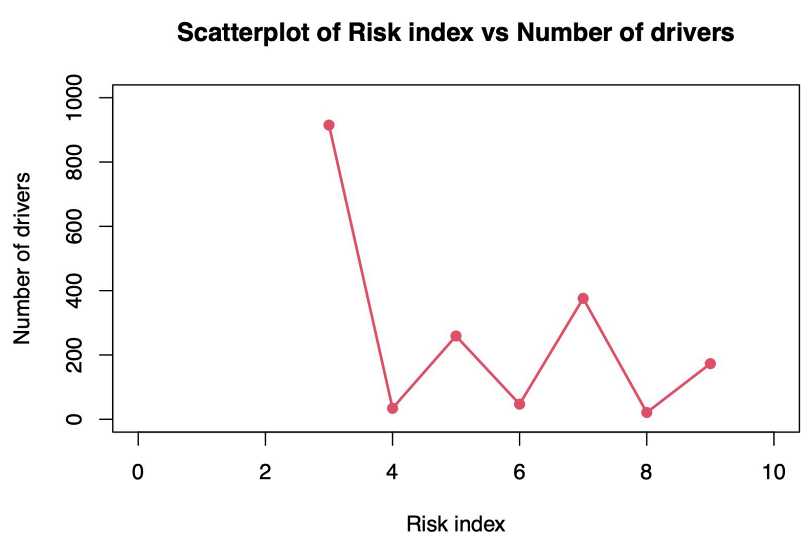


Figure 2: Scatter plot of risk index vs. number of drivers

From the Figure 2, I find that the if the risk index is smaller, the point will mostly locate with the higher location on the Y-axis which represents the number of drivers who experienced the corresponding scenario. It means that I can make an assumption that if the risk index for the specific risky scenario is lower, this type of risky scenario may have a higher possibility to occur during the working time for the drivers who work for the transportation network companies.

2.2 Determination for the Insurance Coverage by Calculating

In order to evaluate whether there actually exists a significant interaction effect for the assumption that I make above, I decide to use the experimental statistics method called two-way factorial design since I realize that there simultaneously exists the effects of the two factors which are the types of the risky scenario and risk index corresponding to each risky scenario in this experiment. Thus, I generate a new data set (Table 3) in order to do the statistics calculation.

Table 3: Data of doing the two-way factorial design for types of risky scenarios and risk index

	Risk Index for the corrsoponding risky scenario						
Risky Scenario	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Poor weather	915	0	0	0	0	0	0
Fail Test	0	34	0	0	0	0	0
Aggressive Driving	0	0	259	0	0	0	0
Less experience	0	0	0	47	0	0	0
Exceed speeds	0	0	0	0	376	0	0
Pedestrian accidents	0	0	0	0	0	21	0
Cause accidents	0	0	0	0	0	0	173

ANOVA table (Table 4) can be generated, and the calculation process showed in Appendix B. From the ANOVA table, I can know that the F-value for the interaction effect is 2.396077.

Table 4: ANOVA table for two-way factorial design

	Sum of Squares	Degrees of Freedom	Mean Squares	F0
Variation for A	86230.49	6	14371.748	1.477608
Variation for B	86230.49	6	14371.748	1.477608
Variation for AB	838984.08	36	23305.113	2.396077
Variation for Error	953183.40	98	9726.361	NA
Variation for Total	1964628.46	110	NA	NA

Therefore, assuming the level of significance is 95%, by the principle of the hypothesis test for a significant interaction effect between the two factors, I can just use the F-value which is calculated in Appendix B to compare with the critical value which calculated in Appendix C to determine whether I should reject the null hypothesis at the level of significance which is 95%. The principle of the hypothesis test describes that if the F-value is greater than its critical value, then the null hypothesis should be rejected at the specific level of significance and it means that there exists a significant influence between the two factors. Otherwise, the null hypothesis should not be rejected, and there does not exist a significant relationship. According to the calculation process for the critical value shown in Appendix C, I can get the critical value of 1.53753 at the level of significance of 95%. Since the F-value is 2.396077 which is greater than the critical value of 1.53753, it shows that we need to reject the null hypothesis at the level of significance of 95%. Due to the rejection of the null hypothesis, it means that there is a significant interaction effect at the 95% level of significance.

