Homework 4

Group 1

Contents

Prepared for:

Dr. Nathan Bastian
City University of New York, School of Professional Studies - Data 621

Prepared by:

Group 1

Senthil Dhanapal
Yadu Chittampalli
Christophe Hunt

1 Introduction

Consumers who own a car are often required to purchase car insurance to protect themselves from serious financial repercussions of being involved in a car accident. Insurance Providers must determine the risk of offering insurance coverage to a new customer through accurate statistical models that evaluate the consumers propensity for accidents. Since Insurance Providers are motivated by collecting the maximum amount of revenue from consumers while returning the lowest amount in accident claims, statistical modeling provides Insurance Providers with insight into the consumers behavior and the most appropriate pricing schemes[1].

2 Statement of the Problem

The purpose of this report is to develop statistical models to make inference into the likelihood of a customer being involved in a car accident and the cost associated of a customer being involved in a car accident.

3 Data Exploration

3.1 Variables Explained

The variables provided in the `Insurance Training Data Set` are explained below:

Variable Code	Definition
INDEX	Identification Variable (do not use)
TARGET_FLAG	Was Car in a crash? 1=YES 0=NO
TARGET_AMT	If car was in a crash, what was the cost
AGE	Age of Driver
BLUEBOOK	Value of Vehicle
CAR_AGE	Vehicle Age
CAR_TYPE	Type of Car
CAR_USE	Vehicle Use
CLM_FREQ	# Claims (Past 5 Years)
EDUCATION	Max Education Level
HOMEKIDS	# Children at Home
HOME_VAL	Home Value
INCOME	Income
KIDSDRIV	# Driving Children
MSTATUS	Marital Status
MVR_PTS	Motor Vehicle Record Points
OLDCLAIM	Total Claims (Past 5 Years)
PARENT1	Single Parent
RED_CAR	A Red Car
REVOKED	License Revoked (Past 7 Years)
SEX	Gender
TIF	Time in Force
TRAVTIME	Distance to Work
URBANICITY	Home/Work Area
YOJ	Years on Job

[1]"Insider Information: How Insurance Companies Measure Risk - Insurance Companies.com." Insurance Companiescom. N.p., n.d. Web. 06 Nov. 2016.

3.1.1 Nominal Variables

We first look at our nominal variables and their applicable proportions. Interestingly, we see that in this data set only a quarter of the customer records indicate an accident occurred. Also, the majority of consumers in this data set have no kids at home, are married, more than a high school education but less than a PhD, use their car for private purposes, typically own a SUV or minivan, and also live in an urban environment. This provides an interesting insight to the type of customer this data set represents and should be considered when further interpreting our statistical model. Additionally, we should be mindful of any selection biases in this data set as consumers with extremely risky histories are likely to have not been extended insurance coverage.

Table 2: Table of nominal variables

Variable	Levels	n	%	$\sum$%
TARGET_FLAG	0	6008	73.6	73.6
	1	2153	26.4	100.0
	all	8161	100.0	
KIDSDRIV	0	7180	88.0	88.0
	1	636	7.8	95.8
	2	279	3.4	99.2
	3	62	0.8	100.0
	4	4	0.0	100.0
	all	8161	100.0	
HOMEKIDS	0	5289	64.8	64.8
	1	902	11.1	75.9
	2	1118	13.7	89.6
	3	674	8.3	97.8
	4	164	2.0	99.8
	5	14	0.2	100.0
	all	8161	100.0	
PARENT1	No	7084	86.8	86.8
	Yes	1077	13.2	100.0
	all	8161	100.0	
MSTATUS	No	3267	40.0	40.0
	Yes	4894	60.0	100.0
	all	8161	100.0	
SEX	F	4375	53.6	53.6
	M	3786	46.4	100.0
	all	8161	100.0	
EDUCATION	Less Than High School	1203	14.7	14.7
	High School	2330	28.6	43.3
	Bachelors	2242	27.5	70.8
	Masters	1658	20.3	91.1
	PhD	728	8.9	100.0
	all	8161	100.0	
JOB		526	6.4	6.4
	Blue Collar	1825	22.4	28.8
	Clerical	1271	15.6	44.4
	Doctor	246	3.0	47.4
	Home Maker	641	7.8	55.2
	Lawyer	835	10.2	65.5
	Manager	988	12.1	77.6
	Professional	1117	13.7	91.3
	Student	712	8.7	100.0
	all	8161	100.0	
CAR_USE	Commercial	3029	37.1	37.1
	Private	5132	62.9	100.0
	all	8161	100.0	
CAR_TYPE	Minivan	2145	26.3	26.3
	Panel Truck	676	8.3	34.6
	Pickup	1389	17.0	51.6
	Sports Car	907	11.1	62.7
	SUV	2294	28.1	90.8

Table 2: Table of nominal variables

Variable	Levels	n	%	$\sum$%
	Van	750	9.2	100.0
	all	8161	100.0	
RED_CAR	no	5783	70.9	70.9
	yes	2378	29.1	100.0
	all	8161	100.0	
CLM_FREQ	0	5009	61.4	61.4
	1	997	12.2	73.6
	2	1171	14.3	88.0
	3	776	9.5	97.5
	4	190	2.3	99.8
	5	18	0.2	100.0
	all	8161	100.0	
REVOKED	No	7161	87.8	87.8
	Yes	1000	12.2	100.0
	all	8161	100.0	
URBANICITY	Highly Rural/ Rural	1669	20.4	20.4
	Highly Urban/ Urban	6492	79.5	100.0
	all	8161	100.0	

3.1.2 Continuous and Discrete Variables

We can see that in our continuous and discrete variables there is some additional variability. The median claim amount (`TARGET_AMT`) is 0 which would coincide with only a quarter for records indicating an accident. However, the spread is large since the average payout is only $1,504.30 but the maximum payout was $107,586.10. Surprisingly, the median `AGE` is 45 and the average `AGE` is 44.8 years, while we expected a lower average it could be due to simple selection bias in the data set source or the aging US population bringing this average higher [2]. We also noticed that an `INCOME` of $0.00 seems unwise because it is unclear how the individual would be able to cover their premium costs without parental support. Finally, we should note that the data set has as `CAR_AGE` of -3, which is impossible and will need to be removed.

There are many missing values for this portion of our data set, we have over 400 values missing for years on the job, income, home value, and car age. Due to these missing values we will need to impute to complete our statistical model.

Variable	n	Min	q_1	$\widetilde{x}$	$\bar{x}$	q_3	Max	s	IQR	#NA
TARGET_AMT	8161	0	0	0	1504.3	1036	107586.1	4704.0	1036	0
TIF	8161	1	1	4	5.4	7	25.0	4.1	6	0
AGE	8155	16	39	45	44.8	51	81.0	8.6	12	6
YOJ	7707	0	9	11	10.5	13	23.0	4.1	4	454
INCOME	7716	0	28097	54028	61898.1	85986	367030.0	47572.7	57889	445
HOME_VAL	7697	0	0	161160	154867.3	238724	885282.0	129123.8	238724	464
TRAVTIME	8161	5	22	33	33.5	44	142.0	15.9	22	0
BLUEBOOK	8161	1500	9280	14440	15709.9	20850	69740.0	8419.7	11570	0
OLDCLAIM	8161	0	0	0	4037.1	4636	57037.0	8777.1	4636	0
MVR_PTS	8161	0	0	1	1.7	3	13.0	2.1	3	0
CAR_AGE	7651	-3	1	8	8.3	12	28.0	5.7	11	510

Table 3:

[2] Ortman, Jennifer M., Victoria A. Velkoff, and Howard Hogan. "An aging nation: the older population in the United States." Washington, DC: US Census Bureau (2014): 25-1140.

