Paper 873-2017

Auto Telematics: Deviations Drive Success

Kelsey Osterloo, State Farm Insurance and Deovrat Kakde, SAS Institute Inc.

ABSTRACT

The use of telematics data within the insurance industry is becoming prevalent as insurers use this data to give discounts, categorize drivers, and provide feedback to improve customer's driving. The data captured through in-vehicle or mobile devices includes acceleration, braking, mileage, and others. Data elements are analyzed to determine 'unsafe' events such as rapid acceleration, hard braking, quick turning, etc. The time between these successive unsafe events is often a function of the mileage driven and time in the telematics program.

Our discussion highlights analysis of unsafe events using recurrent event data analysis techniques available in the RELIABILITY procedure within the SAS/QC® software. The RELIABILITY procedure is used to determine a mean cumulative function (MCF) of unsafe events as a function of mileage, which describes the average driving behavior. The MCF is compared with each driver's cumulative history function of unsafe events to rank their driving behavior. We also discuss the utility of analysis results for providing drivers feedback about their driving.

SECTION 1: INTRODUCTION

State Farm is a leading insurance provider of Auto insurance. As part of the ever-changing demands of the insurance market over the years, State Farm began offering a telematics program (Drive Safe & Save™) in 2012. Under this voluntary program, a driver's driving behaviors are captured through an in-vehicle device or more recently, through a mobile app. Customers in this program are able to receive a discount on their policy depending on how well they drive. An example graphic is displayed below (Figure 1) showing the customer using the Drive Safe & Save™ mobile application on their smartphone. The Bluetooth beacon sits inside the customer's vehicle and communicates with the phone – telling it to begin collecting driving data. The phone then transmits the data back to State Farm where the discount will ultimately be calculated and applied (State Farm 2017).

Figure 1: Graphic of Drive Safe & Save™ Mobile

The in-vehicle device or mobile app is able to capture different driving behaviors as a customer drives their vehicle. Some of the behaviors recorded are force of acceleration and deceleration, time of drive, turning, and many others. Every time a driver uses their car, a new trip is started and all behaviors for that trip are recorded. After every trip is recorded and the data is transmitted back to State Farm servers at periodic

intervals, the raw data is transformed into meaningful driver characteristics. These can include number of hard braking events (based on g forces), percent of time driving during a certain time of day, quick left turns, etc. The trip level data is summarized up to the day level such that every day is a single observation in the dataset.

In this paper we show how to analyze such rich driver behavior data and extract valuable information about individual driving behavior as compared to an average driver.

The rest of the paper is organized as follows. Section 2 describes the driving behavior data. Section 3 details the analysis methodology. Section 4 provides the results followed by conclusions in Section 5.

SECTION 2: DRIVING BEHAVIOR DATA

The driving behavior data of drivers enrolled in the Drive Safe & Save™ program consists of trip level data that is summarized to a day level. In this section we illustrate different driving behaviors of two customers. Table 1 shows the driving behavior data for a customer with driver id = 1. This customer enrolled in the program on 08/01/2013. The customer completed one trip of 20 miles on the date of enrollment and had two hard braking events and one rapid acceleration event. On 08/02/2013, the customer had only one trip. This trip was 10 miles long, and there were zero hard braking events and one rapid acceleration event. For this customer, data was available through 12/31/2014.

Driver ID	Date	Miles Driven	# Hard Brakes	# Rapid Accelerations	# Trips
1	08/01/2013	20	2	1	1
1	08/02/2013	10	0	1	1
. .	. .	. .	. .	. .	. .
1	12/31/2014	25	1	2	5

Table 1: Driving Behavior Data for Driver 1

Table 2 shows the driving behavior data for customer with driver id=2.

Driver ID	Date	Miles Driven	# Hard Brakes	# Rapid Accelerations	# Trips
2	11/01/2013	60	0	1	3
2	10/02/2013	80	1	2	9
. .	. .	. .	. .	. .	. .
2	12/31/2014	40	0	0	4

Table 2: Driving Behavior Data for Driver 2

Tables 1 and 2 indicate that both drivers had unsafe driving events. Table 3 below summarizes the driving behavior of the two drivers. Driver 1 drove 12,000 miles over a period of 17 months whereas driver 2 drove 18,000 miles in 14 months. Driver 1 had a higher count of both the number of hard brakes as well as the number of rapid accelerations.

Statistic	Driver 1	Driver 2
Period Observed	17 Months	14 Months
Number of Trips	2,200	3,500
Miles Driven	12,000	18,000
Number of Hard Brakes	225	110
Number of Rapid Accelerations	220	200

Table 3: Driving Behavior Summary

Table 3 provides simple summary statistics for each driver based on the entire time period the driver was enrolled in the Drive Safe & Save™ program. Although the numbers summarize the observation period, there are some important questions which such a summary fails to address. These include:

- Which driver can be considered an unsafe or safe driver compared to the average driver?
- Do drivers change their driving behaviors over time? Does a driver that begins the program as a safe driver gradually become an unsafe driver or vice versa?
- How do we effectively rank drivers based on their long-term and short-term driving behavior?

Answers to such questions are important for understanding driver behavior over time and for providing feedback to the customers regarding their driving behavior.