Finally, the conclusion can be made owing to the result of statistics calculation which reveals that more serious results and more losses regarding property and life can be caused if drivers who work for ride-share apps has been commonly experienced the dangerous situation of driving with a higher risk index. In a nutshell, auto insurance companies should classify the insurance coverage by supervising the drivers' daily driving habits and let the coverage proportion become malleable in order to be able to decide how extent the insurance companies should be responsible for at any time to not only protecting the property and life for both drivers and customers but also minimizing the losses for the auto insurance companies.

3.0 ANALYZING THE CHANGING FOR THE GOVERNMENT REGULATION POLICIES

To decide and evaluate whether the original and recent government supervision policies can effectively monitor the defects of the prescription and minimize the people's losses regarding both life and property, I use the typical cases and the latest peer-viewed report to analyze whether the changing for the policies can meet the public expectations and can remedy the loopholes of stipulation with respect to supervise the implementation effects of protecting the people when they experienced the car accidents.

According to Shields (2016), a severe pedestrian accident happened to a little girl caused by a ride-share app driver without any protection from auto insurance which caused vast property and health losses and resulted in a permanent injury for the little girl. The occurrence of this tragedy is due to the unclear and unforced regulations for insurance policies regarding insurance coverage during the period of the just emergence of new ways of doing the ride-share. In order to avoid the happening of the same situation, since the drivers are initially not covered by the transportation network company's commercial insurance policy if they use their car for commercial purposes, the government implements the approach that the drivers who log onto the transportation network companies' app have to buy the commercial insurance first. Also, after the negotiation with the transportation network companies, the transportation network companies also ask their driver to prevent using their own personal insurance policy. It advocates differential insurance requirements during the three stages of the drivers' work, letting the transportation network companies share the responsibility of the coverage with the drivers.

Fortunately, the changing regarding the government's regulations actually performed effectively and minimized the people's loss to a certain degree. As stated by Colier et al. (2018), the report pointed that the government had already implemented a mandatory policy for the transportation network company to let them have to meet all the requirements of insurance coverage for the ride-share. It lets the transportation network company has to maintain valid commercial liability insurance and give the minimum liability fund to the government in order to get the insurance certificate. Additionally, the government requires transportation network companies to provide commercial insurance coverage to all their drivers in order to be able to cover any portion of the claim. Due to the implementation of these mandatory policies, people's property and life are more likely to be protected and the amount of losses caused by car accidents that occur by operating ride-share services has decreased sharply in recent years.

Therefore, as discussed above, it shows that if the government persists in implementing the compulsive insurance coverage provision and the same intensive monitoring policies for the insurance requirements for the transportation network companies, both losses for the insurance companies and people who declare the policies can be minimized.

4.0 CONCLUSION

In order to face the new challenges brought by the emergence of new types of ride-share ways owing to the appearance of electronic techniques, auto insurance companies should implement clear boundaries in determining whether the drivers should buy personal or commercial insurance like the drivers who work for the transportation network companies and monitor the drivers' daily driving situation at any operation time to be able to change their insurance coverage constantly to reduce the losses for auto insurance companies at the greatest extent. After doing the calculation process for analyzing the data set of the proposed survey using the statistical method, I realized the importance of constantly knowing the current drivers' driving situation in determining insurance coverage.

However, the expectation described above cannot be realized without coercive administration control in implementing the insurance requirements for the transportation network companies. Thus, it means that government regulation for insurance also plays a crucial role in protecting people's property and life.

Due to the changes in the policies, it stimulates the disruptive innovation of insurance coverage determination in the traditional auto insurance industry and provokes people's understanding of the importance of insurance. The research may inspire the next generation to study new technical insurance in order to face the challenge caused by the rapid development of technology in the traditional insurance market and implement timely policy responses to the emerging market as long as it becomes popular.

Also, I realize that this research strongly connects with the career I want to pursue in the future. With the increasing number of ride-share apps, more and more people will choose ride-share as the standard way to travel. However, people's personal and property safety cannot be protected if poor insurance regulations and unclear insurance coverage classifications exist. Minimizing people's loss and protecting their life are the expectations I insistently want to achieve.