3.2 Imputting Missing Values

In order to address the missing values in our variables we used a non-parametric imputation method (Random Forest) using the `missForest` package. The function is particularly useful in that it can handle any type of input data and it will make as few assumptions about the structure of the data as possible.[3]

Table 2 : Imputed Descriptive Statistics
25 Variables 8161 Observations

TARGET_FLAG

n	missing	distinct	Info	Sum	Mean	Gmd
8161	0	2	0.583	2153	0.3	0.4

TARGET_AMT

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	1949	0.601	1504	2574	0	0	0	0	1036	4904	6452

```
lowest :     0.00000    30.27728    58.53106    95.56732   108.74150
highest:  73783.46592  77907.43028  78874.19056  85523.65335 107586.13616
```

KIDSDRIV

n	missing	distinct	Info	Mean	Gmd
8161	0	5	0.318	0.2	0.3

```
lowest : 0 1 2 3 4, highest: 0 1 2 3 4

0 (7180, 0.880), 1 (636, 0.078), 2 (279, 0.034), 3 (62, 0.008), 4 (4, 0.000)
```

AGE

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	66	0.999	45	10	30	33	39	45	51	56	59

```
lowest : 16 17 18 19 20, highest: 72 73 76 80 81
```

HOMEKIDS

n	missing	distinct	Info	Mean	Gmd
8161	0	6	0.723	0.7	1

```
lowest : 0 1 2 3 4, highest: 1 2 3 4 5

0 (5289, 0.648), 1 (902, 0.111), 2 (1118, 0.137), 3 (674, 0.083), 4 (164, 0.020), 5 (14,
0.002)
```

YOJ

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	446	0.991	10	4	0	5	9	11	13	14	15

```
lowest :  0.00  0.15  0.20  0.26  0.27, highest: 16.00 17.00 18.00 19.00 23.00
```

INCOME

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	7057	1	61569	50845	0e+00	5e+03	3e+04	5e+04	9e+04	1e+05	2e+05

```
lowest :      0.00      5.00      7.00     18.00     26.33
highest: 306277.00 309628.00 320127.00 332339.00 367030.00
```

PARENT1

n	missing	distinct
8161	0	2

```
No (7084, 0.868), Yes (1077, 0.132)
```

HOME_VAL

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	5570	0.978	2e+05	1e+05	0e+00	0e+00	0e+00	2e+05	2e+05	3e+05	4e+05

```
lowest :      0.000   4176.960   8196.080   8263.335   9438.009
highest: 657804.000 682634.000 738153.000 750455.000 885282.000
```

[3] Stekhoven, Daniel J., and Peter Bühlmann. "MissForest-non-parametric missing value imputation for mixed-type data." Bioinformatics 28.1 (2012): 112-118.

MSTATUS

n	missing	distinct
8161	0	2

```
No (3267, 0.4), Yes (4894, 0.6)
```

SEX

n	missing	distinct
8161	0	2

```
F (4375, 0.536), M (3786, 0.464)
```

EDUCATION

n	missing	distinct
8161	0	5

```
lowest : Bachelors             High School           Less Than High School Masters               PhD
highest: Bachelors             High School           Less Than High School Masters               PhD

Bachelors (2242, 0.275), High School (2330, 0.286), Less Than High School (1203, 0.147),
Masters (1658, 0.203), PhD (728, 0.089)
```

JOB

n	missing	distinct
8161	0	8

```
lowest : Blue Collar  Clerical     Doctor       Home Maker   Lawyer
highest: Home Maker   Lawyer       Manager      Professional Student

Value       Blue Collar     Clerical       Doctor   Home Maker       Lawyer      Manager
Frequency          1830         1273          254          643          865         1412
Proportion        0.224        0.156        0.031        0.079        0.106        0.173

Value      Professional      Student
Frequency          1172          712
Proportion        0.144        0.087
```

TRAVTIME

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	97	1	33	18	7	13	22	33	44	54	60

```
lowest :   5   6   7   8   9, highest: 103 113 124 134 142
```

CAR_USE

n	missing	distinct
8161	0	2

```
Commercial (3029, 0.371), Private (5132, 0.629)
```

BLUEBOOK

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	2789	1	15710	9354	4900	6000	9280	14440	20850	27460	31110

```
lowest :  1500  1520  1530  1540  1590, highest: 57970 61050 62240 65970 69740
```

TIF

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	23	0.961	5	5	1	1	1	4	7	11	13

```
lowest :  1  2  3  4  5, highest: 19 20 21 22 25
```

CAR_TYPE

n	missing	distinct
8161	0	6

```
lowest : Minivan     Panel Truck Pickup      Sports Car  SUV
highest: Panel Truck Pickup      Sports Car  SUV         Van

Minivan (2145, 0.263), Panel Truck (676, 0.083), Pickup (1389, 0.170), Sports Car (907,
0.111), SUV (2294, 0.281), Van (750, 0.092)
```

RED_CAR

n	missing	distinct
8161	0	2

```
no (5783, 0.709), yes (2378, 0.291)
```

OLDCLAIM

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	2857	0.769	4037	6563	0	0	0	0	4636	9583	27090

```
lowest :     0   502   506   518   519, highest: 52507 53477 53568 53986 57037
```

CLM_FREQ

n	missing	distinct	Info	Mean	Gmd
8161	0	6	0.763	0.8	1

```
lowest : 0 1 2 3 4, highest: 1 2 3 4 5

0 (5009, 0.614), 1 (997, 0.122), 2 (1171, 0.143), 3 (776, 0.095), 4 (190, 0.023), 5 (18,
0.002)
```

REVOKED

n	missing	distinct
8161	0	2

```
No (7161, 0.877), Yes (1000, 0.123)
```

MVR_PTS

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	13	0.9	2	2	0	0	0	1	3	5	6

```
lowest :  0  1  2  3  4, highest:  8  9 10 11 13

Value          0     1     2     3     4     5     6     7     8     9    10    11    13
Frequency   3712  1157   948   758   599   399   266   167    84    45    13    11     2
Proportion 0.455 0.142 0.116 0.093 0.073 0.049 0.033 0.020 0.010 0.006 0.002 0.001 0.000
```

CAR_AGE

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
8161	0	507	0.985	8	6	1	1	4	8	12	16	18

```
lowest :  0.000  1.000  2.000  2.035  2.890, highest: 24.000 25.000 26.000 27.000 28.000
```

URBANICITY

n	missing	distinct
8161	0	2

```
Highly Rural/ Rural (1669, 0.205), Highly Urban/ Urban (6492, 0.795)
```

3.3 Exploration of Variables

4 Data Transformation

4.1 Outliers Treatment

We chose winsorizing as the method to address outliers. Instead of trimming values, winsorizing uses the interquantile range to replace values that are above or below the interquantile range multiplied by a factor. Those values above or below the range multiplied by the factor are then replaced with max and min value of the interquantile range. Using the factor 2.2 for winsorizing outliers is a method developed my Hoaglin and Iglewicz and published Journal of American Statistical Association in 1987[4].

The below table is the summary results of the winsorizing of the data.

[4] Hoaglin, D. C., and Iglewicz, B. (1987), Fine tuning some resistant rules for outlier labeling, Journal of American Statistical Association, 82, 1147-1149.

