Data Integration for Drivers Telematics with Selection Biases

Hashan Peiris [1] **Himchan Jeong** [1] **Jae-Kwang Kim** [2]

1. Data Structure

Driver telematics data refers to a variety of data collected from driving records of a vehicle either with a physical device attached or an app on a real-time basis as in Figure 1.

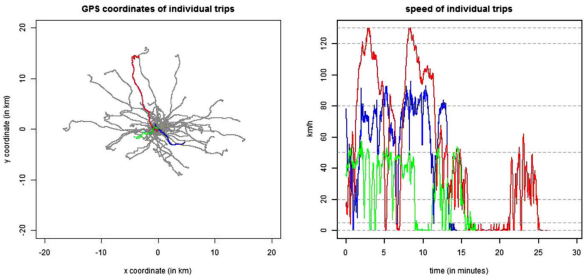

Figure 1. [Left panel] 200 individual trips of a car driver with [Right panel] resulting speed profiles of the three colored trips (Wüthrich, 2017)

While raw telematics data are collected in a real-time basis, usually automobile insurance premium is assessed per every period of time with a tabular dataset. In this regard, we focus our discussion on the integration of the so-called traditional dataset and a telematics dataset where telematics features are pre-processed in a tabular format tailored to be used in fitting GLMs, which are widely used in actuarial practice for ratemaking. We define $\mathcal{S}_0$ as a small dataset with M_0 observations that contains both telematics ($\mathbf{x}_{i2}$) and traditional features, and $\mathcal{S}_1$ as a large dataset with M_1 number of observations that contains only traditional features ($\mathbf{x}_{i1}$) as visualized in Figure 2. We also assume that the finite population $\mathcal{S}$ consists of $\mathcal{S}_0$ and $\mathcal{S}_1$ and the total number of observations in $\mathcal{S}$ is M.

2. Problem Description and Data Integration Approaches

As we work with two datasets both in a tabular format, here we want to estimate $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ in a Poisson regression model[1] $E(N_i \mid \mathbf{x}_i) = \exp(\mathbf{x}_i\boldsymbol{\beta}) = \exp(\mathbf{x}_{i1}\boldsymbol{\beta}_1 + \mathbf{x}_{i2}\boldsymbol{\beta}_2)$ where N_i is the response variable (claim frequency) and observed throughout the population. In this case, the census estimating equation for $\boldsymbol{\beta}$ can be written as $\sum_{i=1}^{M} U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) = 0$ where $U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) = \{n_i - \exp(\mathbf{x}_i\boldsymbol{\beta})\}\mathbf{x}_i$.

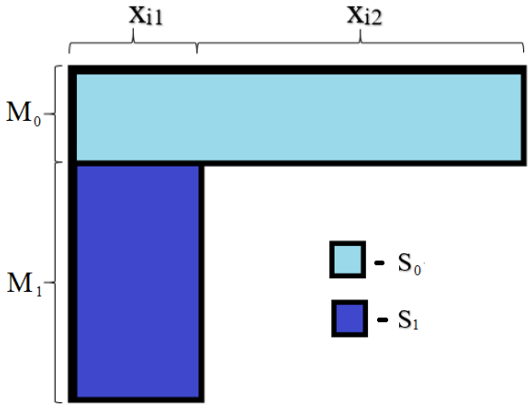

Figure 2. Pictorial visualization of $\mathcal{S}_0$, $\mathcal{S}_1$, $\mathbf{x}_{i1}$, and $\mathbf{x}_{i2}$

Due to missingness of $\mathbf{x}_{i2}$ in $\mathcal{S}_1$, we solve the following equation instead of the census estimating equation:

$$\sum_{i=1}^{M} \delta_i \omega_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) = 0, \tag{1}$$

where $\delta_i = \mathbb{I}(i \in \mathcal{S}_0)$. For comparison, we consider the following approaches to estimate $\boldsymbol{\beta}$ for some ω_i and subsequently predict the claim frequency N_i:

- **Complete case method**: Solve (1) for $\boldsymbol{\beta}$ assuming that $\omega_i = 1$ for all i.
- **Traditional method**: Solve (1) for $\boldsymbol{\beta}_1$ assuming that $\delta_i\omega_i = 1$ for all i and $\boldsymbol{\beta}_2 = 0$.
- **Full method**: Solve (1) for $\boldsymbol{\beta}$ assuming that $\delta_i\omega_i = 1$ for all i. So that it is expected to provide the best prediction performance but may not be available in practice.