SECTION 3: ANALYSIS METHODOLOGY

For the methodology used in this study, the concept of recurrent events is very important to analyze driver behavior data. Recurrent events are similar events that keep on occurring over time. Examples of recurrent events in other contexts include repair events for automobile transmission systems, recurrences of bladder cancer tumors (Nelson 2003), etc. The recurrent event data includes a subject identifier and age of the subject at the time of the event. In the context of driving, drivers are the subjects and hard braking (or other events of interest) are recurrent events since drivers repeatedly engage in such behaviors. The recurrent event data is often censored since we observe the subject only through the time of analysis. The event history beyond the censoring age is unknown at the time of data analysis (Nelson 2003).

For recurrent data analysis, generally the age of the subject at the time of the event is recorded. Hence the gap between two successive events is expressed in terms of time. In the case of driving behavior data, expressing the gap between successive unsafe events in terms of time units can be misleading. Such analysis will not be able to address the fact that the vehicle is not driven continuously (i.e. every hour of the day, every day of the week). To address this shortcoming, we performed analysis using mileage instead of age of the vehicle at the unsafe event.

There are two popular ways for depicting the recurrent data, the event plot and the cumulative history function (CHF) plot. Figure 2 provides an example of an event plot for four drivers. The X axis shows the mileage and markers indicate the mileage at which an unsafe event took place.

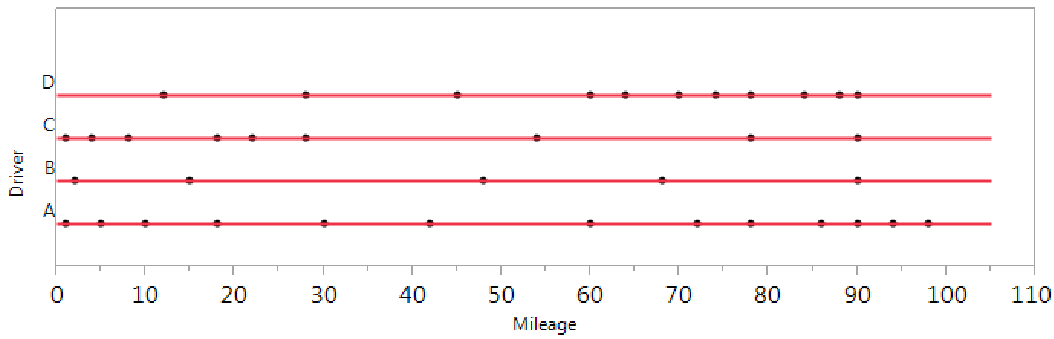


Figure 2: Example of an Event Plot

An alternate display of recurrent event data is using a cumulative history function plot, shown in Figure 3. This plot shows the cumulative number of events on the Y axis and mileage on the X axis for two drivers.

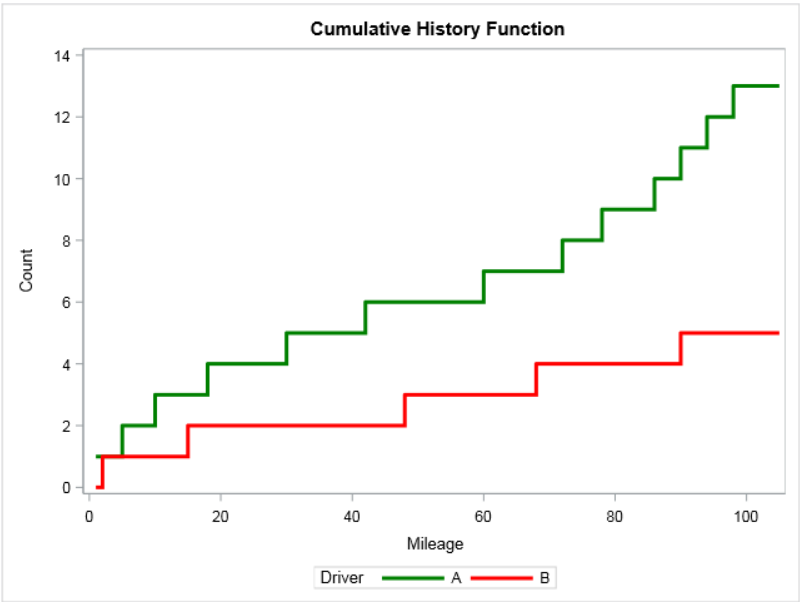


Figure 3: Example of a Cumulative History Function Plot

A mean cumulative function (MCF) is often used to describe a population of subjects who experience recurrent events. The value of the MCF at any mileage *t* is obtained by taking an average across all values of the cumulative history functions at mileage *t.*

If *n* drivers are observed for mileage of more than *t* miles, then the value of the MCF at mileage *t* is computed as:

$$MCF_t = \frac{\sum_{i=1}^{n} CF_{it}}{n}$$

where CF_{it} is the cumulative event count for subject *i* at mileage *t*. The function is therefore a mean curve. Figure 4 illustrates a MCF.

The MCF shown in Figure 4 resembles a staircase function similar to the cumulative history function shown in Figure 3. When the number of subjects used for constructing the MCF is large, the curve starts to lose its staircase shape and becomes smoother.

