Does aggressive driving result in more car accidents?
Experimental Design
Prepared by: Prashant Sanghal

SIADS 688 Assignment 4: Analysis Plan

Section 1. Introduction (3 points)

Write a 200-word introduction that defines the research topic and answers the question, "Why does this topic matter?" Read the Introduction section of the relevant papers we have covered to learn how to write an effective introduction.

In this section, you should do the following:

1. **(1 point) Establish the problem leading to the study. Please narrow your topic as much as you can.**
2. **(1 point) At some point in your discussion, identify the questions you will try to answer through your research. Make these as concrete and focused as possible, so that your reader can figure out exactly what you want to learn.**
3. **(1 point) The question "Why does this topic matter?" implies the sub-question "To whom does this matter?" What's your audience, and why should they care?**

As per road rage statistics, 66% of the traffic fatalities are caused by aggressive driving behaviours such as speeding, tailgating (driving too close to a car in front), road rage, traffic congestion, drunk driving etc. as reported by (NHTSA) National Highway Traffic Safety Administration. This is a significant number and imposes negative externalities on other drivers *(week1-lecture Dr.Chen)* which can hurt many lives on the road, and probably can be saved if we can find an effective way of controlling aggressive driving behaviour before it happens. Hence, this study is motivated to conduct an experiment and see what would happen if our cars were equipped with a sensing device which could detect inappropriate driving behaviour beforehand, and provide real-time feedback to the drivers. Would it contribute to lower car accidents in real traffic conditions, and if yes then by how much, before and after receiving the feedback? One of the assumptions, in this study is that real-time feedback treatment effect should provide a long term and sustainable solution to this problem when compared to just receiving a technical report in the mail of previous traffic violations and aggressive behaviour management tutorials from the police department.

The target audience in this case study would be the drivers who have a history of traffic violations in the past 24 months with a minimum of 1 accident. They can be either males or females, with or without self-aggression status in the age group of 18-45 and 46-60 years old, randomly assigned to control and treatment groups in this study.

The participants of this study would be given a free sensing device at no cost to them as an incentive for participating in the study regardless of which ever group they are assigned to.

“One aggressive driver changed, can save two car accidents”

Section 2. Literature Review (3 points)

Write a 300-word literature review for your chosen topic that answers the question, “What do we already know about this topic, and what do we still need to find out?” The purpose of the literature review is to understand how the new knowledge you will be acquiring fits into what is already known.

You need to discuss at least 3 items from the literature. You can start with the classic papers in an area of research, then do a Google Scholar search to find the most relevant papers.

Please avoid writing three paragraphs, each summarizing one paper.

There were a number of studies conducted in the past to understand aggressive driving behavior among people. For example, factors that could trigger aggressive driving behaviours are getting stuck in the traffic for long hours, cutting lanes frequently, driving to work late, driving too close to a vehicle in front etc. It was interesting to note that men were more likely than women to commit aggressive driving. Also, drivers in the age group of 45 years old or older were less likely to drive aggressively than the younger drivers. By varying driving conditions (number of cars on the road, driver's age, day/time of driving) aggressive driving behaviour can vary and may respond differently to the treatments applied. *(David et al, 2004)* In one of the studies researchers used driving simulator as a treatment to detect variation in the aggressive behavior. They hired a group of university students (one with self-reported history of aggressive behaviour and other with non-aggressive behaviour) to go through a series of frustrating events in the driving environment related to other driver's behaviour or traffic conditions and then see how it impacts the driving behaviour right after watching the driving simulation? Even though drivers with self-reported aggression history were more likely to exhibit strong aggressive tendencies but the focus of this experimental set up was to test the extent of aggression it would trigger under different conditions *(Maya et al, 2011)*. An experimental study conducted in Beijing seems to correlate well with the effect of driving simulation results just discussed. The three most common aggressive driving behaviours found among this population was occupying bus lanes, frequently changing lanes and jumping queue at the crossing. It was found that all three behaviours were significantly higher during the rush-hour than those during the off-peak hours. While, occupying bus lane was found relatively fewer than the other two driving behaviours among the population but the presence of such driving behaviours were seemed to have put driver's safety at increased risk as well as was one of the primary cause behind ongoing traffic congestion in the city *(Qin H et al, 2013)*

Before we conduct our own experimental design, the three literatures discussed above provide sufficient evidence through past studies that driving conditions (example: traffic congestion, rush-hour, speeding) likely influence aggressive behavior among young male drivers who have a history of self-reported aggression when compared to other drivers. So, the experimental design question is, would real-time feedback help reduce car accidents among aggressive drivers and by how much?