[1]Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada [2]Department of Statistics, Iowa State University, 2438 Osborn Drive, Ames, IA 50011, USA. Correspondence to: Himchan Jeong <himchan_jeong@sfu.ca>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

[1]Note that uses of all the following methods are not restricted to Poisson distribution but applicable to any types of discrete distribution.

- **Boosting method**: Solve (1) for $\boldsymbol{\beta}_2$ assuming that $\omega_i = 1$ for all i and $\boldsymbol{\beta}_1$ equals to its estimate from the traditional method as in Ayuso et al. (2019).
- **Propensity score method**: Estimate ω_i so that $\omega_i - 1 \propto \exp(\phi_0 + \sum_{l=1}^{L} b_{li}\phi_l)$ and $\mathbb{E}[\delta_i\omega_i|\mathbf{x}_{1i}, n_i] = 1$ to handle the possible selection biases, where b_{li} are formed only using the traditional features ($\mathbf{x}_{1i}$) and observed number of claims (n_i). After that, solve (1) for $\boldsymbol{\beta}$ by letting $\omega_i = \hat{\omega}_i$ for all i. See Appendix A for details of this approach.

3. Data Analysis and Discussions

To quantify the impact of possible adverse selection in data integration, we analyze two datasets here; (i) a simulated dataset that mimics an actual insurance claims portfolio and (ii) a synthetic dataset that is originally introduced by So et al. (2021) and modified accordingly. (see Appendix B for detailed descriptions of the datasets) Both datasets include $100,000$ observations and are treated as finite populations of the automobile insurance policyholders with policy characteristics and claims information. After that, we split them into three disjoint samples; a small training set $\mathcal{S}_0$ with $\mathbf{x}_1$, $\mathbf{x}_2$, and N, a large training set $\mathcal{S}_1$ only with $\mathbf{x}_1$ and N, and a test set $\mathcal{T}$ under the following three scenarios on selection biases (**random**, **age**, and **adverse** selections) to test the impacts of possible selection biases on each of the data integration approaches. Firstly, 100,000 of data points are bootstrapped at random as $\mathcal{T}$ for out-of-sample validation, where $\{N_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}\}$ are all available. In this regard, $\mathcal{T}$ is homogenous to the (unknown) population with telematics features. After that, 100,000 of the data points are bootstrapped as $\mathcal{S}_0$ with a sampling probability p_i. Depending on the availability of telematics information, we applied the following sampling schemes of $\mathcal{S}_0$:

- **Random selection**: The data points assigned to $\mathcal{S}_0$ are chosen at random, equivalently with $p_i = 0.1$.
- **Age selection**: Each data point assigned to $\mathcal{S}_0$ is chosen with $p_i = \frac{1}{1+\exp(\texttt{Insured.Age})}$.
- **Adverse selection**: Each data point assigned to $\mathcal{S}_0$ is chosen with $p_i = \frac{1}{1+\exp(N_i)}$.

Lastly, 800,000 of data points are bootstrapped as $\mathcal{S}_1$ with a sampling probability $1 - p_i$. Table 1 showcases the out-of-sample validation results with the five approaches for data integration under different scenarios on selection biases, which are measured by mean (and standard error) of Poisson deviance in the test sets generated by 500 data splits. It is observed that in general, the traditional method (which excludes uses of telematics features) underperforms the other methods regardless of possible selection biases, which means simply discarding the telematics features from the analysis may harm predictive performance severely. On the other hand, the full method outperforms all the other methods in all situation while it might not be available in practice. Further, prediction performance of the complete case method (which excludes uses of $\mathcal{S}_1$, an external dataset with more data points but fewer features) is acceptable in the cases of no selection biases or selection bias on the observable covariates (driver's age in our case), however, its prediction performance is quite worse compared to the full and propensity methods in the case of adverse selection. Lastly, the propensity method could be a reasonable alternative of the full method as it shows comparable predictive performance to that of the full method.