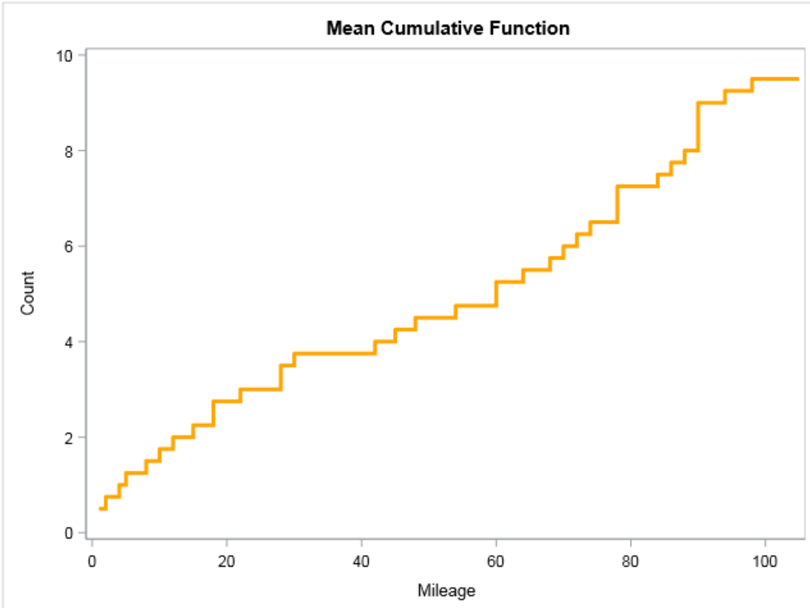
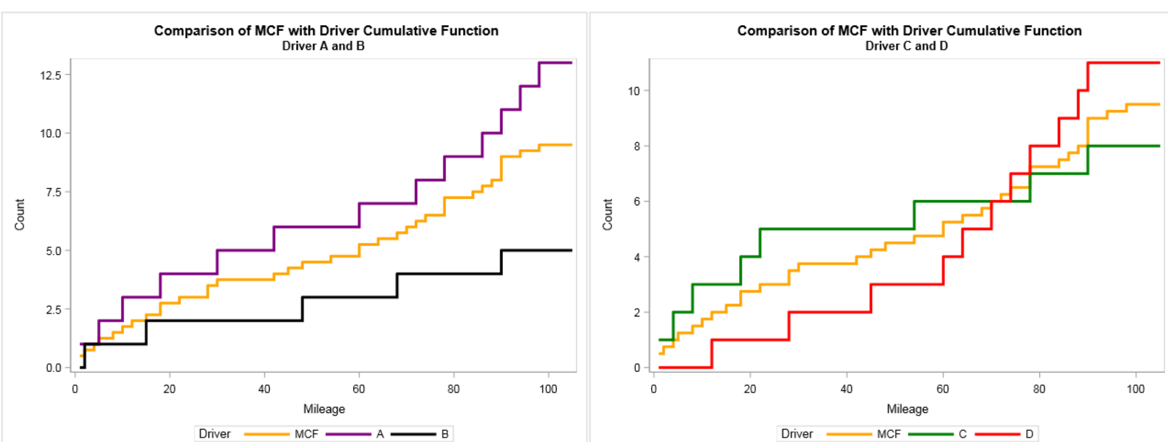


Figure 4: Example of a Mean Cumulative Function

The comparison of the mean cumulative function of unsafe events with an individual driver's cumulative history function provides useful insights. Figures 5 (a) and 5 (b) shows the MCF of unsafe events along with the cumulative history function of four drivers with different driving behavior. In Figure 5 (a), the cumulative history function of driver A is always above the MCF, indicating driver A is an unsafe driver compared to the average driver. On the other hand, driver B is a safer driver compared to an average driver. Figure 5 (b) illustrates driving behavior of two drivers which changes over time. Driver C starts out as an unsafe driver with the cumulative history function above the MCF, but over time as driver C accumulates mileage, the cumulative history function drops below the MCF indicating safer driving behavior compared to the average. Driver D starts out as a safe driver with his cumulative history function below the MCF, but over time as D accumulates mileage, his cumulative history function goes above the MCF indicating unsafe driving behavior compared to the average. Both plots in Figure 5 indicate that the MCF and cumulative history function together form an excellent tool to collectively describe the average driving behavior and contrast individual driving behavior.



(a) (b)

Figures 5 (a) & 5 (b): Comparison of MCF with Driver Cumulative History Function

ANALYSIS OF DRIVER BEHAVIOR DATA USING MCF

The individual driver's driving behavior data such as depicted in Tables 1 and 2 was used to obtain a MCF which describes behavior of an average driver with respect to unsafe events as a function of cumulative mileage. The results reported in this paper are based on driving behavior of 100 drivers observed between January 2013 and December 2014. Table 5 summarizes the key data statistics across all 100 drivers in the sample.