REFERENCES

Collier, R. B., Dubal, V. B., & Carter, C. L. (2018). Disrupting regulation, regulating disruption: The Politics of Uber in the United States: Perspectives on politics. *Cambridge Core*, 1-19. Retrieved from: https://www.rstreet.org/wp-content/uploads/2018/04/RSTREET28-1.pdf

Lehmann, R. J. (2021). Blurred lines: Insurance challenges in the ride-sharing market. *R Street*, 1-14. Retrieved from: https://www.cambridge.org/core/journals/perspectives-on-politics/article/disrupting-regulation-regulating-disruption-the-politics-of-uber-in-the-united-states/B7691A244DDF05EF1AE23FE82326FC04

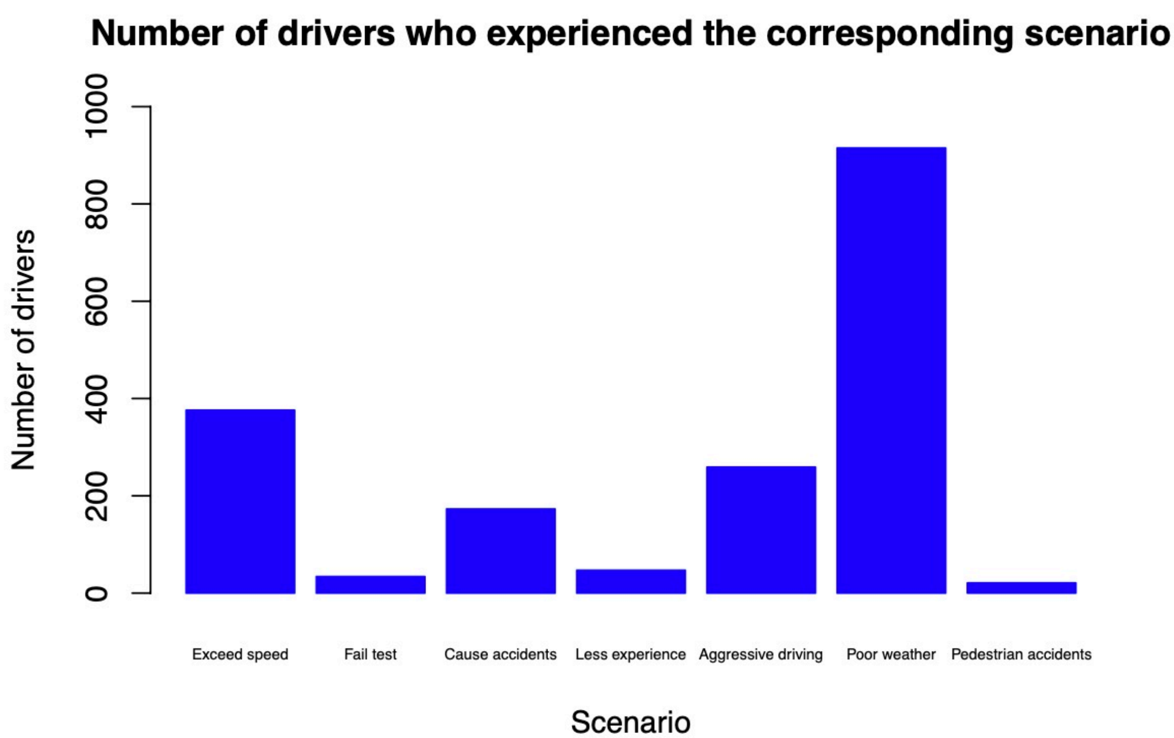
Shields, L. (2016). Driving decision-making: An analysis of policy diffusion and its role in the development and implementation of ridesharing regulations in four Canadian municipalities. *Scholarship@Western*, 1-49. Retrieved from: https://ir.lib.uwo.ca/lgp-mrps/150/