Table 1. Out-of-sample validation performance

	RANDOM	AGE	ADVERSE
SIMULATED DATASET			
COMPLETE	49.65 (0.91)	49.65 (0.91)	65.86 (2.20)
TRADITIONAL	52.55 (0.93)	52.55 (0.93)	52.55 (0.93)
BOOSTING	49.77 (0.83)	49.77 (0.83)	51.06 (0.90)
FULL	49.60 (0.91)	49.60 (0.91)	49.60 (0.91)
PROPENSITY	49.61 (0.91)	49.62 (0.91)	49.67 (0.91)
SYNTHETIC DATASET			
COMPLETE	23.89 (0.28)	23.91 (0.28)	25.36 (0.37)
TRADITIONAL	26.74 (0.30)	26.74 (0.30)	26.74 (0.30)
BOOSTING	24.07 (0.27)	24.07 (0.28)	25.35 (0.37)
FULL	23.87 (0.28)	23.87 (0.28)	23.87 (0.28)
PROPENSITY	23.88 (0.28)	23.88 (0.28)	24.24 (0.71)

4. Future Research Directions

Possible future research directions are suggested in threefold. Firstly, as pre-processing of the telematics features in a tabular format might lose their richness inherited from the data generation scheme on a real-time basis, one can consider integration of telematics data that is not summarized in a tabular format with a traditional insurance dataset. It would be more challenging as the traditional dataset is still collected in a tabular format. Secondly, it can be worthy to explore uses of neural network or tree-based models for unbalanced data integration as they can handle the high-dimensionality on its own. Lastly, the aforementiond data structure can describe different problems such as health insurance cost prediction, where one can observe basic health information (such as age, gender, and BMI) for a large number of policyholders whereas only a few policyholders provide detailed health information collected by digital devices (such as dietary habits, work-out and sleep patterns). To this end, we expect that this research can be an invitation to encourage the ML community to consider this issue, data integration with possible selection biases with more discussions and advanced methodologies, which eventually would end up with a better incentive structure for risk controls and improvement of affordability and equity in insurance provision.

References

Ayuso, M., Guillen, M., and Nielsen, J. P. Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. *Transportation*, 46(3):735–752, 2019.

Deville, J. C. and Särndal, C. E. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382, 1992.

Imai, K. and Ratkovic, M. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76:243–263, 2014.

Jeong, H. Dimension reduction techniques for summarized telematics data. *Journal of Risk Management*, 33(4):1–25, 2022.

Rubin, D. B. Inference and missing data. *Biometrika*, 63:581–592, 1976.

So, B., Boucher, J.-P., and Valdez, E. A. Synthetic dataset generation of driver telematics. *Risks*, 9(4):58, 2021.

Wang, H. and Kim, J. K. Information projection approach to propensity score estimation for handling selection bias under missing at random. *arXiv e-prints*, pp. arXiv–2104, 2021.

Wüthrich, M. V. Covariate selection from telematics car driving data. *European Actuarial Journal*, 7:89–108, 2017.

A. Propsensity Score Estimation of ω_i via Information Projection

Here we observe $(\mathbf{x}_{i1}, n_i)$ in the sample $\mathcal{S}_1$, whereas we observe $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, n_i)$ in the sample $\mathcal{S}_0$. In this case, we wish to construct the propensity weight $\omega_i = \omega(\mathbf{x}_{i1}, n_i)$ in $\mathcal{S}_0$ such that

$$\sum_{i\in\mathcal{S}_0} \omega_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) = \sum_{i=1}^{M} \left[\delta_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) + (1-\delta_i)\bar{U}(\boldsymbol{\beta}; \mathbf{x}_{i1}, n_i)\right], \tag{2}$$

where $\delta_i = \mathbb{I}(i \in \mathcal{S}_0)$ and $\bar{U}(\boldsymbol{\beta}; \mathbf{x}_{i1}, n_i) = E\{U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) \mid \mathbf{x}_{i1}, n_i\}$. The propensity score (PS) estimating equation satisfying (2) is called self-efficient, as it leads to an efficient estimation of $\boldsymbol{\beta}$ as long as the conditional expectation in $E\{U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) \mid \mathbf{x}_{i1}, n_i\}$ is correct. Here, we assume that the sampling mechanism for $\mathcal{S}_0$ is missing at random (MAR) in the sense of Rubin (1976). That is, we assume

$$\delta \perp \mathbf{x}_2 \mid (n, \mathbf{x}_1).$$

To find ω_i satisfying (2), we first find the basis functions satisfying

$$E\{U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) \mid \mathbf{x}_{i1}, n_i\} \in \text{span}\{b_1(\mathbf{x}_{i1}, n_i), \ldots, b_L(\mathbf{x}_{i1}, n_i)\}. \tag{3}$$