Variable	Min	Median	Mean	Max
Number of Days Driven	1	259	251	626
Number of Stops Taken per Day	0	14	18	242
Number of Miles Driven per Day	0	14	23	1034
Number of Trips Taken per Day	0	3	4	32

Table 4: Summary Statistics

For each driver, we used the data discussed in Tables 1 and 2 to calculate an individual driver's cumulative mileage and cumulative unsafe event count. The cumulative events were subsequently distributed in mileage intervals of 5 miles such as 0-5, 5-10 etc. assuming equal probability of events happening at any mileage in a given trip. A new variable named censored was created which had a value of 1 if the censoring mileage of a vehicle fell in the mileage interval, 0 otherwise. In other words, if the driver drove a cumulative total of 26 miles, the censored variable would equal 1 in the 25-30 interval, but 0 in the 0-5, 5-10, 10-15, 15-20, and 20-25 intervals. The above computations were performed for each driver individually. Subsequently, the unsafe event count and censor columns were summarized across drivers. The summarized dataset provided a total count of recurrences and censors across all mileage intervals. Such data is referred to as interval mileage data. Figure 6 summarizes the steps utilized to construct the interval mileage data.

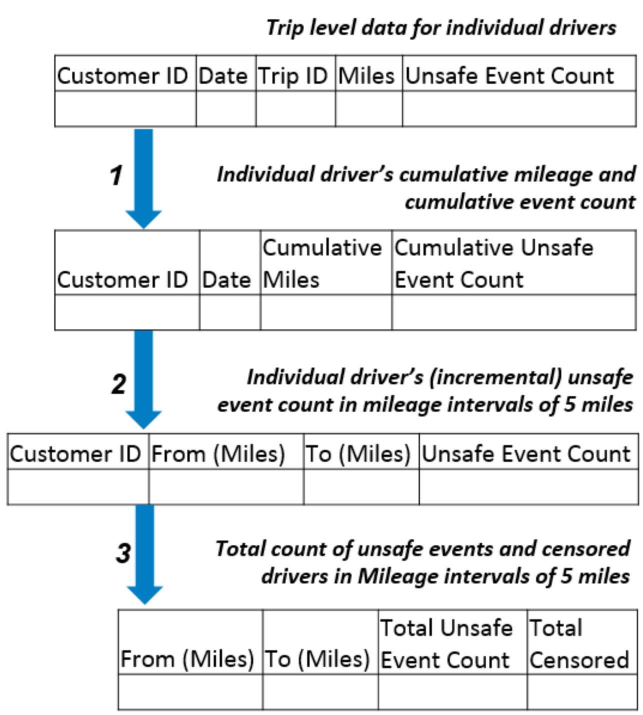


Figure 6: Steps Used to Construct Interval Mileage Data

The interval mileage dataset was analyzed using the RELIABILITY Procedure available in the SAS/QC® software to obtain the mean cumulative function (MCF). The RELIABILITY procedure provides a MCFPLOT statement which can be used to compute the MCF.

The MCF computation was followed by comparison of the individual driver's cumulative count function (CF) with the overall MCF. The difference $\nabla_{i,k}$ between the individual driver's cumulative event count function and the MCF curve at mileage interval *k* was computed. This difference was used to quantify departure of driver's driving behavior from the average driving behavior observed at mileage interval *k*. This difference is computed as:

$$\nabla_{i,k} = MCF_k - CF_{ik}$$

To rank drivers at any mileage interval, the percentile values of both the $(\nabla_{i,k} - \nabla_{i,k-1})$ and $\nabla_{i,k}$were computed at each mileage interval and were plotted as a function of mileage. The $\left(\nabla_{i,k} - \nabla_{i,k-1}\right)$ percentile is termed as the incremental percentile whereas the $\nabla_{i,k}$ percentile is termed as the cumulative percentile. The incremental percentile is useful to rank drivers based on their short term driving behavior whereas the cumulative percentile provides a robust ranking of drivers based on their entire driving history. Drivers in higher percentiles engage in unsafe driving behaviors more frequently compared with other drivers in the group. Figure 7 shows results for one of the drivers. The unsafe event is a quick left turn. The X axis shows the mileage. The Y axis on the left shows the cumulative count of the unsafe event where as the Y axis on the right shows the percentile value. The red line indicates the mean cumulative function of the unsafe event counts as a function of mileage. The blue line indicates the cumulative count for the individual driver. The green line indicates the incremental percentile value and the brown line indicates cumulative percentile value. The red and blue lines correspond to the left Y axis whereas the brown and green lines correspond to the right Y axis. As an example, at 1,500 miles, the incremental percentile was between the 50 to 60 percentile whereas the cumulative percentile was in the 90 to 100 percentile range. The incremental percentile indicates that the driver improved his driving between 1,250 to 1,500 miles. Although, this improvement did not significantly change the driver's ranking based on the cumulative percentile at 1,500 miles.