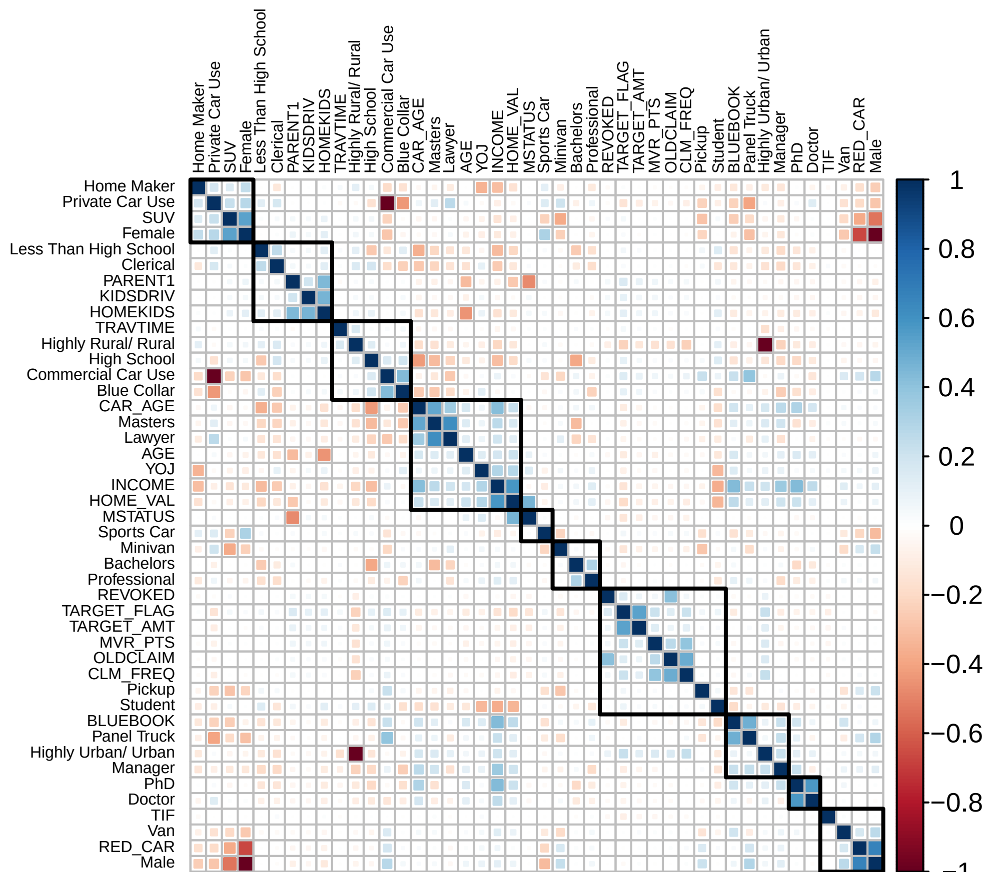


Figure 1: Correlation Plot of Training Data Set

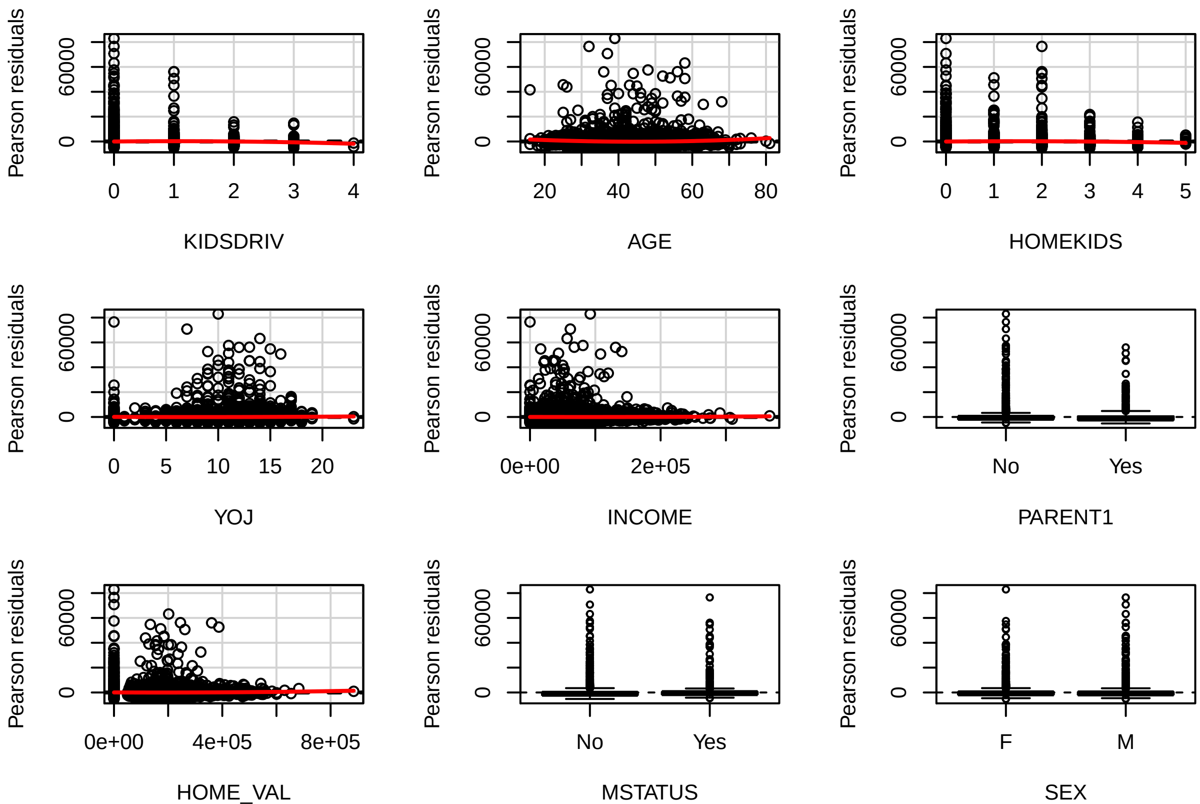
Table 4:

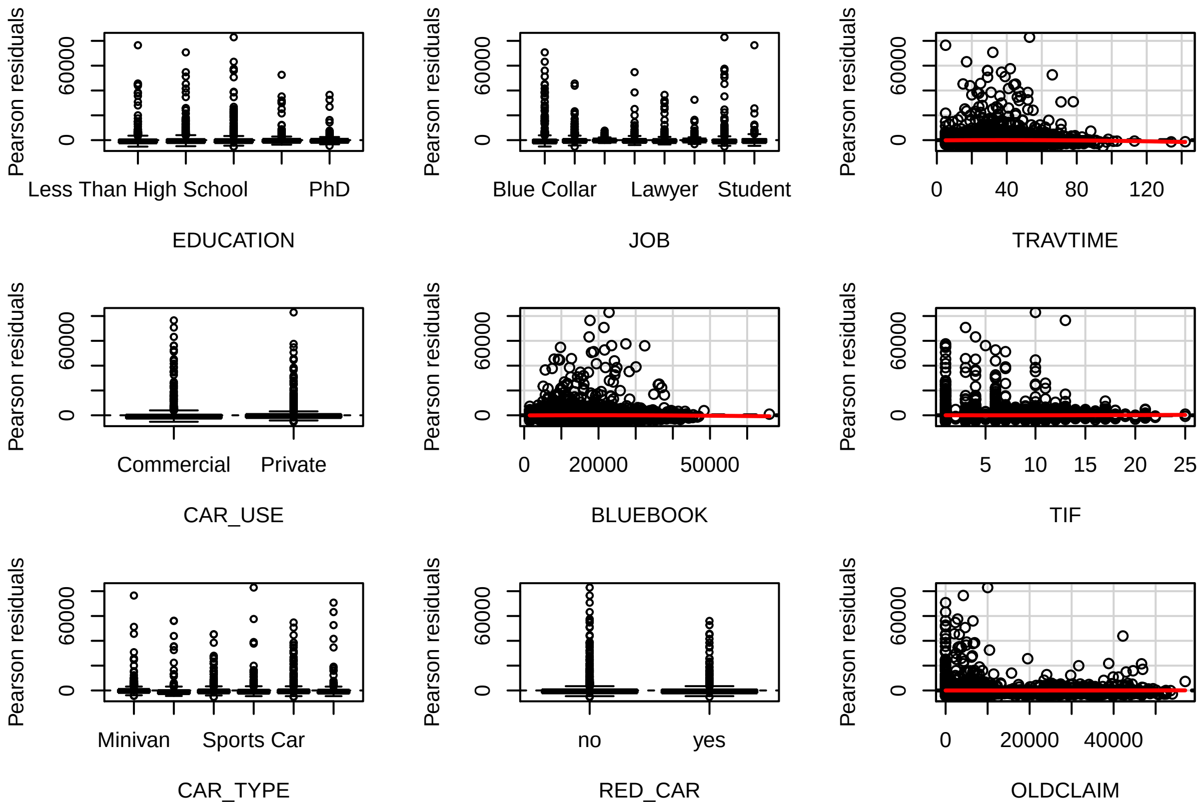
Statistic	N	Mean	St. Dev.	Min	Max
TARGET_FLAG	8,161	0.264	0.441	0	1
TARGET_AMT	8,161	1,504.325	4,704.027	0.000	107,586.100
KIDSDRIV	8,161	0.171	0.512	0	4
AGE	8,161	44.782	8.630	16.000	81.000
HOMEKIDS	8,161	0.721	1.116	0	5
YOJ	8,161	10.498	4.037	0.000	23.000
INCOME	8,161	61,568.910	47,152.790	0.000	367,030.000
HOME_VAL	8,161	154,998.700	127,487.100	0.000	885,282.000
TRAVTIME	8,161	33.486	15.908	5	142
BLUEBOOK	8,161	15,709.900	8,419.734	1,500	69,740
TIF	8,161	5.351	4.147	1	25
OLDCLAIM	8,161	4,037.076	8,777.139	0	57,037
CLM_FREQ	8,161	0.799	1.158	0	5
MVR_PTS	8,161	1.696	2.147	0	13
CAR_AGE	8,161	8.350	5.602	0.000	28.000

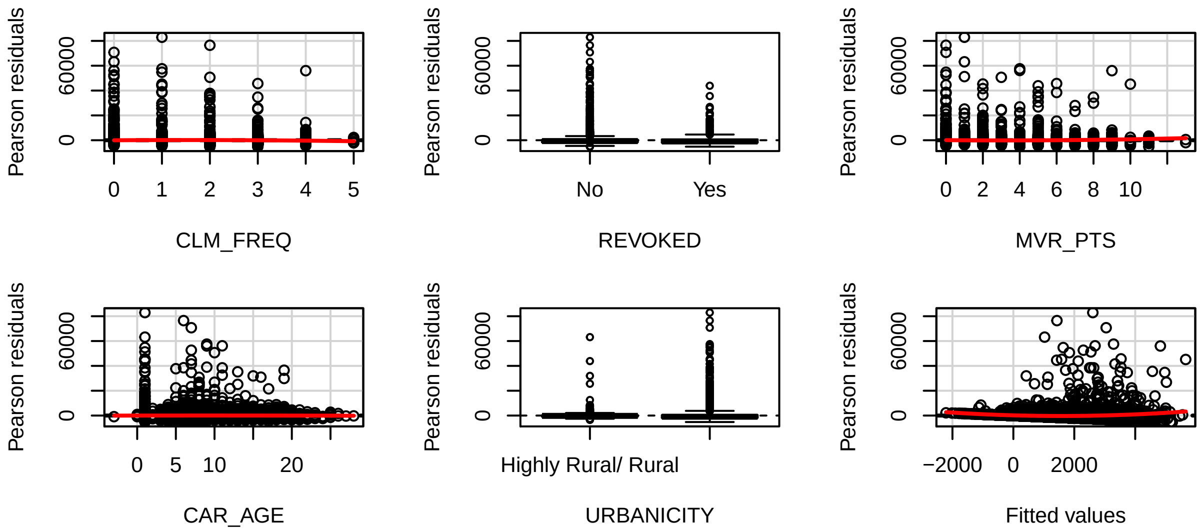
4.2 BoxCox Transformations

The Box-Cox transformations were done only on three of the input variables - income, house value, and the total number of claims during the past 5 years. These transformations were done based on the residual plots. In the residual plots, these three variables showed a great deal of non-constant variance because the plots were funnel-shaped.

```
## Non-constant Variance Score Test
## Variance formula: ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 + HOME_VAL + MSTATUS + SEX +
## Chisquare = 3755.09    Df = 36     p = 0
```




Pearson residuals
0
60000
Less Than High School
PhD
EDUCATION
Blue Collar
Lawyer
Student
JOB
0
40
80
120
TRAVTIME
Commercial
Private
CAR_USE
0
20000
50000
BLUEBOOK
5
10
15
20
25
TIF
Minivan
Sports Car
CAR_TYPE
no
yes
RED_CAR
0
20000
40000
OLDCLAIM




```
##              Test stat Pr(>|t|)
## KIDSDRIV        -1.733    0.083
## AGE              4.364    0.000
## HOMEKIDS        -2.420    0.016
## YOJ              1.003    0.316
## INCOME           0.513    0.608
## PARENT1             NA       NA
## HOME_VAL         1.483    0.138
## MSTATUS             NA       NA
## SEX                 NA       NA
## EDUCATION           NA       NA
## JOB                 NA       NA
## TRAVTIME        -1.033    0.302
## CAR_USE             NA       NA
## BLUEBOOK        -0.618    0.536
## TIF              0.467    0.640
## CAR_TYPE            NA       NA
## RED_CAR             NA       NA
## OLDCLAIM         0.171    0.864
## CLM_FREQ        -1.784    0.074
## REVOKED             NA       NA
## MVR_PTS          1.987    0.047
## CAR_AGE         -0.277    0.782
## URBANICITY          NA       NA
## Tukey test       8.214    0.000
```

Using the `BoxCox.lambda` function from the `forecast` package we are able to determine our necessary transformations to our independent variables.

λ	Variables
0.268842617694589	INCOME
0.505233636014921	HOME_VAL
0.0650958785108903	OLDCLAIM

Utilizing the below table of common transformations based on the lambda value of the BoxCox we further transform our independent variables.

Common Box-Cox Transformations[5] [6]

λ	Y'
-2	$Y^{-2} = \frac{1}{Y^2}$
-1	$Y^{-1} = \frac{1}{Y^1}$
-0.5	$Y^{-0.5} = \frac{1}{\sqrt{(Y)}}$
0	$\log(Y)$
.25	$\sqrt[4]{Y}$
0.5	$Y^{0.5} = \sqrt{(Y)}$
1	$Y^1 = Y$
2	Y^2

Lambda values were truncated to the nearest tenth that match a common transformation as per the below table.

variable	variable transformation
INCOME	$\sqrt[4]{INCOME}$
HOME VAL	$\sqrt{(HOME\,VAL)}$
OLDCLAIM	$\log(OLDCLAIM)$

[5] By Understanding Both the Concept of Transformation and the Box-Cox Method, Practitioners Will Be Better Prepared to Work with Non-normal Data. . “Making Data Normal Using Box-Cox Power Transformation.” ISixSigma. N.p., n.d. Web. 29 Oct. 2016.