Now, using the basis functions in (3), we impose

$$\sum_{i\in\mathcal{S}_0} \omega_i [1, b_{1i}, \cdots, b_{Li}] = \sum_{i=1}^{M} [1, b_{1i}, \cdots, b_{Li}] \tag{4}$$

as a constraint for propensity weights ω_i, where $b_{li} = b_l(\mathbf{x}_{i1}, n_i)$. Constraint (4) is often called the covariate-balancing property (Imai & Ratkovic, 2014) or calibration property (Deville & Särndal, 1992).

Now, as long as (4) is satisfied, we can express

$$\begin{aligned}
\sum_{i\in S_0} \omega_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) &= \sum_{i=1}^{M} \delta_i \omega_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) + \sum_{i=1}^{M} (1 - \delta_i \omega_i) \sum_{k=0}^{L} \alpha_k b_{ki} \\
&= \sum_{i=1}^{M} \left\{ \delta_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) + (1-\delta_i) \sum_{k=0}^{L} \alpha_k b_{ki} \right\} + \sum_{i=1}^{M} \delta_i (\omega_i - 1) \left\{ U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) - \sum_{k=0}^{L} \alpha_k b_{ki} \right\}
\end{aligned}$$

for any $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_L)$. Thus, for the choice of $\hat{\boldsymbol{\alpha}}$ satisfying

$$\sum_{i=1}^{M} \delta_i(\omega_i - 1) \left\{ U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) - \sum_{k=0}^{L} \hat{\alpha}_k b_{ki} \right\} = 0, \tag{5}$$

we can obtain

$$\sum_{i \in S_0} \omega_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) = \sum_{i=1}^{M} \left\{ \delta_i U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) + (1 - \delta_i) \sum_{k=0}^{L} \hat{\alpha}_k b_{ki} \right\}. \tag{6}$$

Furthermore, the condition in (5) under model (3) implies that $\sum_{k=0}^{L} \hat{\alpha}_k b_{ki}$ is an estimator of $E\{U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) \mid \mathbf{x}_{i1}, n_i\}$. Thus, we can see that (6) shows the self-efficiency in (2). That is, the calibration condition (4) on the basis functions in (3) is a sufficient condition for self-efficiency.

Now, to uniquely determine ω_i, we can use the information projection of Wang & Kim (2021) under the constraint (4) to get

$$\omega_i = 1 + \frac{M_1}{M_0} \exp\left\{ \phi_0 + \phi_1 b_{1i} + \cdots + \phi_L b_{Li} \right\}, \tag{7}$$

where $M_0 = \sum_{i=1}^{M} \delta_i$, $M_1 = M - M_0$ and $\boldsymbol{\phi} = (\phi_0, \cdots, \phi_L)$ is an unknown parameter. The parameters are estimated by solving the calibration equation in (4).

Once $\phi_0, \cdots, \phi_L$ is estimated by (4), we can use

$$\hat{\omega}_i = 1 + \frac{M_1}{M_0} \exp\left\{ \hat{\phi}_0 + \hat{\phi}_1 b_{1i} + \cdots + \hat{\phi}_L b_{Li} \right\}$$

as the final propensity weights for estimating $\boldsymbol{\beta}$ using (8):

$$\sum_{i \in \mathcal{S}} \delta_i \hat{\omega}_i(\boldsymbol{\phi}) U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) = 0, \tag{8}$$

where $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1$ be the combined sample. Because the propensity weights satisfy the calibration equation in (4), it satisfies the self-efficiency without estimating the regression coefficients $\hat{\boldsymbol{\alpha}}$ in the regression model

$$E\{U(\boldsymbol{\beta}; \mathbf{x}_i, n_i) \mid \mathbf{x}_{i1}, n_i\} = \sum_{k=1}^{L} \alpha_k b_l(\mathbf{x}_{i1}, n_i).$$