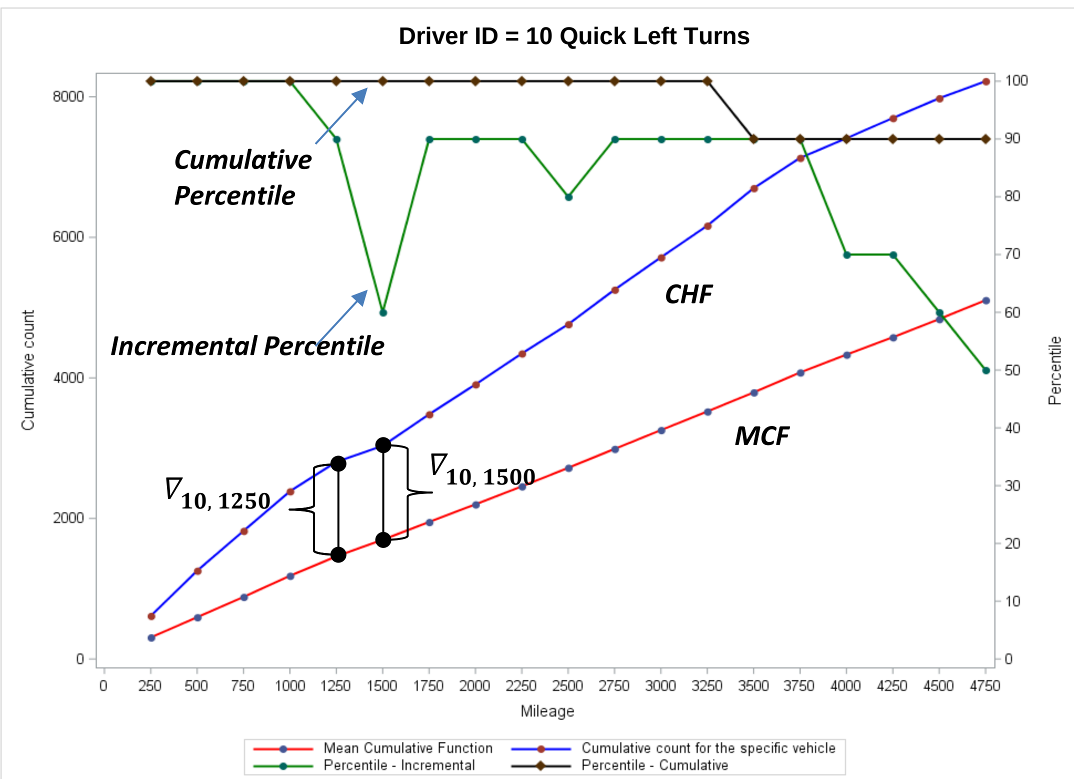


Figure 7: Driving Behavior Analysis: Key Statistics

SECTION 4: RESULTS

The methodology outlined in Section 3 was used to analyze driving behavior data. A SAS macro was created to analyze individual driver behavior as compared to the average. Comparison of the driver cumulative history function with respect to the mean cumulative function of the group allows identification of safe and unsafe drivers. A driver with a cumulative history function which is consistently above the mean cumulative function can be categorized as an unsafe driver compared to the average driver. Similarly, a driver with a cumulative history function curve consistently below the mean cumulative function, can be categorized as a safer driver compared to the average. The incremental and cumulative percentiles allow ranking drivers based on their short term and long term deviations from the MCF. Some typical plots of driving behavior plots are presented below. The unsafe event plotted here is again quick left turns. The following graphs illustrate how individual driver's behavior over time can be characterized using the analysis methodology outlined in this paper.

Safe Driver: The graph below (Figure 8) shows the cumulative history function of an individual driver below the MCF, indicating that the driver is a safe driver compared to the average driver. The incremental and cumulative percentile values are on the lower side as well.