[6] Osborne, Jason W. “Improving your data transformations: Applying the Box-Cox transformation.” Practical Assessment, Research & Evaluation 15.12 (2010): 1-9.

5 Models Built

5.1 Model 1 - Backwards Selection Method

In the backward step selection model The resulting AIC was .

6 TODO UPDATE THIS

6.0.1 Model Metrics for Backwards Selection

We first use an established threshold of .50 to determine our best possible threshold.

7 TODO Add Model Metrics

Table 6:

	fullModel
KIDSDRIV	0.065^{***}
	(0.009)
I(INCOME^(1/4))	-0.008^{***}
	(0.002)
PARENT1Yes	0.084^{***}
	(0.015)
I(sqrt(HOME_VAL))	-0.0001^{***}
	(0.00003)
MSTATUSYes	-0.059^{***}
	(0.012)
EDUCATIONHigh School	0.062^{***}
	(0.013)
EDUCATIONLess Than High School	0.054^{***}
	(0.016)
EDUCATIONMasters	0.023
	(0.017)
EDUCATIONPhD	0.042^{*}
	(0.022)
JOBClerical	0.012
	(0.016)
JOBDoctor	-0.134^{***}
	(0.037)
JOBHome Maker	-0.054^{**}
	(0.025)
JOBLawyer	-0.051^{**}
	(0.026)
JOBManager	-0.137^{***}
	(0.019)
JOBProfessional	-0.028
	(0.018)
JOBStudent	-0.069^{***}
	(0.023)
TRAVTIME	0.002^{***}
	(0.0003)
CAR_USEPrivate	-0.126^{***}
	(0.014)
BLUEBOOK	-0.00000^{***}
	(0.00000)
TIF	-0.008^{***}
	(0.001)
CAR_TYPEPanel Truck	0.070^{***}
	(0.022)
CAR_TYPEPickup	0.074^{***}
	(0.015)
CAR_TYPESports Car	0.132^{***}
	(0.016)
CAR_TYPESUV	0.097^{***}
	(0.012)
CAR_TYPEVan	0.080^{***}
	(0.018)
CLM_FREQ	0.024^{***}
	(0.004)
REVOKEDYes	0.128^{***}
	(0.013)
MVR_PTS	0.020^{***}
	(0.002)
URBANICITYHighly Urban/ Urban	0.299^{***}
	(0.012)
Constant	0.171^{***}
	(0.034)
N	8,161
Log Likelihood	$-3,826.099$
Akaike Inf. Crit.	7,712.198
Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

7.0.1 Multicollinearity for Backwards Selection

rn	GVIF	Df	GVIF^(1/(2*Df))
KIDSDRIV	1.077269	1	1.037916
I(INCOME^(1/4))	3.303477	1	1.817547
PARENT1	1.411040	1	1.187872
I(sqrt(HOME_VAL))	2.093420	1	1.446866
MSTATUS	2.012248	1	1.418537
EDUCATION	6.363135	4	1.260256
JOB	25.141215	7	1.259005
TRAVTIME	1.037090	1	1.018376
CAR_USE	2.409461	1	1.552244
BLUEBOOK	1.671917	1	1.293026
TIF	1.007017	1	1.003502
CAR_TYPE	2.483562	5	1.095236
CLM_FREQ	1.262974	1	1.123821
REVOKED	1.020286	1	1.010092
MVR_PTS	1.213958	1	1.101798
URBANICITY	1.250699	1	1.118347