Section 3. Analysis Plan (12 points)

This section should contain the following key points, written in coherent paragraphs:

1. Hypotheses (2 points): for each hypothesis, state both the null and the alternative hypothesis. For each hypothesis, state where it comes from, either theory or past empirical or experimental findings. Cite relevant papers which support each hypothesis.

2. **Data Description (4 points): This section should provide summary statistics and introduce the experimental variations (if applicable). You should answer the following questions:**
    1. **(a) Who are your subjects or what is your population? If applicable, provide summary statistics for an overview.**
    2. **(b) What are the treatment and control conditions? You may have multiple treatment groups, and control groups.**
    3. **(c) Are the experimental conditions randomly assigned? If applicable, test for randomization (Ref: Chen et al. 2017, Table 2 from Week 1).**
    4. **(d) What are the outcomes you observe? Interpret the outcomes you plan to measure.**
    5. **(e) What are the independent variables in your analysis?**
    6. **(f) What are the potential confounders in your analysis plan and how can you address it?**
3. **Empirical method (6 points): How will you test your hypothesis using the existing dataset? You should include:**
    1. **(a) A regression model with necessary explanations for the independent variables and coefficients.**
    2. **(b) The empirical strategy to estimate the model.**
    3. **(c) Preliminary result, if applicable.**

1. Hypotheses 1:

Null Hypothesis: Driving conditions do not impact aggressive driving behaviour among drivers.

Alternative Hypothesis: Driving conditions impact aggressive driving behaviour among drivers.

Above hypothesis is derived from *(Maya et al, 2011)*, where researchers ran a driving simulator experiment on a group of university students to assess whether change in driving conditions impacted aggressive behaviour among drivers (both aggressive and non-aggressive) or not. It was learned that by changing driving conditions (traffic congestion, time pressure, cutting lanes) increased aggressive driving tendencies among both aggressive and non-aggressive drivers.

Hypotheses 2:

Null Hypothesis: Real-time feedback on aggressive driving behaviour will not help reduce car accidents in every day driving conditions.

Alternative Hypothesis: Real time feedback on aggressive driving behaviour will help reduce car accidents in every day driving conditions.

In this reference, a study was published on how SVM classifier was utilized in a real traffic conditions to detect aggressive behavior with 93.1% accuracy. However, this study focusses on the technology aspect and does not quantify overall percentage reduction in the car accidents as a result of this study *(*Akar et al*, 2016).*

There were other articles , however, that suggest that lane departure warning system which would be one of the components of the above sensing device saw a 11% reduction in car crashes. *(Crash Statistics, ITS Deployment Evaluation, 2009-2015)*

2. Data Description

The participants of this study are registered drivers (males or females) in the age group between 18-45 and 46-60 years old with a history of traffic violations in the past 24 months with a minimum of 1 accident. To invite participants to this study, a letter describing the outcome of the study (i.e. controlling aggressive driving behaviour in every day driving to prevent a car crash) will be sent out in the mail. Participants who agree to participate in the study will be assigned to the control and treatment groups randomly.

Control Group: Contains drivers with no contact.

Treatment Group: Contains driver who will receive 2 treatments:

First treatment group will receive a technical report from the police department advising them to drive safely which would be based on their 2 year's worth of traffic violations, and a minimum 1 reported accident on their driver's log. This treatment will be coded as 'treat1'

Second treatment group will receive a sensing device fitted in the car which will "beep and display" whenever any aggressive driving behaviour (limited to speeding and tailgating) is detected while driving. This treatment will be coded as 'treat2' and the sensor log report will be available for future analysis on the cloud platform. Reason, this will help us understand how many time the sensor was activated during the experiment, and if there was any change in driver's aggressive behaviour over time. (i.e. is there a declining trend in sensor activation).

Sample Size Calculation:

Since, we do not have any prior data to conduct this experiment. We will use power analysis to determine the sample size for this study. It will help us determine smallest sample size that will be required before data collection that can detect the effect of test at the desired significance level, and hence minimize cost related to data collection.

Here are the parameters for power analysis:

Beta*(Power level)* = 0.*8 (i.e. 80% probability that type 2 error will not be committed and we will correctly reject null at significance level stated below)*

Alpha*(significance level)* = 0.05 *(i.e. 5% type 1 error to be tolerated)*

if p-value < 0.05 then test results found will be statistically significant at 5% and null can be rejected. Otherwise if pvalue>0.05 then test results will not be statistically significant at 5% and null cannot be rejected.

