Homework #4: Insurance Claim Prediction

Data 621 Business Analytics and Data Mining

Aadi Kalloo, Nathan Lim, Asher Meyers, Daniel Smilowitz, Logan Thomson

Due July 10, 2016

Contents

1 Data Exploration **2**

2 Data Preparation **8**

3 Model Creation **11**

3.1 Multiple Linear Regression 11

3.2 Binary Logistic Regression 14

3.3 **(Intercept)** -2.666 0.1688 -15.8 3.083e-56 22

4 Model Selection & Prediction **23**

Appendix A: Index-wise Results from Predictive Model **24**

Appendix B: R Code **61**

1 Data Exploration

The dataset of interest contains information about customers of an auto insurance company. The dataset has 8161 rows (each representing a customer) and 25 variables. There are 23 predictor variables and 2 response variables: `TARGET_FLAG`, a binary categorical variable representing whether each customer has been in an accident; and `TARGET_AMT`, a numerical variable indicating the cost of a crash that a customer was in. The class of variables read in from the dataset is presented below:

	Class	Levels
TARGET_FLAG	integer	-
TARGET_AMT	numeric	-
KIDSDRIV	integer	-
AGE	integer	-
HOMEKIDS	integer	-
YOJ	integer	-
INCOME	factor	6613
PARENT1	factor	2
HOME_VAL	factor	5107
MSTATUS	factor	2
SEX	factor	2
EDUCATION	factor	5
JOB	factor	9
TRAVTIME	integer	-
CAR_USE	factor	2
BLUEBOOK	factor	2789
TIF	integer	-
CAR_TYPE	factor	6
RED_CAR	factor	2
OLDCLAIM	factor	2857
CLM_FREQ	integer	-
REVOKED	factor	2
MVR_PTS	integer	-
CAR_AGE	integer	-
URBANICITY	factor	2

The very high number of levels for four of the variables (`INCOME`, `HOME_VAL`, `BLUEBOOK`, and `OLDCLAIM`) indicates that these variables are not in fact factors; investigation of the dataset indicates that these are dollar values interpreted as strings due to the presence of dollar signs and commas. The numerical values are extracted for these variables.

Additionally, there are 7 variables with only two levels. These are recast as binary variables as follows:

- `PARENT1`, `MSTATUS`, `RED_CAR`, and `REVOKED`: using 1 to indicate Yes
- `SEX`: using 1 to indicate Male
- `CAR_USE`: using 1 to indicate Commercial
- `URBANICITY`: using 1 to indicate Highly Urban/ Urban

Finally, there are three categorical variables – factors with more than two levels. Dummy variables are created for each of these, as follows:

- `EDUCATION`: 5 dummy variables
- `CAR_TYPE`: 6 dummy variables
- `JOB`: 8 dummy variables

A summary of each variable is presented below:

	MEAN	MEDIAN	IQR	SKEW	r_{FLAG}	r_{AMT}	NAs
TARGET_FLAG	0.26	0	1	1.07	1	0.54	0
TARGET_AMT	1504	0	1036	8.71	0.54	1	0
KIDSDRIV	0.17	0	0	3.35	0.09	0.05	0
AGE	44.79	45	12	-0.03	-0.11	-0.05	6
HOMEKIDS	0.72	0	1	1.34	0.11	0.06	0
YOJ	10.5	11	4	-1.2	-0.07	-0.02	454
INCOME	61898	54028	57889	1.19	-0.14	-0.06	445
PARENT1	0.13	0	0	2.17	0.16	0.1	0
HOME_VAL	154867	161160	238724	0.49	-0.18	-0.09	464
MSTATUS	0.6	1	1	-0.41	-0.13	-0.1	0
SEX	0.46	0	1	0.14	-0.02	0.01	0
TRAVTIME	33.49	33	22	0.45	0.05	0.03	0
CAR_USE	0.37	0	1	0.53	0.14	0.1	0
BLUEBOOK	15710	14440	11570	0.79	-0.11	0	0
TIF	5.35	4	6	0.89	-0.08	-0.04	0
RED_CAR	0.29	0	1	0.92	-0.02	0	0
OLDCLAIM	4037	0	4636	3.12	0.14	0.08	0
CLM_FREQ	0.8	0	2	1.21	0.22	0.12	0
REVOKED	0.12	0	0	2.3	0.15	0.06	0
MVR_PTS	1.7	1	3	1.35	0.23	0.14	0
CAR_AGE	8.33	8	11	0.28	-0.11	-0.06	510
URBANICITY	0.8	1	0	-1.46	0.22	0.12	0
HSDropout	0.15	0	0	1.99	0.06	0.04	0
HS	0.29	0	1	0.95	0.11	0.04	0
Bachelors	0.27	0	1	1.01	-0.05	-0.02	0
Masters	0.2	0	0	1.48	-0.09	-0.05	0
PhD	0.09	0	0	2.88	-0.06	-0.02	0
Minivan	0.26	0	1	1.08	-0.14	-0.08	0
Panel_Truck	0.08	0	0	3.03	0	0.04	0
Pickup	0.17	0	0	1.75	0.05	0.02	0
Sports_Car	0.11	0	0	2.47	0.06	0.03	0
Van	0.09	0	0	2.82	0	0.01	0
SUV	0.28	0	1	0.97	0.05	0.01	0
Blank_Job	0.06	0	0	3.55	-0.01	0.01	0
Professional	0.14	0	0	2.11	-0.04	0	0
Blue_Collar	0.22	0	0	1.33	0.1	0.07	0
Clerical	0.16	0	0	1.9	0.04	0	0
Doctor	0.03	0	0	5.49	-0.05	-0.03	0
Lawyer	0.1	0	0	2.62	-0.06	-0.03	0
Manager	0.12	0	0	2.32	-0.12	-0.07	0
Home_Maker	0.08	0	0	3.13	0.01	0	0
Student	0.09	0	0	2.92	0.07	0.02	0

From the table above, it is clear that there are four variables with missing values, with the proportion of values missing ranging from less than $< 0.1\%$ to roughly 6.2%; these missing values will need to either be imputed or excluded from the dataset before modeling. The variables exhibit varying levels of skewness, with a few extreme values.

The large number of binary variables in the dataset makes graphical visualization of the distribution of all variables not particularly useful. The proportion of binary variables having a value of 0 or 1 is presented in the table below:

	0	1
TARGET_FLAG	0.74	0.26
PARENT1	0.87	0.13
MSTATUS	0.4	0.6
SEX	0.54	0.46
CAR_USE	0.63	0.37
RED_CAR	0.71	0.29
REVOKED	0.88	0.12
URBANICITY	0.2	0.8
HSDropout	0.85	0.15
HS	0.71	0.29
Bachelors	0.73	0.27
Masters	0.8	0.2
PhD	0.91	0.09
Minivan	0.74	0.26
Panel_Truck	0.92	0.08
Pickup	0.83	0.17
Sports_Car	0.89	0.11
Van	0.91	0.09
SUV	0.72	0.28
Blank_Job	0.94	0.06
Professional	0.86	0.14
Blue_Collar	0.78	0.22
Clerical	0.84	0.16
Doctor	0.97	0.03
Lawyer	0.9	0.1
Manager	0.88	0.12
Home_Maker	0.92	0.08
Student	0.91	0.09

The remaining variables are vizualized below in boxplots:

The boxplots illustrate the high skewness of the distributions of the predictors `KIDSDRIV`, `INCOME`, `HOME_VAL`, `TRAVTIME`, `BLUEBOOK`, `TIF`, `OLDCLAIM` and `MVR_PTS`. The target `TARGET_AMT` is also highly skewed – this makes sense, as this value is 0 for any customers without claims. Density plots of these variables are presented below:

Since `TARGET_AMT` and `OLDCLAIM` have such high concentrations at values of zero, separate plots for these two variables are created with values of zero removed:

The 8 predictors with the highest correlation to `TARGET_FLAG` and the 8 predictors with the highest correlation to `TARGET_AMT` share 7 predictors. The correlation between these variables is investigated and plotted below:

There appears to be evidence of possible multicollinearity between `HOME_VAL` and `OLDCLAIM`.

Finally, boxplots are also prepared for non-binary variables split by `TARGET_FLAG`:

Interestingly, there are only a few variables with immediately visible difference in median value based on `TARGET_FLAG` values, while the range of predictor values for each flag value differs noticeably.