7.1 Model 2 - Forwards Selection Method

```
## Start:  AIC=-13372.08
## TARGET_FLAG ~ 1
##
##                      Df Sum of Sq    RSS    AIC
## + I(log1p(OLDCLAIM))  1    91.126 1493.9 -13853
## + URBANICITY          1    79.707 1505.3 -13791
## + MVR_PTS             1    76.155 1508.8 -13772
## + CLM_FREQ            1    74.084 1510.9 -13761
## + I(sqrt(HOME_VAL))   1    51.744 1533.3 -13641
## + JOB                 7    46.882 1538.1 -13603
## + PARENT1             1    39.379 1545.6 -13575
## + REVOKED             1    36.591 1548.4 -13561
## + CAR_USE             1    32.264 1552.7 -13538
## + EDUCATION           4    33.068 1551.9 -13536
## + CAR_TYPE            5    33.092 1551.9 -13534
## + I(INCOME^(1/4))     1    30.058 1555.0 -13526
## + MSTATUS             1    28.940 1556.1 -13520
## + HOMEKIDS            1    21.189 1563.8 -13480
## + CAR_AGE             1    17.493 1567.5 -13461
## + AGE                 1    17.312 1567.7 -13460
## + KIDSDRIV            1    17.034 1568.0 -13458
## + BLUEBOOK            1    16.941 1568.1 -13458
## + TIF                 1    10.754 1574.2 -13426
## + YOJ                 1     8.904 1576.1 -13416
## + TRAVTIME            1     3.708 1581.3 -13389
## + SEX                 1     0.704 1584.3 -13374
## <none>                            1585.0 -13372
## + RED_CAR             1     0.076 1584.9 -13370
##
## Step:  AIC=-13853.3
```

```
## TARGET_FLAG ~ I(log1p(OLDCLAIM))
##
##                     Df Sum of Sq    RSS    AIC
## + URBANICITY         1    43.757 1450.1 -14094
## + JOB                7    42.071 1451.8 -14072
## + I(sqrt(HOME_VAL))  1    39.338 1454.5 -14069
## + PARENT1            1    32.703 1461.2 -14032
## + EDUCATION          4    29.452 1464.4 -14008
## + I(INCOME^(1/4))    1    25.386 1468.5 -13991
## + CAR_USE            1    24.356 1469.5 -13986
## + CAR_TYPE           5    25.360 1468.5 -13983
## + REVOKED            1    23.703 1470.2 -13982
## + MVR_PTS            1    23.659 1470.2 -13982
## + MSTATUS            1    22.170 1471.7 -13973
## + HOMEKIDS           1    17.235 1476.6 -13946
## + CAR_AGE            1    15.920 1478.0 -13939
## + AGE                1    14.340 1479.5 -13930
## + BLUEBOOK           1    13.822 1480.1 -13927
## + KIDSDRIV           1    13.795 1480.1 -13927
## + TIF                1     8.937 1484.9 -13900
## + YOJ                1     7.773 1486.1 -13894
## + TRAVTIME           1     3.924 1490.0 -13873
## + SEX                1     0.748 1493.1 -13855
## + CLM_FREQ           1     0.630 1493.2 -13855
## <none>                           1493.9 -13853
## + RED_CAR            1     0.176 1493.7 -13852
##
## Step:  AIC=-14093.91
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY
##
##                     Df Sum of Sq    RSS    AIC
## + JOB                7    81.521 1368.6 -14552
## + EDUCATION          4    52.640 1397.5 -14388
## + I(sqrt(HOME_VAL))  1    51.137 1399.0 -14385
## + I(INCOME^(1/4))    1    43.607 1406.5 -14341
## + PARENT1            1    35.824 1414.3 -14296
## + CAR_TYPE           5    30.054 1420.1 -14255
## + CAR_AGE            1    27.930 1422.2 -14251
## + CAR_USE            1    27.331 1422.8 -14247
## + MSTATUS            1    23.265 1426.9 -14224
## + HOMEKIDS           1    22.001 1428.1 -14217
## + MVR_PTS            1    21.724 1428.4 -14215
## + REVOKED            1    20.418 1429.7 -14208
## + BLUEBOOK           1    19.549 1430.6 -14203
## + AGE                1    17.793 1432.3 -14193
## + KIDSDRIV           1    16.441 1433.7 -14185
## + YOJ                1    11.590 1438.5 -14157
## + TRAVTIME           1     9.962 1440.2 -14148
## + TIF                1     9.578 1440.5 -14146
## + SEX                1     1.504 1448.6 -14100
## + RED_CAR            1     0.505 1449.6 -14095
## + CLM_FREQ           1     0.448 1449.7 -14094
## <none>                           1450.1 -14094
##
```

```
## Step:  AIC=-14552.1
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB
##
##                     Df Sum of Sq    RSS    AIC
## + PARENT1            1   28.2209 1340.4 -14720
## + I(sqrt(HOME_VAL))  1   26.9240 1341.7 -14712
## + MSTATUS            1   26.7036 1341.9 -14711
## + CAR_TYPE           5   25.3117 1343.3 -14694
## + REVOKED            1   17.9545 1350.7 -14658
## + MVR_PTS            1   17.6223 1351.0 -14656
## + CAR_USE            1   16.1366 1352.5 -14647
## + KIDSDRIV           1   12.3905 1356.2 -14624
## + I(INCOME^(1/4))    1   10.7114 1357.9 -14614
## + TIF                1    9.4325 1359.2 -14606
## + HOMEKIDS           1    8.7559 1359.8 -14602
## + TRAVTIME           1    7.5972 1361.0 -14596
## + BLUEBOOK           1    4.5644 1364.0 -14577
## + AGE                1    4.4488 1364.2 -14577
## + EDUCATION          4    5.1250 1363.5 -14575
## + YOJ                1    3.6560 1364.9 -14572
## + CAR_AGE            1    0.5663 1368.0 -14554
## + CLM_FREQ           1    0.5374 1368.1 -14553
## <none>                          1368.6 -14552
## + SEX                1    0.3280 1368.3 -14552
## + RED_CAR            1    0.0989 1368.5 -14551
##
## Step:  AIC=-14720.14
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1
##
##                     Df Sum of Sq    RSS    AIC
## + CAR_TYPE           5   24.7941 1315.6 -14862
## + CAR_USE            1   16.8799 1323.5 -14822
## + REVOKED            1   16.1451 1324.2 -14817
## + MVR_PTS            1   15.7526 1324.6 -14815
## + I(sqrt(HOME_VAL))  1   14.3825 1326.0 -14806
## + I(INCOME^(1/4))    1    9.6391 1330.7 -14777
## + TIF                1    9.4139 1331.0 -14776
## + MSTATUS            1    8.8757 1331.5 -14772
## + TRAVTIME           1    8.6134 1331.8 -14771
## + KIDSDRIV           1    6.5708 1333.8 -14758
## + EDUCATION          4    5.7573 1334.6 -14747
## + BLUEBOOK           1    4.0308 1336.3 -14743
## + YOJ                1    2.9466 1337.4 -14736
## + CLM_FREQ           1    0.5568 1339.8 -14722
## + HOMEKIDS           1    0.4799 1339.9 -14721
## + CAR_AGE            1    0.4700 1339.9 -14721
## <none>                          1340.4 -14720
## + AGE                1    0.2876 1340.1 -14720
## + SEX                1    0.0316 1340.3 -14718
## + RED_CAR            1    0.0074 1340.4 -14718
##
## Step:  AIC=-14862.52
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE
```

```
##
##                      Df Sum of Sq    RSS    AIC
## + REVOKED             1   15.3532 1300.2 -14956
## + I(sqrt(HOME_VAL))  1   14.7624 1300.8 -14953
## + MVR_PTS             1   14.6198 1301.0 -14952
## + CAR_USE             1   12.7107 1302.9 -14940
## + TIF                 1    9.7233 1305.9 -14921
## + I(INCOME^(1/4))     1    9.5224 1306.1 -14920
## + MSTATUS             1    9.1986 1306.4 -14918
## + TRAVTIME            1    8.5626 1307.0 -14914
## + KIDSDRIV            1    6.5791 1309.0 -14901
## + BLUEBOOK            1    4.6785 1310.9 -14890
## + EDUCATION           4    5.0314 1310.5 -14886
## + YOJ                 1    2.7159 1312.9 -14877
## + SEX                 1    1.3675 1314.2 -14869
## + CLM_FREQ            1    0.4980 1315.1 -14864
## + CAR_AGE             1    0.4966 1315.1 -14864
## + RED_CAR             1    0.4599 1315.1 -14863
## + AGE                 1    0.3742 1315.2 -14863
## + HOMEKIDS            1    0.3657 1315.2 -14863
## <none>                            1315.6 -14862
##
## Step:  AIC=-14956.32
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED
##
##                      Df Sum of Sq    RSS    AIC
## + MVR_PTS             1   15.0377 1285.2 -15049
## + I(sqrt(HOME_VAL))  1   13.9393 1286.3 -15042
## + CAR_USE             1   12.3640 1287.9 -15032
## + I(INCOME^(1/4))     1    9.1154 1291.1 -15012
## + TIF                 1    9.0088 1291.2 -15011
## + MSTATUS             1    8.7226 1291.5 -15009
## + TRAVTIME            1    8.6019 1291.6 -15008
## + KIDSDRIV            1    5.9732 1294.3 -14992
## + BLUEBOOK            1    4.5516 1295.7 -14983
## + EDUCATION           4    4.7785 1295.5 -14978
## + YOJ                 1    2.6325 1297.6 -14971
## + CLM_FREQ            1    1.3260 1298.9 -14963
## + SEX                 1    1.2547 1299.0 -14962
## + CAR_AGE             1    0.5034 1299.7 -14958
## + RED_CAR             1    0.4159 1299.8 -14957
## <none>                            1300.2 -14956
## + AGE                 1    0.2962 1299.9 -14956
## + HOMEKIDS            1    0.2766 1300.0 -14956
##
## Step:  AIC=-15049.25
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS
##
##                      Df Sum of Sq    RSS    AIC
## + I(sqrt(HOME_VAL))  1   13.3944 1271.8 -15133
## + CAR_USE             1   11.3872 1273.8 -15120
## + MSTATUS             1    8.7625 1276.4 -15103
```

```
## + I(INCOME^(1/4))    1    8.2944 1276.9 -15100
## + TIF                1    8.2887 1276.9 -15100
## + TRAVTIME           1    8.1748 1277.0 -15099
## + KIDSDRIV           1    5.4184 1279.8 -15082
## + BLUEBOOK           1    4.3746 1280.8 -15075
## + EDUCATION          4    4.7229 1280.5 -15071
## + YOJ                1    2.2444 1283.0 -15062
## + CLM_FREQ           1    1.3511 1283.8 -15056
## + SEX                1    1.3307 1283.9 -15056
## + CAR_AGE            1    0.5563 1284.6 -15051
## + RED_CAR            1    0.4122 1284.8 -15050
## <none>                           1285.2 -15049
## + HOMEKIDS           1    0.1984 1285.0 -15048
## + AGE                1    0.1456 1285.0 -15048
##
## Step:  AIC=-15132.75
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL))