Effect Size:

Say, we assume effect size between 'treat1' and 'treat2', delta = sigma/5, and the reduction in the number of car accidents between the two treatment conditions, sigma = 0.2 or 20% *(I & Bos et al, 2000)*

Then using the formula

$$n_0^* = n_1^* = n^* = 2(t_{\alpha/2} + t_\beta)^2 \left(\frac{\sigma}{\delta}\right)^2$$

We will get:

2*(1.96+0.84)^2 * (0.2/(0.2/5))*2 = 392*5 = 1,960 for each control, and two treatment groups.

Hence, total sample size = 1960+1960+1960 = 5,880

Test for randomization:

Now, we will calculate p-values between the control and the pooled treatment groups (treat1 and treat2) for the list of covariates to see if p-value is less than 10% significance level or not. We expect that p-values will be greater than 10% significance level to say that our randomization worked. *(Chen et al. 2017, Table 2 from Week 1)*

Planned Outcomes:

Using total sample size of 5,880, we plan to run the experiment to test our 2 hypothesis. First, does participants demonstrate aggressive driving behaviour with the change in driving conditions. It's most likely as we have seen the driving simulator study discussed in literature review above. Secondly, if the first hypothesis is found statistically significant, then do we observe any variation (reduction) in car crashes over one-year period when compared to the controlled group condition.

Independent Variables:

This may include driver's information such as Opened Letter (Yes, No), Gender (Male, Female), Age Group (18-45, 46-60), History of Aggression (Yes, No), Driving Condition (peak_on, peak_off), Observed Aggression (Yes, No).

In addition, we will also include Violations Before Treatment (over 2 years period) and Accidents Before Treatment (minimum 1, over 2 years period) as a part of controlled group collection. Similarly, we will also gather Violations After Treatment and Accidents After Treatment as a part of treatment group collection. This will help us calculate "Violations Difference" and "Accidents Difference" on an average for Year 1 before and after the treatment is applied. We will split Year 1 in to 6 months and 12 months period to capture trend.

Furthermore, we assume that driving conditions during 'peak_on' hours in a real traffic condition will naturally increase traffic congestion, jumping lanes, tailgating behaviour when compared to peak_off hours. Hence, during this time we will try to observe through sensing device specifically 'treat2' condition whether sensor picked up any aggressive driving behaviour (speeding, tailgating) from the driver or not. However, getting the same information for 'treat1' and 'control' conditions will be dependent on driver's own behaviour and will be very difficult for the researchers to capture it otherwise, unless they use a camera to monitor irregular (car speeds or movement) as a feature to record behaviour variation.

Please note, if we decide to use a camera as a monitoring tool, then drivers in 'control' and 'treat1' conditions will be informed in order to remain compliant with 'treat1' and 'control' design conditions i.e camera will only record driver's aggressive behaviour but will not have any sensor-like features ('beep & display') which only applicable for 'treat 2' condition.

Dependent Variables:

Join Study (Yes, No), Average Violations Difference (6 Months), Average Violations Difference (12 Months), Average Accidents Difference (6 Months), Average Accidents Difference (12 Months).

By studying above dependent variables, we will be able to explore whether decrease in violations also lower car accidents or not?

Possible Confounders:

Weather could be a possible confounder that can influence driving behaviour. For example, during slippery road conditions due to snow or rain, the cars can collide even when drivers are not driving aggressively.

Self-mood could be another confounder that can force drivers to be upset about something else at work or home, and not particularly angry about traffic conditions or how others are driving on the road. Hence, emotional status before they drive can also trigger aggression not related to driving but other personal factors.

By randomly assigning participants to control and treatment groups, we would be able to minimize the effect of confounders on our study. As well as, by avoiding doing this experiment during extreme weather conditions, we can also minimize the impact of weather on driving conditions. The goal of this experiment is to create awareness among drivers that how aggressive driving is putting their life and lives of others at high risk.

3. Regression Model:

We will use instrument variable estimation using two stage regression model (IV2SLS) to estimate difference in violations and accidents before and after receiving the treatments. But first, we would need to check whether our instrument variable "Mail" is a valid instrument or not. Here are the two properties which our IV will need to satisfy:

a) Inclusion restriction: During First Stage Regression, does F-statistics between outcome variable "Join Study" and our instrument variable "Mail" is greater than 10 or not?
b) Exclusion restriction: Secondly, does instrument variable "Mail" have any direct impact on **other outcome variables** as under. That is F-statistics for outcome variables is less than 10 or not?