2 Data Preparation

As stated in Part 1, four predictor variables have missing values. Because these missing values represent, in many cases, a non-menial proportion of the dataset at large, they are imputed so the cases containing missing values can be included in modeling. Due to the skewness illustrated by some of the variables with missing data (`INCOME` and `HOME_VAL`), the median is used to avoid any bias introduced into the mean by the skewness of these variables' distribution. For the remaining two variables, the median is also used for imputation for consistency.

Additionally, there is one instance of a vehicle's `CAR_AGE` being -3. Since this variable represents the age of the car in years, this is a nonsensical value. This instance of this variable is removed before imputation.

A summary of the variables with imputation of median values is presented below:

	MEAN	MEDIAN	IQR	SKEW	r_{FLAG}	r_{AMT}	NAs
TARGET_FLAG	0.26	0	1	1.07	1	0.53	0
TARGET_AMT	1504	0	1036	8.71	0.53	1	0
KIDSDRIV	0.17	0	0	3.35	0.1	0.06	0
AGE	44.79	45	12	-0.03	-0.1	-0.04	0
HOMEKIDS	0.72	0	1	1.34	0.12	0.06	0
YOJ	10.53	11	4	-1.26	-0.07	-0.02	0
INCOME	61469	54028	53597	1.24	-0.14	-0.06	0
PARENT1	0.13	0	0	2.17	0.16	0.1	0
HOME_VAL	155225	161160	233352	0.49	-0.18	-0.08	0
MSTATUS	0.6	1	1	-0.41	-0.14	-0.09	0
SEX	0.46	0	1	0.14	-0.02	0.01	0
TRAVTIME	33.49	33	22	0.45	0.05	0.03	0
CAR_USE	0.37	0	1	0.53	0.14	0.1	0
BLUEBOOK	15710	14440	11570	0.79	-0.1	0	0
TIF	5.35	4	6	0.89	-0.08	-0.05	0
RED_CAR	0.29	0	1	0.92	-0.01	0.01	0
OLDCLAIM	4037	0	4636	3.12	0.14	0.07	0
CLM_FREQ	0.8	0	2	1.21	0.22	0.12	0
REVOKED	0.12	0	0	2.3	0.15	0.06	0
MVR_PTS	1.7	1	3	1.35	0.22	0.14	0
CAR_AGE	8.31	8	8	0.3	-0.1	-0.06	0
URBANICITY	0.8	1	0	-1.46	0.22	0.12	0
HSDropout	0.15	0	0	1.99	0.05	0.03	0
HS	0.29	0	1	0.95	0.11	0.04	0
Bachelors	0.27	0	1	1.01	-0.04	-0.02	0
Masters	0.2	0	0	1.48	-0.08	-0.04	0
PhD	0.09	0	0	2.88	-0.07	-0.02	0
Minivan	0.26	0	1	1.08	-0.14	-0.08	0
Panel_Truck	0.08	0	0	3.03	0	0.03	0
Pickup	0.17	0	0	1.75	0.06	0.02	0
Sports_Car	0.11	0	0	2.47	0.06	0.02	0
Van	0.09	0	0	2.82	0	0.02	0
SUV	0.28	0	1	0.97	0.05	0.01	0
Blank_Job	0.06	0	0	3.55	0	0.02	0
Professional	0.14	0	0	2.11	-0.04	0	0
Blue_Collar	0.22	0	0	1.33	0.1	0.06	0
Clerical	0.16	0	0	1.9	0.03	0.01	0
Doctor	0.03	0	0	5.49	-0.06	-0.03	0
Lawyer	0.1	0	0	2.62	-0.06	-0.03	0
Manager	0.12	0	0	2.32	-0.11	-0.06	0
Home_Maker	0.08	0	0	3.13	0.01	-0.01	0
Student	0.09	0	0	2.92	0.08	0.02	0

The creation of additional or combined predictors did not yield any predictors that would improve models, however due to the skewness of some predictors, log transformations were conducted.