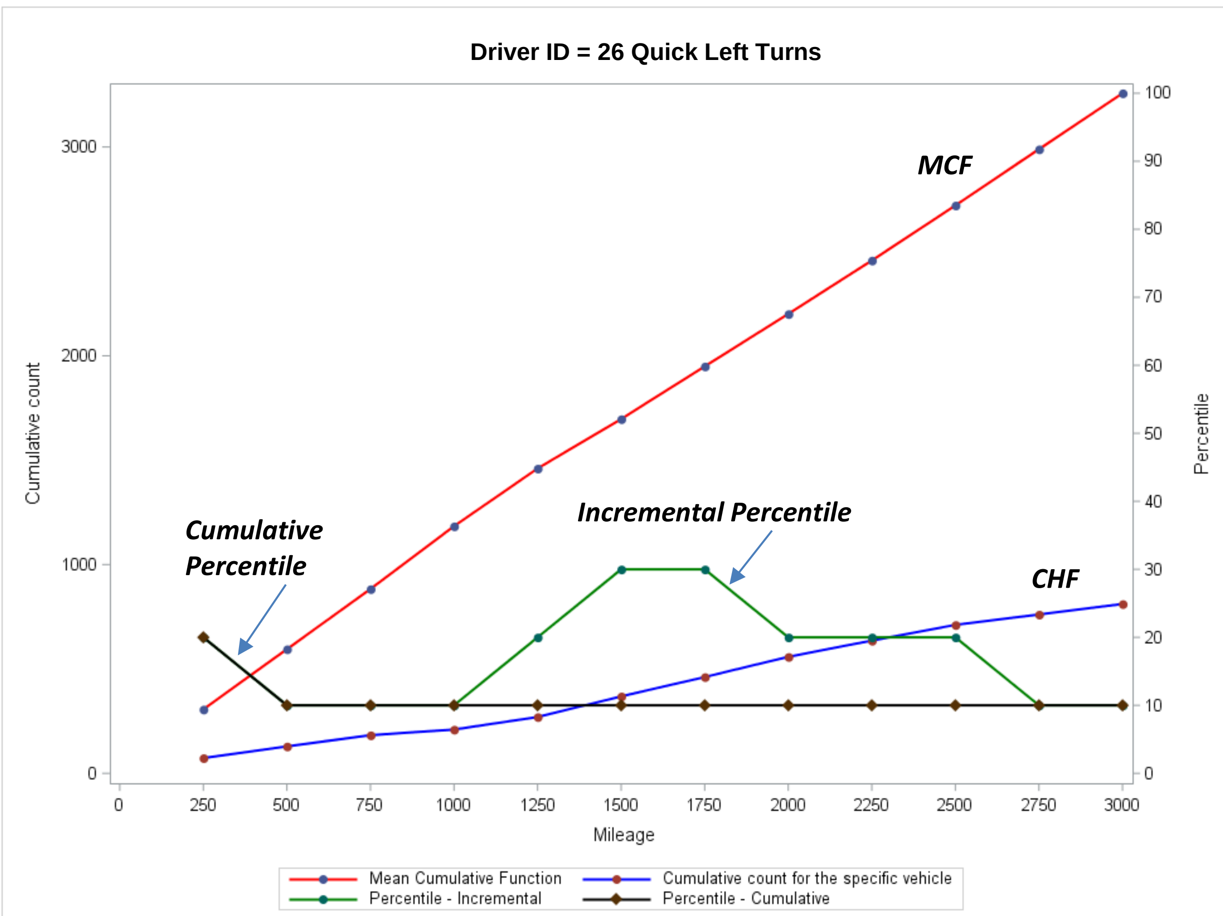


Figure 8: Example of a Safe Driver

Unsafe Driver: The next graph (Figure 9) shows the cumulative history function of an individual driver above the MCF, indicating that the driver is an unsafe driver compared to the average driver. The cumulative percentile values are on the higher side. The incremental percentile drops at certain mileages, but is not enough to change the cumulative percentile ranking of the driver.

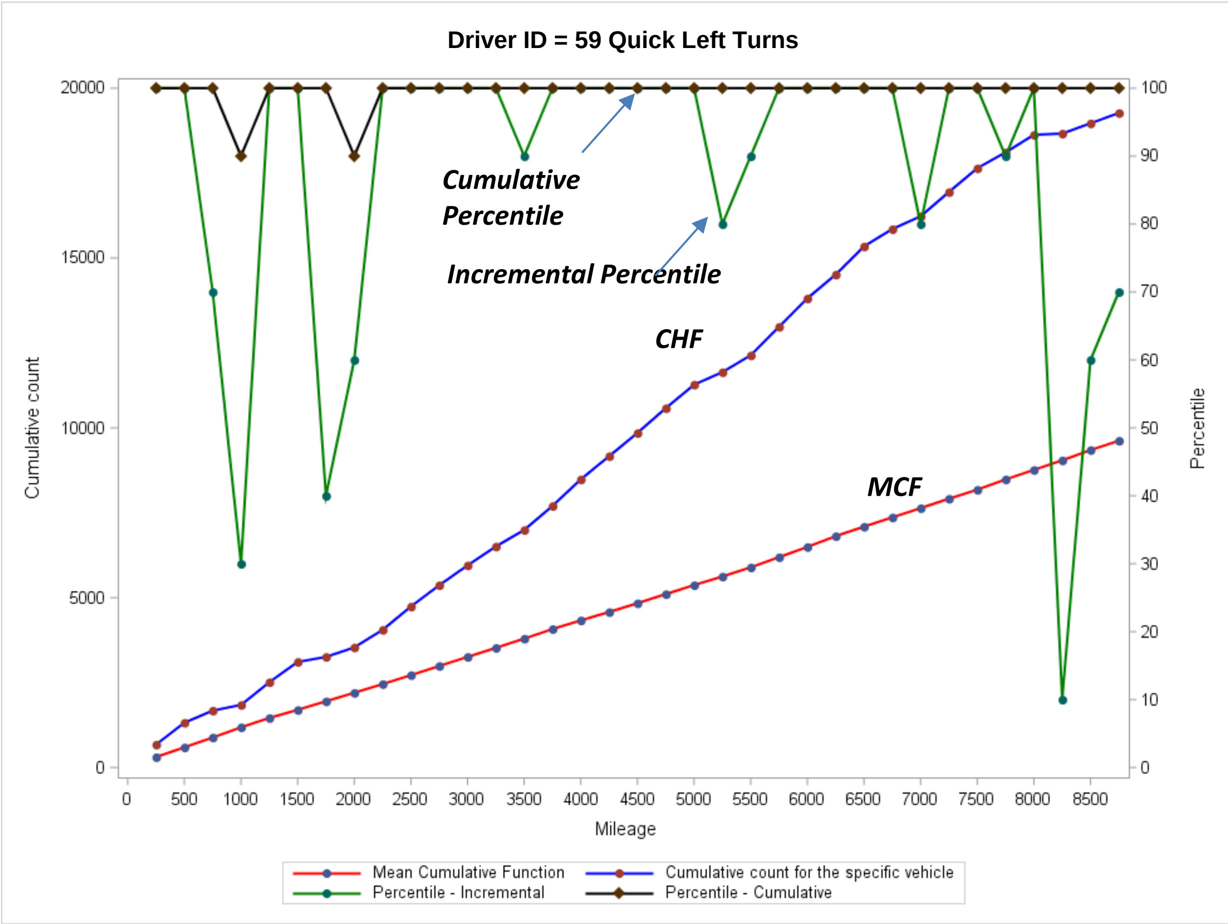


Figure 9: Example of an Unsafe Driver

Average Driver: Figure 10 shows the cumulative history function of an individual driver close to the MCF, indicating that the driver is close to an average driver in the group and stays close the entire time. As more miles are accumulated, the incremental and cumulative percentile values are right around average as well.