To estimate the standard error of the estimates, note that there are two models in this method. One is the PS model (with parameter $\boldsymbol{\phi}$) and the other is the regression outcome model (with parameter $\boldsymbol{\beta}$). We can construct two estimating functions for estimating two parameters as follows.

$$\begin{aligned} \hat{U}_1(\boldsymbol{\phi}) &= \sum_{i \in \mathcal{S}} \{\delta_i \omega_i(\boldsymbol{\phi}) - 1\} \mathbf{b}_i, \\ \hat{U}_2(\boldsymbol{\phi}, \beta) &= \sum_{i \in \mathcal{S}} \delta_i \hat{\omega}_i(\boldsymbol{\phi}) U(\boldsymbol{\beta}; \mathbf{x}_i, n_i), \end{aligned}$$

where $\mathbf{b}_i = (1, b_{1i}, \cdots, b_{Li})'$ and

$$\omega_i(\boldsymbol{\phi}) = 1 + \frac{M_1}{M_0} \exp\left\{ \phi_0 + \phi_1 b_{1i} + \cdots + \phi_L b_{Li} \right\}.$$

The final estimator $\hat{\boldsymbol{\beta}}$ is the solution to the joint estimating equations:

$$\hat{U}_1(\boldsymbol{\phi}) = 0 \quad \text{and} \quad \hat{U}_2(\boldsymbol{\phi}, \boldsymbol{\beta}) = 0.$$

We can treat $\boldsymbol{\theta}' = (\boldsymbol{\phi}', \boldsymbol{\beta}')$ and define

$$\hat{U}(\boldsymbol{\theta}) = \begin{pmatrix} \hat{U}_1(\boldsymbol{\phi}) \\ \hat{U}_2(\boldsymbol{\phi}, \boldsymbol{\beta}) \end{pmatrix}.$$

The variance estimation for $\hat{\boldsymbol{\theta}}$ can be implemented using the Sandwich formula. That is, $V(\hat{\boldsymbol{\theta}}) = \tau^{-1}V(\hat{U})\tau^{-1'}$ where $\tau = E\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\hat{U}(\boldsymbol{\theta})\right\}$.

One can use an empirical estimate of $V(\hat{\boldsymbol{\theta}})$ as follows:

$$\tilde{\tau} = \frac{\partial}{\partial\theta'}\hat{U}(\theta)\bigg|_{\theta=\hat{\theta}} \qquad \text{and } \tilde{V}(\hat{U}) = \sum_{i\in\mathcal{S}}(\tilde{U}_i - \overline{\tilde{U}}_i)(\tilde{U}_i - \overline{\tilde{U}}_i)'$$

as a proxy of τ and $V(\hat{U})$, respectively where $\hat{\theta}' = (\hat{\phi}', \hat{\boldsymbol{\beta}}')$ is the solution of the joint estimating equation and

$$\tilde{U}_i = \begin{pmatrix} \left\{\delta_i\omega_i(\hat{\phi}) - 1\right\}\mathbf{b}_i \\ \\ \delta_i\hat{\omega}_i(\hat{\phi})U(\hat{\boldsymbol{\beta}}; \mathbf{x}_i, y_i) \end{pmatrix}, \quad \overline{\tilde{U}}_i = \frac{1}{M}\sum_{i\in\mathcal{S}}\tilde{U}_i.$$

B. Description of the Simulated and Synthetic Datasets

For the simulation study, we generate a finite population of size $100,000$ with the following specification:

$$\begin{aligned}
&N_i \sim \mathcal{P}(\lambda_i), \quad \log\lambda_i = \mathbf{x}_{i1}\boldsymbol{\beta}_1 + \mathbf{x}_{i2}\boldsymbol{\beta}_2,\ \boldsymbol{\beta}_1 = (\beta_0, \beta_{A1}, \beta_{A2}, \beta_G),\\
&\boldsymbol{\beta}_2 = \beta_T,\ \mathbf{x}_{1i} = (1, x_{Ai}, x_{Ai}^2, x_{Gi}),\ \mathbf{x}_{2i} = x_{Ti},\\
&x_{Ai} \sim \mathcal{U}(0.18, 0.81),\ x_{Gi} \sim \mathcal{B}er(0.6),\ x_{Ti} \sim \mathcal{N}(0,1),\\
&\beta_0 = -1.3,\ \beta_{A1} = -4,\ \beta_{A2} = 3.4,\ \beta_G = 0.1,\ \beta_T = 0.5,
\end{aligned}$$