##
##                   Df Sum of Sq    RSS    AIC
## + CAR_USE          1   11.7618 1260.0 -15207
## + TIF              1    8.3955 1263.4 -15185
## + TRAVTIME         1    7.9114 1263.9 -15182
## + KIDSDRIV         1    6.8245 1265.0 -15175
## + I(INCOME^(1/4))  1    5.2098 1266.6 -15164
## + BLUEBOOK         1    3.5762 1268.2 -15154
## + EDUCATION        4    4.2448 1267.6 -15152
## + CLM_FREQ         1    1.2374 1270.6 -15139
## + MSTATUS          1    1.1890 1270.6 -15138
## + SEX              1    1.1379 1270.7 -15138
## + YOJ              1    1.0236 1270.8 -15137
## + HOMEKIDS         1    0.9844 1270.8 -15137
## + CAR_AGE          1    0.4976 1271.3 -15134
## <none>                          1271.8 -15133
## + RED_CAR          1    0.2279 1271.6 -15132
## + AGE              1    0.0741 1271.7 -15131
##
## Step:  AIC=-15206.58
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE
##
##                   Df Sum of Sq    RSS    AIC
## + TIF              1    8.4863 1251.5 -15260
## + TRAVTIME         1    7.8466 1252.2 -15256
## + KIDSDRIV         1    7.3123 1252.7 -15252
## + I(INCOME^(1/4))  1    5.9798 1254.1 -15243
## + BLUEBOOK         1    4.2151 1255.8 -15232
## + EDUCATION        4    4.6828 1255.4 -15229
## + SEX              1    1.3520 1258.7 -15213
## + HOMEKIDS         1    1.1816 1258.9 -15212
## + CLM_FREQ         1    1.1064 1258.9 -15212
## + CAR_AGE          1    1.0820 1259.0 -15212
## + MSTATUS          1    0.9770 1259.1 -15211
## + YOJ              1    0.8663 1259.2 -15210
```

```
## <none>                          1260.0 -15207
## + RED_CAR          1    0.2745 1259.8 -15206
## + AGE              1    0.1035 1259.9 -15205
##
## Step:  AIC=-15259.73
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF
##
##                   Df Sum of Sq    RSS    AIC
## + TRAVTIME         1    7.7180 1243.8 -15308
## + KIDSDRIV         1    7.3385 1244.2 -15306
## + I(INCOME^(1/4))  1    6.0240 1245.5 -15297
## + BLUEBOOK         1    4.3350 1247.2 -15286
## + EDUCATION        4    4.8001 1246.8 -15283
## + SEX              1    1.3897 1250.2 -15267
## + HOMEKIDS         1    1.2909 1250.3 -15266
## + CLM_FREQ         1    1.1264 1250.4 -15265
## + MSTATUS          1    0.9996 1250.5 -15264
## + CAR_AGE          1    0.9886 1250.6 -15264
## + YOJ              1    0.8347 1250.7 -15263
## <none>                          1251.5 -15260
## + RED_CAR          1    0.2680 1251.3 -15260
## + AGE              1    0.1043 1251.5 -15258
##
## Step:  AIC=-15308.21
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF + TRAVTIME
##
##                   Df Sum of Sq    RSS    AIC
## + KIDSDRIV         1    7.2611 1236.6 -15354
## + I(INCOME^(1/4))  1    6.1080 1237.7 -15346
## + BLUEBOOK         1    4.3495 1239.5 -15335
## + EDUCATION        4    5.0983 1238.7 -15334
## + HOMEKIDS         1    1.3707 1242.5 -15315
## + SEX              1    1.3458 1242.5 -15315
## + CAR_AGE          1    1.0543 1242.8 -15313
## + MSTATUS          1    1.0200 1242.8 -15313
## + CLM_FREQ         1    0.9819 1242.8 -15313
## + YOJ              1    0.8875 1243.0 -15312
## <none>                          1243.8 -15308
## + RED_CAR          1    0.2310 1243.6 -15308
## + AGE              1    0.1371 1243.7 -15307
##
## Step:  AIC=-15353.99
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF + TRAVTIME + KIDSDRIV
##
##                   Df Sum of Sq    RSS    AIC
## + I(INCOME^(1/4))  1    5.9989 1230.6 -15392
## + BLUEBOOK         1    4.4580 1232.1 -15382
## + EDUCATION        4    4.9570 1231.6 -15379
```

```
## + MSTATUS          1    1.9446 1234.6 -15365
## + SEX              1    1.5045 1235.1 -15362
## + YOJ              1    1.2405 1235.3 -15360
## + CAR_AGE          1    1.0094 1235.6 -15359
## + CLM_FREQ         1    0.9471 1235.6 -15358
## + RED_CAR          1    0.3077 1236.3 -15354
## <none>                         1236.6 -15354
## + AGE              1    0.1411 1236.4 -15353
## + HOMEKIDS         1    0.0017 1236.6 -15352
##
## Step:  AIC=-15391.68
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF + TRAVTIME + KIDSDRIV + I(INCOME^(1/4))
##
##             Df Sum of Sq    RSS    AIC
## + MSTATUS    1    3.1003 1227.5 -15410
## + BLUEBOOK   1    2.9133 1227.7 -15409
## + EDUCATION  4    3.6428 1226.9 -15408
## + SEX        1    1.0274 1229.5 -15396
## + CLM_FREQ   1    0.9053 1229.7 -15396
## + CAR_AGE    1    0.5260 1230.0 -15393
## <none>                  1230.6 -15392
## + RED_CAR    1    0.1793 1230.4 -15391
## + AGE        1    0.1099 1230.5 -15390
## + YOJ        1    0.0058 1230.6 -15390
## + HOMEKIDS   1    0.0006 1230.6 -15390
##
## Step:  AIC=-15410.26
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF + TRAVTIME + KIDSDRIV + I(INCOME^(1/4)) + MSTATUS
##
##             Df Sum of Sq    RSS    AIC
## + EDUCATION  4    3.9592 1223.5 -15429
## + BLUEBOOK   1    2.9446 1224.5 -15428
## + SEX        1    0.9623 1226.5 -15415
## + CLM_FREQ   1    0.8507 1226.6 -15414
## + CAR_AGE    1    0.5957 1226.9 -15412
## <none>                  1227.5 -15410
## + AGE        1    0.2229 1227.2 -15410
## + YOJ        1    0.1983 1227.3 -15410
## + HOMEKIDS   1    0.1929 1227.3 -15410
## + RED_CAR    1    0.1689 1227.3 -15409
##
## Step:  AIC=-15428.63
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF + TRAVTIME + KIDSDRIV + I(INCOME^(1/4)) + MSTATUS + EDUCATION
##
##            Df Sum of Sq    RSS    AIC
## + BLUEBOOK  1   2.85270 1220.7 -15446
## + SEX       1   1.00718 1222.5 -15433
## + CLM_FREQ  1   0.86993 1222.7 -15432
```

```
## <none>                   1223.5 -15429
## + AGE       1   0.21601 1223.3 -15428
## + RED_CAR   1   0.20655 1223.3 -15428
## + HOMEKIDS  1   0.17402 1223.3 -15428
## + YOJ       1   0.10093 1223.4 -15427
## + CAR_AGE   1   0.05030 1223.5 -15427
##
## Step:  AIC=-15445.68
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF + TRAVTIME + KIDSDRIV + I(INCOME^(1/4)) + MSTATUS + EDUCATION +
##     BLUEBOOK
##
##            Df Sum of Sq    RSS    AIC
## + CLM_FREQ  1   0.81994 1219.8 -15449
## <none>                   1220.7 -15446
## + HOMEKIDS  1   0.14918 1220.5 -15445
## + SEX       1   0.11415 1220.5 -15444
## + AGE       1   0.08653 1220.6 -15444
## + YOJ       1   0.06180 1220.6 -15444
## + CAR_AGE   1   0.04759 1220.6 -15444
## + RED_CAR   1   0.00336 1220.7 -15444
##
## Step:  AIC=-15449.16
## TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY + JOB + PARENT1 +
##     CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) + CAR_USE +
##     TIF + TRAVTIME + KIDSDRIV + I(INCOME^(1/4)) + MSTATUS + EDUCATION +
##     BLUEBOOK + CLM_FREQ
##
##            Df Sum of Sq    RSS    AIC
## <none>                   1219.8 -15449
## + HOMEKIDS  1  0.159901 1219.7 -15448
## + SEX       1  0.105852 1219.7 -15448
## + AGE       1  0.092803 1219.8 -15448
## + YOJ       1  0.070188 1219.8 -15448
## + CAR_AGE   1  0.048674 1219.8 -15448
## + RED_CAR   1  0.001518 1219.8 -15447

##
## Call:
## lm(formula = TARGET_FLAG ~ I(log1p(OLDCLAIM)) + URBANICITY +
##     JOB + PARENT1 + CAR_TYPE + REVOKED + MVR_PTS + I(sqrt(HOME_VAL)) +
##     CAR_USE + TIF + TRAVTIME + KIDSDRIV + I(INCOME^(1/4)) + MSTATUS +
##     EDUCATION + BLUEBOOK + CLM_FREQ, data = wimputedDfTr)
##
## Coefficients:
##                  (Intercept)             I(log1p(OLDCLAIM))
##                    1.708e-01                      2.468e-03
##  URBANICITYHighly Urban/ Urban                    JOBClerical
##                    2.972e-01                      1.146e-02
##                    JOBDoctor                  JOBHome Maker
##                   -1.334e-01                     -5.387e-02
##                    JOBLawyer                     JOBManager
##                   -5.151e-02                     -1.366e-01
```

```
##                   JOBProfessional                         JOBStudent
##                        -2.813e-02                         -6.884e-02
##                        PARENT1Yes                CAR_TYPEPanel Truck
##                         8.418e-02                          6.940e-02
##                    CAR_TYPEPickup                CAR_TYPESports Car
##                         7.426e-02                          1.319e-01
##                       CAR_TYPESUV                        CAR_TYPEVan
##                         9.639e-02                          8.002e-02
##                        REVOKEDYes                            MVR_PTS
##                         1.253e-01                          1.963e-02
##                 I(sqrt(HOME_VAL))                     CAR_USEPrivate
##                        -1.200e-04                         -1.258e-01
##                               TIF                           TRAVTIME
##                        -7.903e-03                          1.991e-03
##                          KIDSDRIV                   I(INCOME^(1/4))
##                         6.510e-02                         -8.369e-03
##                        MSTATUSYes              EDUCATIONHigh School
##                        -5.888e-02                          6.167e-02
## EDUCATIONLess Than High School                   EDUCATIONMasters
##                         5.323e-02                          2.369e-02
##                      EDUCATIONPhD                           BLUEBOOK
##                         4.210e-02                         -2.846e-06
##                          CLM_FREQ
##                         1.705e-02
```