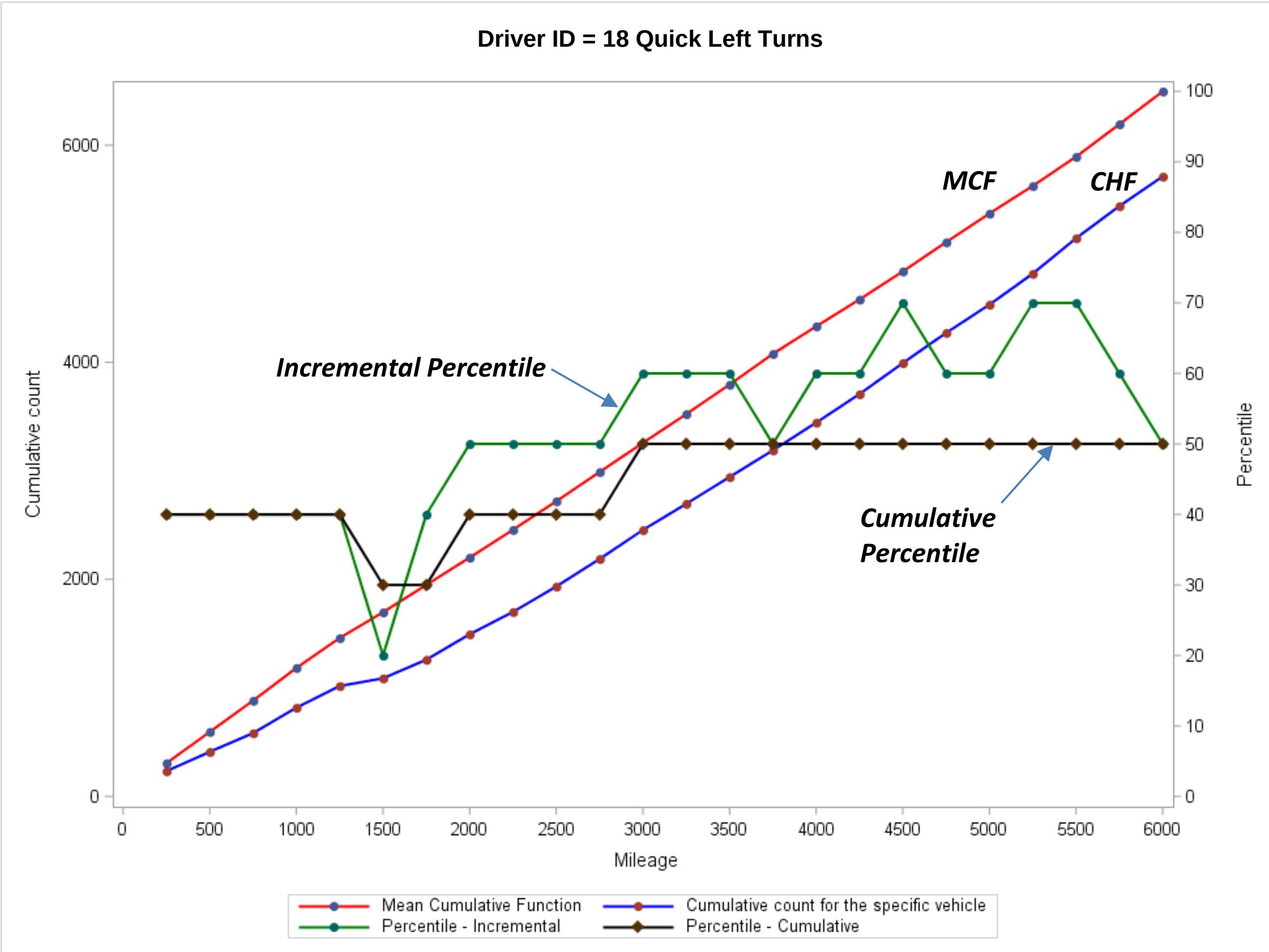


Figure 10: Example of an Average Driver

Driver with Improved Driving Behavior: The last graph (Figure 11) shows the cumulative history function of an individual driver above the MCF first, indicating unsafe driving as compared to the average. Over time, the cumulative history function falls below the MCF, indicating safe driving as compared to the average. This safe driving is evidenced in both the blue line falling well below the red line, but also in that the brown line has decreased as the driver has driven more miles.