APPENDIX A: PROCESS FOR GENERATING THE GRAPHS AND TRENDS

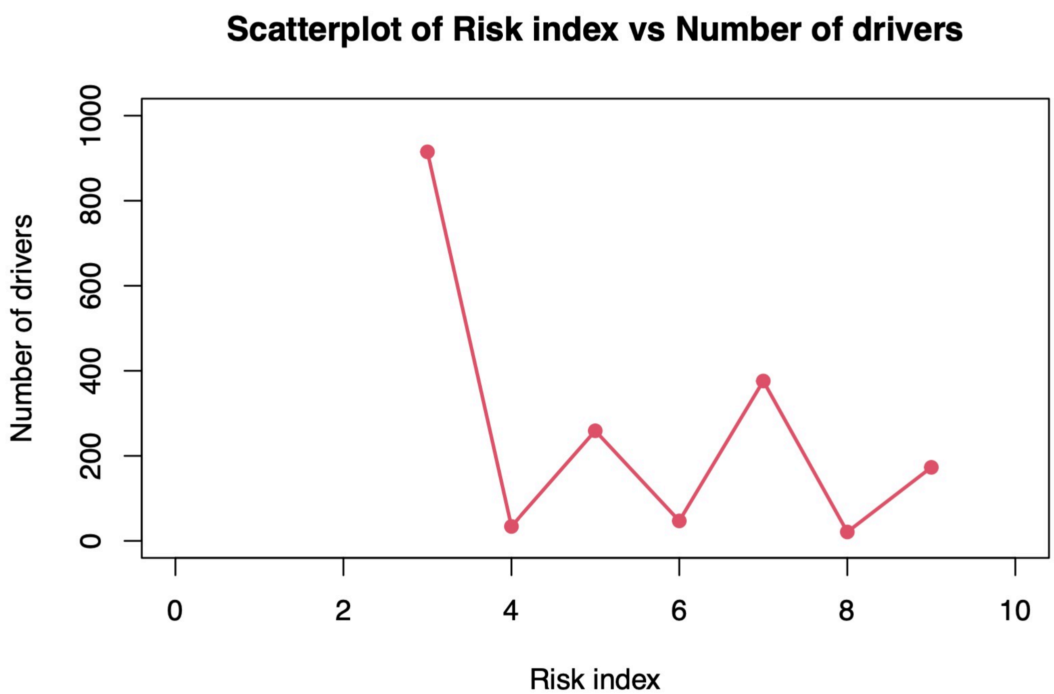
The driving scenario which may cause risk	**Number of drivers who experienced the scenario**	**Risk Index**
Exceeded the speed limit more than 10 times a year	376	7
Failed the written driving test more than three times	34	4
Caused more than 5 accidents each year	173	9
Had the driving experience for less than 1 year	47	6
Experienced aggressive driving more than 10 times	259	5
Driving in poor weather more than 10 times a year	915	3
Experienced pedestrian accidents	21	8

```r
# Import the Survey data in R
survey = read.csv("SURVEY.csv", header = TRUE)
scenario = survey$Scenario
number_drivers = survey$Number.of.experience
risk_index = survey$Risk.Index
data_survey = data.frame(scenario, number_drivers, risk_index)
```

```r
# Plot the bar plot for the number of drivers
barplot(number_drivers,
        xlab = "Scenario",
        ylab = "Number of drivers", ylim = c(0, 1000),
        names.arg = scenario, cex.names = 0.53,
        col = "blue", border = "blue",
        main = "Number of drivers who experienced the corresponding scenario")
```



```r
# Scatterplot for the risk index vs number of drivers
plot(risk_index, number_drivers,
     ylab = "Number of drivers", ylim = c(0, 1000),
     xlab = "Risk index", xlim = c(0, 10),
     main = "Scatterplot of Risk index vs Number of drivers",
     pch = 19, col = 2)
lines(sort(risk_index), c(915, 34, 259, 47, 376, 21, 173),
      col = 2, lwd = 2, pch = 1)
```