Table 8:

	TARGET_FLAG
I(log1p(OLDCLAIM))	0.020
	(0.012)
URBANICITYHighly Urban/ Urban	2.383***
	(0.113)
I(sqrt(HOME_VAL))	−0.001***
	(0.0002)
CAR_USEPrivate	−0.799***
	(0.086)
BLUEBOOK	−0.00002***
	(0.00000)
REVOKEDYes	0.709***
	(0.081)
MVR_PTS	0.102***
	(0.014)
PARENT1Yes	0.427***
	(0.094)
I(INCOME^(1/4))	−0.071***
	(0.012)
JOBClerical	0.114
	(0.106)
JOBDoctor	−0.720***
	(0.233)
JOBHome Maker	−0.450***
	(0.164)
JOBLawyer	−0.224
	(0.141)
JOBManager	−0.801***
	(0.112)
JOBProfessional	−0.266**
	(0.109)
JOBStudent	−0.511***
	(0.151)
TRAVTIME	0.014***
	(0.002)
CAR_TYPEPanel Truck	0.703***
	(0.148)
CAR_TYPEPickup	0.597***
	(0.100)
CAR_TYPESports Car	0.958***
	(0.107)
CAR_TYPESUV	0.726***
	(0.086)
CAR_TYPEVan	0.675***
	(0.121)
TIF	−0.055***
	(0.007)
KIDSDRIV	0.418***
	(0.055)
MSTATUSYes	−0.430***
	(0.085)
CAR_AGE	−0.012*
	(0.007)
CLM_FREQ	0.093**
	(0.043)
YOJ	0.015
	(0.010)
Constant	−1.947***
	(0.227)
N	8,161
Log Likelihood	−3,655.423
Akaike Inf. Crit.	7,368.846
Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

7.1.1 Model Metrics for Forwards Selection

We first use an established threshold of .50 to determine our best possible threshold.

7.1.2 Multicollinearity for Forwards Selection

Here in our forward selection model we find that no variable exceeds our pre-established threshold of 5 for multicollinearity.

rn	GVIF	Df	GVIFˆ(1/(2*Df))
KIDSDRIV	1.077269	1	1.037916
I(INCOMEˆ(1/4))	3.303477	1	1.817547
PARENT1	1.411040	1	1.187872
I(sqrt(HOME_VAL))	2.093420	1	1.446866
MSTATUS	2.012248	1	1.418537
EDUCATION	6.363135	4	1.260256
JOB	25.141215	7	1.259005
TRAVTIME	1.037090	1	1.018376
CAR_USE	2.409461	1	1.552244
BLUEBOOK	1.671917	1	1.293026
TIF	1.007017	1	1.003502
CAR_TYPE	2.483562	5	1.095236
CLM_FREQ	1.262974	1	1.123821
REVOKED	1.020286	1	1.010092
MVR_PTS	1.213958	1	1.101798
URBANICITY	1.250699	1	1.118347

7.2 Model 3 - Subset Selection Method

7.2.1 Subset Variable Selection

Using the `leaps` package and the `regsubsets` function we are able to subset our independent variables by looking at the best model for each predictor.

	1 (1)	2 (1)	3 (1)	4 (1)	5 (1)	6 (1)	7 (1)	8 (1)
KIDSDRIV								*
AGE								
HOMEKIDS								
YOJ								
I(INCOME^(1/4))					*	*	*	*
PARENT1Yes					*	*	*	
I(sqrt(HOME_VAL))		*	*	*				
MSTATUSYes								*
SEXM								
EDUCATIONHigh School								
EDUCATIONLess Than High School								
EDUCATIONMasters								
EDUCATIONPhD								
JOBClerical								
JOBDoctor								
JOBHome Maker								
JOBLawyer								
JOBManager							*	*
JOBProfessional								
JOBStudent								
TRAVTIME								
CAR_USEPrivate				*	*	*	*	*
BLUEBOOK								
TIF								
CAR_TYPEPanel Truck								
CAR_TYPEPickup								
CAR_TYPESports Car								
CAR_TYPESUV								
CAR_TYPEVan								
RED_CARyes								
I(log1p(OLDCLAIM))	*							
CLM_FREQ								
REVOKEDYes						*	*	*
MVR_PTS			*	*	*	*	*	*
CAR_AGE								
URBANICITYHighly Urban/ Urban		*	*	*	*	*	*	*

7.2.2 Subset Model

The variables as indicated in column 8 of the previous table will be further implement into our subset selection model in the following table. We don't see as strong of a relationship in our independent variables to the dependent variable in this model as our previous model. For example, the coefficient for tax was as high as 106 in the forward selection model but it is -5 in this model. However, our intercept in this model is larger than any other model.