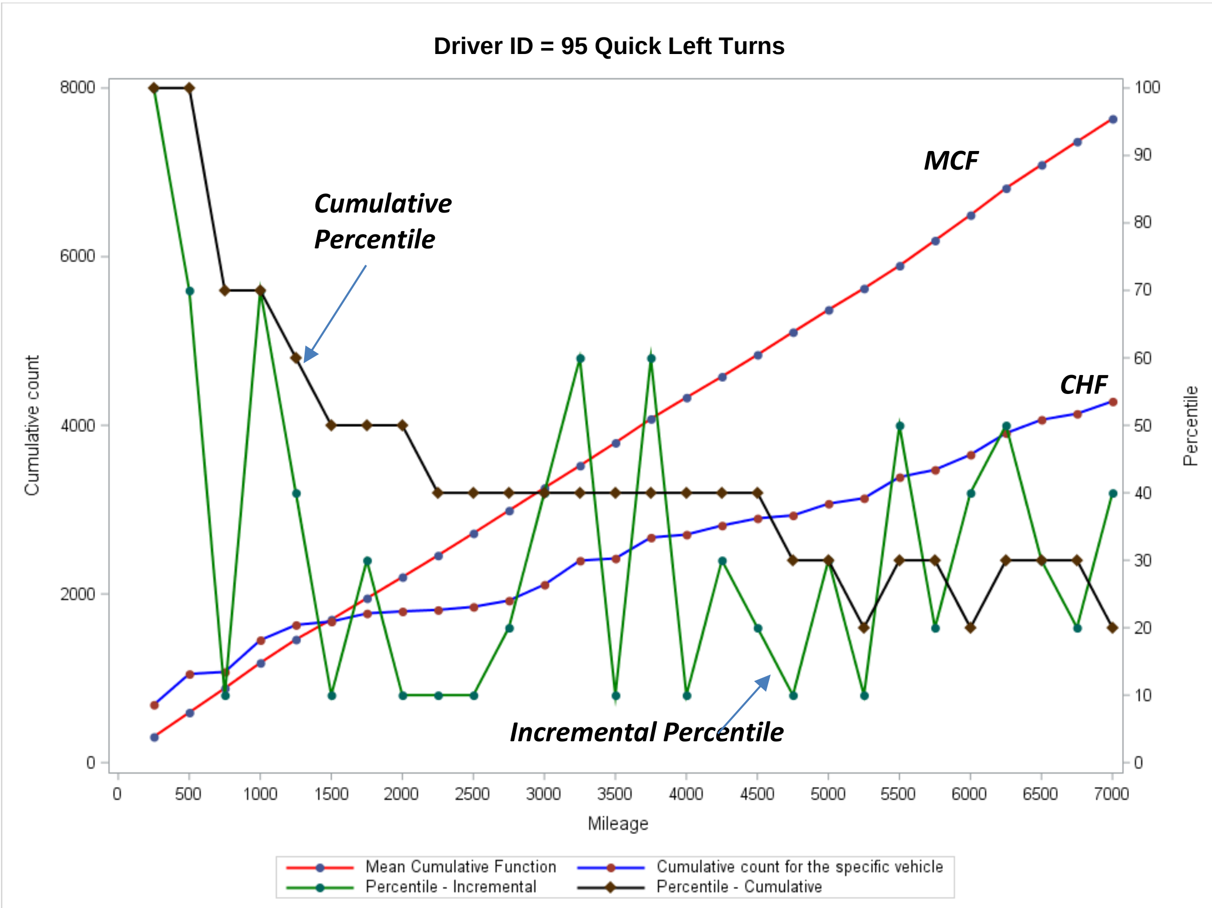


Figure 11: Example of a Driver with Improved Behavior

All the plots here are depicting quick left turns, but the same analysis could be done for all unsafe events (or any event of interest). For example, we can do the same thing for quick right turns, rapid accelerations or decelerations, unsafe speeds, unsafe times of the day, etc. where the individual driver's behavior is compared to the average driver in the population. Additionally, we could subset the data on certain characteristics to allow the driver to see how they compare to other subgroups. This could include a comparison of an individual female driver to the average female driver, a driver with a sports car to the average driver with a sports car, a driver with a new vehicle to the average driver with a new vehicle, and so on.

SECTION 5: CONCLUSION

A simple method for analyzing driving behavior is presented. The method uses mean cumulative functions to characterize driving behavior of an average driver and utilizes deviations of individual drivers from the average for ranking the drivers. The results obtained from such analysis can be communicated back to the

customer on a regular basis. Improvement in driving behavior as a result of feedback can be tracked using the methodology presented in this paper. In addition to driver feedback, the calculation of the deviations of individual drivers from the average driver in the population is a way to quantify a policyholder's risk. There is potential that this information could be used as additional regressors in a model that is trying to capture the riskiness of a customer. Overall, this analysis can help achieve many insurer's goals of better segmenting the population so the customer is charged the correct rate while also being able to provide feedback to customers in hopes of them being safer on the roads.

REFERENCES

Nelson, W. B. (2003). *Recurrent events data analysis for product repairs, disease recurrences, and other applications*. Society for Industrial and Applied Mathematics.

State Farm Mutual Automobile Insurance Company (2017). "Drive Safe & Save™ Mobile". Retrieved from https://www.statefarm.com/insurance/auto/discounts/drive-safe-save/mobile-app

RECOMMENDED READING

SAS Institute Inc. 2013. SAS/QC® 13.1 User's Guide. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Kelsey Osterloo
Property and Casualty Actuarial Department
State Farm Insurance Companies
One State Farm Plaza
Bloomington, Illinois 61710-001
kelsey.osterloo.u15h@statefarm.com

Deovrat Kakde
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Dev.Kakde@sas.com