where $\mathcal{P}$, $\mathcal{U}$, $\mathcal{B}er$ and $\mathcal{N}$ refer to Poisson, uniform, Bernoulli, and normal distributions, respectively. Here, x_{Ai} refers to a traditional continuous variable with quadratic effect (e.g., driver's age), x_{Gi} refers to a traditional binary variable (e.g., geographic location - urban/rural), and x_{Ti} refers to a telematics variable of significant impact on the risk profile.

The synthetic dataset is an emulated dataset from an actual insurance claims data with telematics features by adding perturbation. While it still preserves characteristics of the original dataset, it is not identical to the original dataset so that it is publicly available at https://www2.math.uconn.edu/~valdez/telematics_syn-032021.csv without proprietery issues or privacy concerns. It contains traditional characteristics (including age of the drivers), telematic characteristics, and the response variable, which is the claim counts. Note that all of the aforementioned data integration approaches is based on estimating equations (and equivalently GLMs) so that it lacks the ability to handle high-dimensionality on its own, unlike neural network models or tree-based models. In this regard, we further processesed the original dataset of So et al. (2021) to handle such high-dimensionality issues as described in Jeong (2022). For the resulting variables and their descriptions, see Table 2.

Table 2. Variable descriptions of the pre-processed synthetic dataset

VARIABLE	DESCRIPTION
TRADITIONAL	
DURATION	DURATION OF THE INSURANCE COVERAGE OF A GIVEN POLICY, IN DAYS
INSURED.AGE	AGE OF INSURED DRIVER, IN YEARS
INSURED.SEX	SEX OF INSURED DRIVER (MALE/FEMALE)
CAR.AGE	AGE OF VEHICLE, IN YEARS
MARITAL	MARITAL STATUS (SINGLE/MARRIED)
CAR.USE	USE OF VEHICLE: PRIVATE, COMMUTE, FARMER, COMMERCIAL
CREDIT.SCORE	CREDIT SCORE OF INSURED DRIVER
REGION	TYPE OF REGION WHERE DRIVER LIVES: RURAL, URBAN
ANNUAL.MILES.DRIVE	ANNUAL MILES EXPECTED TO BE DRIVEN DECLARED BY DRIVER
YEARS.NOCLAIMS	NUMBER OF YEARS WITHOUT ANY CLAIMS
TERRITORYEMB	EMBEDDED VALUE FROM THE TERRITORIAL LOCATION OF VEHICLE
TELEMATICS	
ANNUAL.PCT.DRIVEN	ANNUALIZED PERCENTAGE OF TIME ON THE ROAD
TOTAL.MILES.DRIVEN	TOTAL DISTANCE DRIVEN IN MILES
PCT.DRIVE.XXX	PERCENT OF DRIVING DAY XXX OF THE WEEK: MON/TUE/.../SUN
PCT.DRIVE.RUSH.AM	PERCENT OF DRIVING DURING AM RUSH HOURS
PCT.DRIVE.RUSH.PM	PERCENT OF DRIVING DURING PM RUSH HOURS
AVGDAYS.WEEK	MEAN NUMBER OF DAYS USED PER WEEK
ACCEL.06MILES	NUMBER OF SUDDEN ACCELERATION 6MPH/S PER 1000MILES
BRAKE.06MILES	NUMBER OF SUDDEN BRAKES 6MPH/S PER 1000MILES
ACBR.OTHERS	TOTAL NUMBER OF SUDDEN ACCELERATION AND BRAKES 8/9/.../14 MPH/S PER 1000MILES
LEFT.TURNS	NUMBER OF LEFT TURN PER 1000MILES WITH INTENSITY GREATER THAN EQUAL TO 8
RIGHT.TURNS	NUMBER OF RIGHT TURN PER 1000MILES GREATER THAN EQUAL TO 8
RESPONSE	
NB_CLAIM	NUMBER OF OBSERVED CLAIMS