In the table below, we can see the effect of each of the log transformation on the correlations to the two target variables. For many variables, the transformations do not improve the correlation, but the log transformations of `INCOME` and `HOME_VAL` may fit our models better.

	r_{FLAG}	$\log r_{FLAG}$	r_{AMT}	$\log r_{AMT}$
TARGET_FLAG	1	1	0.53	0.53
TARGET_AMT	0.53	1	1	0.58
KIDSDRIV	0.1	0.11	0.06	0.06
AGE	-0.1	-0.12	-0.04	-0.05
HOMEKIDS	0.12	0.13	0.06	0.07
YOJ	-0.07	-0.08	-0.02	-0.02
INCOME	-0.14	-0.11	-0.06	-0.03
PARENT1	0.16	0.16	0.1	0.1
HOME_VAL	-0.18	-0.15	-0.08	-0.07
MSTATUS	-0.14	-0.14	-0.09	-0.09
SEX	-0.02	-0.02	0.01	0.01
TRAVTIME	0.05	0.06	0.03	0.03
CAR_USE	0.14	0.14	0.1	0.1
BLUEBOOK	-0.1	-0.11	0	0
TIF	-0.08	-0.08	-0.05	-0.05
RED_CAR	-0.01	-0.01	0.01	0.01
OLDCLAIM	0.14	0.24	0.07	0.13
CLM_FREQ	0.22	0.24	0.12	0.13
REVOKED	0.15	0.15	0.06	0.06
MVR_PTS	0.22	0.17	0.14	0.11
CAR_AGE	-0.1	-0.09	-0.06	-0.05
URBANICITY	0.22	0.22	0.12	0.12
HSDropout	0.05	0.05	0.03	0.03
HS	0.11	0.11	0.04	0.04
Bachelors	-0.04	-0.04	-0.02	-0.02
Masters	-0.08	-0.08	-0.04	-0.04
PhD	-0.07	-0.07	-0.02	-0.02
Minivan	-0.14	-0.14	-0.08	-0.08
Panel_Truck	0	0	0.03	0.03
Pickup	0.06	0.06	0.02	0.02
Sports_Car	0.06	0.06	0.02	0.02
Van	0	0	0.02	0.02
SUV	0.05	0.05	0.01	0.01
Blank_Job	0	0	0.02	0.02
Professional	-0.04	-0.04	0	0
Blue_Collar	0.1	0.1	0.06	0.06
Clerical	0.03	0.03	0.01	0.01
Doctor	-0.06	-0.06	-0.03	-0.03
Lawyer	-0.06	-0.06	-0.03	-0.03
Manager	-0.11	-0.11	-0.06	-0.06
Home_Maker	0.01	0.01	-0.01	-0.01
Student	0.08	0.08	0.02	0.02

3 Model Creation

3.1 Multiple Linear Regression

3.1.1 Full Model

	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	314.2	113.2	2.777	0.005507
AGE	5.221	7.066	0.7389	0.46
HOMEKIDS	77.64	65.35	1.188	0.2349
YOJ	-3.947	15.09	-0.2617	0.7936
INCOME	-0.004414	0.001805	-2.445	0.0145
PARENT1	576.1	202	2.852	0.00436
HOME_VAL	-0.0005574	0.0005908	-0.9436	0.3454
MSTATUS	-570.3	144.9	-3.937	8.326e-05
SEX	368.5	183.8	2.005	0.04498
TRAVTIME	11.95	3.222	3.709	0.0002093
CAR_USE	779.6	164.4	4.742	2.156e-06
BLUEBOOK	0.01433	0.008623	1.662	0.09657
TIF	-48.2	12.18	-3.958	7.629e-05
RED_CAR	-48.22	149.1	-0.3235	0.7463
OLDCLAIM	-0.01056	0.007436	-1.42	0.1557
CLM_FREQ	141.7	55.03	2.575	0.01003
REVOKED	549.6	173.5	3.167	0.001545
MVR_PTS	175.3	25.92	6.764	1.436e-11
CAR_AGE	-26.9	12.8	-2.102	0.0356
URBANICITY	1665	139.4	11.94	1.322e-32
HSDropout	-287.1	356	-0.8065	0.42
HS	-375.9	322.4	-1.166	0.2437
Bachelors	-545.6	288.8	-1.889	0.05895
Masters	-262.8	254.1	-1.035	0.3009
Minivan	-751.6	179.3	-4.192	2.797e-05
Panel_Truck	-486.5	340.6	-1.428	0.1532
Pickup	-375.9	199.2	-1.888	0.0591
Sports_Car	270.1	178.9	1.51	0.1311
Van	-237	267.1	-0.8873	0.3749
Blank_Job	-287.6	374	-0.7691	0.4418
Professional	169	273	0.6189	0.536
Blue_Collar	220	233.4	0.9427	0.3459
Clerical	241.5	236.8	1.02	0.3078
Doctor	-787.4	466.7	-1.687	0.09159
Lawyer	-56.84	352.3	-0.1613	0.8718
Manager	-766.4	290.9	-2.635	0.008431
Home_Maker	64.64	270.2	0.2393	0.8109
(Intercept)	149	553.9	0.2691	0.7879