APPENDIX B: PROCESS FOR DOING THE ANOVA TABLE CALCULATION

```
# Let the risky scenario be treatment A and let risk index be treatment B
a = 7
b = 7
n = 1

y_xxx = (1 / (a * b * n)) * (915 + 34 + 259 + 47 + 376 + 21 + 173)
y_1xx = (1 / (b * n)) * 915
y_2xx = (1 / (b * n)) * 34
y_3xx = (1 / (b * n)) * 259
y_4xx = (1 / (b * n)) * 47
y_5xx = (1 / (b * n)) * 376
y_6xx = (1 / (b * n)) * 21
y_7xx = (1 / (b * n)) * 173
y_x1x = (1 / (a * n)) * 915
y_x2x = (1 / (a * n)) * 34
y_x3x = (1 / (a * n)) * 259
y_x4x = (1 / (a * n)) * 47
y_x5x = (1 / (a * n)) * 376
y_x6x = (1 / (a * n)) * 21
y_x7x = (1 / (a * n)) * 173
y_11x = (1 / n) * 915
y_21x = (1 / n) * 0
y_31x = (1 / n) * 0
y_41x = (1 / n) * 0
y_51x = (1 / n) * 0
y_61x = (1 / n) * 0
y_71x = (1 / n) * 0
y_12x = (1 / n) * 0
y_22x = (1 / n) * 34
y_32x = (1 / n) * 0
y_42x = (1 / n) * 0
y_52x = (1 / n) * 0
y_62x = (1 / n) * 0
y_72x = (1 / n) * 0
y_13x = (1 / n) * 0
y_23x = (1 / n) * 0
y_33x = (1 / n) * 259
y_43x = (1 / n) * 0
y_53x = (1 / n) * 0
y_63x = (1 / n) * 0
y_73x = (1 / n) * 0
y_14x = (1 / n) * 0
y_24x = (1 / n) * 0
y_34x = (1 / n) * 0
y_44x = (1 / n) * 47
y_54x = (1 / n) * 0
y_64x = (1 / n) * 0
y_74x = (1 / n) * 0
y_15x = (1 / n) * 0
y_25x = (1 / n) * 0
y_35x = (1 / n) * 0
y_45x = (1 / n) * 0
y_55x = (1 / n) * 376
```

```
y_65x = (1 / n) * 0
y_75x = (1 / n) * 0
y_16x = (1 / n) * 0
y_26x = (1 / n) * 0
y_36x = (1 / n) * 0
y_46x = (1 / n) * 0
y_56x = (1 / n) * 0
y_66x = (1 / n) * 21
y_76x = (1 / n) * 0
y_17x = (1 / n) * 0
y_27x = (1 / n) * 0
y_37x = (1 / n) * 0
y_47x = (1 / n) * 0
y_57x = (1 / n) * 0
y_67x = (1 / n) * 0
y_77x = (1 / n) * 173

SS_A = b * n * (((y_1xx - y_xxx)^2) + ((y_2xx - y_xxx)^2) +
                 ((y_3xx - y_xxx)^2) + ((y_4xx - y_xxx)^2) +
                 ((y_5xx - y_xxx)^2) + ((y_6xx - y_xxx)^2) +
                 ((y_7xx - y_xxx)^2))
SS_B = a * n * (((y_x1x - y_xxx)^2) + ((y_x2x - y_xxx)^2) +
                 ((y_x3x - y_xxx)^2) + ((y_x4x - y_xxx)^2) +
                 ((y_x5x - y_xxx)^2) + ((y_x6x - y_xxx)^2) +
                 ((y_x7x - y_xxx)^2))
SS_AB = n * (((y_11x - y_1xx - y_x1x + y_xxx)^2) +
  ((y_21x - y_2xx - y_x1x + y_xxx)^2) + ((y_31x - y_3xx - y_x1x + y_xxx)^2) +
  ((y_41x - y_4xx - y_x1x + y_xxx)^2) + ((y_51x - y_5xx - y_x1x + y_xxx)^2) +
  ((y_61x - y_6xx - y_x1x + y_xxx)^2) + ((y_71x - y_7xx - y_x1x + y_xxx)^2) +
  ((y_12x - y_1xx - y_x2x + y_xxx)^2) + ((y_22x - y_2xx - y_x2x + y_xxx)^2) +
  ((y_32x - y_3xx - y_x2x + y_xxx)^2) + ((y_42x - y_4xx - y_x2x + y_xxx)^2) +
  ((y_52x - y_5xx - y_x2x + y_xxx)^2) + ((y_62x - y_6xx - y_x2x + y_xxx)^2) +
  ((y_72x - y_7xx - y_x2x + y_xxx)^2) + ((y_13x - y_1xx - y_x3x + y_xxx)^2) +
  ((y_23x - y_2xx - y_x3x + y_xxx)^2) + ((y_33x - y_3xx - y_x3x + y_xxx)^2) +
  ((y_43x - y_4xx - y_x3x + y_xxx)^2) + ((y_53x - y_5xx - y_x3x + y_xxx)^2) +
  ((y_63x - y_6xx - y_x3x + y_xxx)^2) + ((y_73x - y_7xx - y_x3x + y_xxx)^2) +