Average Violations Difference (6 Months), Average Violations Difference (12 Months), Average Accidents Difference (6 Months), Average Accidents Difference (12 Months)

To satisfy exclusion restriction, we can use previous study as a resource *(I & Bos et al, Table 2)* to regress each outcomes variable (as shown in red italics) using OLS regression on both control and treatment conditions ("treat1", "treat2") to check whether F-statistics is under 10 and if p-values for "treat1" & "treat2" are statistically significant at given significance level or not, before conducting second stage regression using IV2SLS regression model. The coefficients of "treat1" and "treat2" will be used as p-values for this assessment. The coefficient for control condition will likely be greater than the significance level.

As noted earlier, previous studies suggested 20% reduction in accidents due to in-car data recorders. It was also noted from driving simulation experiment that change in driving conditions increase aggressive behaviour among drivers. Hence, running above OLS regression using data from these studies, will be helpful in proving that alternate hypothesis indeed exists, and that change in driving conditions and real-time feedback will produce statistically significant results suggesting reduction in both traffic violations and car accidents, not directly linked to the mails sent to the drivers.

Exogeneous variables will be all our "**Independent Variables**" as shown below including a constant

Opened Letter (Yes, No), Gender (Male, Female), Age Group (18-45, 46-60), History of Aggression (Yes, No), Driving Condition (peak_on, peak_off), Observed Aggression (Yes, No), Violations Before Treatment (over 2 years period) and Accidents Before Treatment (minimum 1, over 2 years period), constant

Endogenous variables will be "**Other outcome variables**" as shown above in red italics.

Instrument Variable will be Mail including a constant.

Summary of Regression Model: "smf", "ols" , "gnn" are from statsmodel library.

First Stage Regression = smf.ols('Join Study~Mail' , data).fit() require F-stage >10 to satisfy inclusion restriction

Pre-Second Stage: smf.ols('Average Differences in Traffic Violations (6 & 12 months) ~ control+treat1+treat2, data_previous_study).fit() will require F-Statistics <10, and p-values<0.05 for both alternate hypothesis to exist and satisfy exclusion restriction.

Second stage Regression: We will run the model as under:

Model = gmm**.**IV2SLS(endog, exog, instrument)**.**fit() for each endogenous variable to note the coefficients of average difference in the traffic violations and car accidents over 6 to 12 months period to see if there is a declining pattern or not.

Section 4. Conclusions and Limitations (1 point)

This experiment will help us assess how we can effectively reduce aggressive behaviour among drivers while driving in a day-to-day real traffic conditions. Which strategies would be very beneficial?

A) Control (Do nothing)

B) Apply Treatment (Send Technical Report) or Install Sensing device in the car which can beep and display aggressive driving behaviour such as speeding or tailgating to the driver.

Based on previous studies and based on the test results (when data is collected), we are hopeful that we will see more than 20% reduction in car accidents as well as traffic violations as we include other sensing features such as "road rage" and "drunk driving" to our list.

However, one of the main limitations of this study, we believe would be seen when car sensors malfunctions and detects aggressive behaviour incorrectly (example in slow moving bumper to bumper traffic), if sensor starts to beep then it can be seen as a distraction than a useful tool. Hence, it's really important that we fine tune the sensors at par with the real-traffic conditions to avoid noise in the experiment.

Having said above, we feel that this experiment can help save many lives that are lost as a result of aggressive driving every day. Sometimes all we need is a reminder to 'Stop' which works for many of us.

References (1 point)

This section should contain at least the three references used in your literature review.

1. Shinar, David & Compton, Richard. (2004). Aggressive Driving: An Observational Study of Driver, Vehicle, and Situational Variables. Accident; analysis and prevention. 36. 429-37. 10.1016/S0001-4575(03)00037-X.

2. Abou-Zeid, Maya & Kaysi, Isam & Al-Naghi, Hani. (2011). Measuring Aggressive Driving Behaviour using a driving simulator: A exploratory Study.

3. Qin H., Zhu H., Huang R. (2013) Study on Aggressive Driving Activities at Crossroads in Beijing. In: Rau P.L.P. (eds) Cross-Cultural Design. Cultural Differences in Everyday Life. CCD 2013. Lecture Notes in Computer Science, vol 8024. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39137-8_36

4. Kumtepe, Omurcan & Akar, Gozde & Yuncu, Enes. (2016). Driver aggressiveness detection via multisensory data fusion. EURASIP Journal on Image and Video Processing. 2016. 10.1186/s13640-016-0106-9.

5. Wouters, I & Bos, John. (2000). Traffic accident reduction by monitoring driver behaviour with in-car data recorders. Accident; analysis and prevention. 32. 643-50. 10.1016/S0001-4575(99)00095-0.