  ((y_14x - y_1xx - y_x4x + y_xxx)^2) + ((y_24x - y_2xx - y_x4x + y_xxx)^2) +
  ((y_34x - y_3xx - y_x4x + y_xxx)^2) + ((y_44x - y_4xx - y_x4x + y_xxx)^2) +
  ((y_54x - y_5xx - y_x4x + y_xxx)^2) + ((y_64x - y_6xx - y_x4x + y_xxx)^2) +
  ((y_74x - y_7xx - y_x4x + y_xxx)^2) + ((y_15x - y_1xx - y_x5x + y_xxx)^2) +
  ((y_25x - y_2xx - y_x5x + y_xxx)^2) + ((y_35x - y_3xx - y_x5x + y_xxx)^2) +
  ((y_45x - y_4xx - y_x5x + y_xxx)^2) + ((y_55x - y_5xx - y_x5x + y_xxx)^2) +
  ((y_65x - y_6xx - y_x5x + y_xxx)^2) + ((y_75x - y_7xx - y_x5x + y_xxx)^2) +
  ((y_16x - y_1xx - y_x6x + y_xxx)^2) + ((y_26x - y_2xx - y_x6x + y_xxx)^2) +
  ((y_36x - y_3xx - y_x6x + y_xxx)^2) + ((y_46x - y_4xx - y_x6x + y_xxx)^2) +
  ((y_56x - y_5xx - y_x6x + y_xxx)^2) + ((y_66x - y_6xx - y_x6x + y_xxx)^2) +
  ((y_76x - y_7xx - y_x6x + y_xxx)^2) + ((y_17x - y_1xx - y_x7x + y_xxx)^2) +
  ((y_27x - y_2xx - y_x7x + y_xxx)^2) + ((y_37x - y_3xx - y_x7x + y_xxx)^2) +
  ((y_47x - y_4xx - y_x7x + y_xxx)^2) + ((y_57x - y_5xx - y_x7x + y_xxx)^2) +
  ((y_67x - y_6xx - y_x7x + y_xxx)^2) + ((y_77x - y_7xx - y_x7x + y_xxx)^2))
SS_Err = ((915 - y_xxx)^2) + ((34 - y_xxx)^2) + ((259 - y_xxx)^2) +
  ((47 - y_xxx)^2) + ((376 - y_xxx)^2) + ((21 - y_xxx)^2) + ((173 - y_xxx)^2)
SS_Tot = SS_A + SS_B + SS_AB + SS_Err
```

```r
# Construct the ANOVA table for two-way factorial design
df_A = a - 1
df_B = b - 1
df_AB = (a - 1) * (b - 1)
df_Err = a * b * (n + 1)
df_Tot =df_A + df_B + df_Err
MS_A = SS_A / df_A
MS_B = SS_B / df_B
MS_AB = SS_AB / df_AB
MS_Err = SS_Err / df_Err
F0_A = MS_A / MS_Err
F0_B = MS_B / MS_Err
F0_AB = MS_AB / MS_Err

ss = c(SS_A, SS_B, SS_AB, SS_Err, SS_Tot)
df = c(df_A, df_B, df_AB, df_Err, df_Tot)
ms = c(MS_A, MS_B, MS_AB, MS_Err)
F0 = c(F0_A, F0_B, F0_AB)

max_ln = length(ss)
name_row = c("Variation for A", "Variation for B", "Variation for AB",
             "Variation for Error", "Variation for Total")
name_col = c("Sum of Squares", "Degrees of Freedom", "Mean Squares", "F0")

ANOVA_data = setNames(data.frame(col1 = c(ss, rep(NA, max_ln - length(ss))),
                      col2 = c(df, rep(NA, max_ln - length(df))),
                      col3 = c(ms, rep(NA, max_ln - length(ms))),
                      col4 = c(F0, rep(NA, max_ln - length(F0))),
                      row.names = name_row), name_col)

library(knitr)
knitr::kable(ANOVA_data, "pipe")
```

	Sum of Squares	Degrees of Freedom	Mean Squares	F0
Variation for A	86230.49	6	14371.748	1.477608
Variation for B	86230.49	6	14371.748	1.477608
Variation for AB	838984.08	36	23305.113	2.396077
Variation for Error	953183.40	98	9726.361	NA
Variation for Total	1964628.46	110	NA	NA

APPENDIX C: CALCULATION PROCESS FOR THE CRITICAL VALUE

```
# Find the critical value
a_c = 7
b_c = 7
n_c = 1
df_AB_c = (a_c - 1) * (b_c - 1)
df_Err_c = a_c * b_c * (n_c + 1)
alpha_s = 1 - 0.95
qf(alpha_s, df_AB_c, df_Err_c, lower.tail = FALSE)
```

```
## [1] 1.53753
```