(Dispersion parameter for gaussian family taken to be 20649542)

Null deviance:	1.806e+11 on 8160 degrees of freedom
Residual deviance:	1.677e+11 on 8123 degrees of freedom

3.1.2 Log-Transformed Full Model

	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	0.7714	0.1248	6.179	6.747e-10

	Estimate	Std. Error	t value	Pr(>|t|)
AGE	-0.00102	0.007795	-0.1308	0.8959
HOMEKIDS	0.09037	0.0721	1.253	0.2101
YOJ	-0.02749	0.01664	-1.652	0.09865
INCOME	-5.141e-06	1.991e-06	-2.581	0.00986
PARENT1	0.9998	0.2229	4.486	7.353e-06
HOME_VAL	-2.144e-06	6.517e-07	-3.29	0.001006
MSTATUS	-0.9227	0.1598	-5.774	8.02e-09
SEX	0.2003	0.2027	0.9877	0.3233
TRAVTIME	0.02571	0.003555	7.232	5.211e-13
CAR_USE	1.545	0.1814	8.518	1.918e-17
BLUEBOOK	-3.107e-05	9.513e-06	-3.266	0.001094
TIF	-0.1027	0.01344	-7.642	2.385e-14
RED_CAR	-0.03648	0.1644	-0.2218	0.8244
OLDCLAIM	-2.968e-05	8.203e-06	-3.619	0.0002981
CLM_FREQ	0.4048	0.06071	6.668	2.755e-11
REVOKED	1.955	0.1914	10.21	2.481e-24
MVR_PTS	0.2805	0.02859	9.81	1.364e-22
CAR_AGE	-0.004524	0.01412	-0.3205	0.7486
URBANICITY	3.834	0.1538	24.93	2.716e-132
HSDropout	0.29	0.3927	0.7383	0.4603
HS	0.3683	0.3556	1.036	0.3004
Bachelors	-0.4561	0.3186	-1.431	0.1524
Masters	-0.1411	0.2803	-0.5035	0.6146
Minivan	-1.353	0.1978	-6.84	8.497e-12
Panel_Truck	-0.6247	0.3758	-1.662	0.09649
Pickup	-0.4291	0.2197	-1.953	0.05085
Sports_Car	0.4992	0.1974	2.529	0.01145
Van	-0.4035	0.2947	-1.369	0.1709
Blank_Job	-0.9216	0.4126	-2.234	0.02551
Professional	-0.2786	0.3012	-0.9249	0.355
Blue_Collar	0.09541	0.2575	0.3705	0.711
Clerical	0.2844	0.2613	1.089	0.2764
Doctor	-1.434	0.5148	-2.785	0.005368
Lawyer	-0.5799	0.3887	-1.492	0.1357
Manager	-1.731	0.3209	-5.394	7.08e-08
Home_Maker	-0.001317	0.2981	-0.00442	0.9965
(Intercept)	-3.596	0.611	-5.885	4.127e-09

(Dispersion parameter for gaussian family taken to be 25.13234)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	204150 on 8123 degrees of freedom

3.1.3 Bayesian Information Criteria

```
Reordering variables and trying again:
```

```
[1] 9
```

	Estimate	Std. Error	t value	Pr(>|t|)
BLUEBOOK	-6.701e-05	7.448e-06	-8.997	2.835e-19
(Intercept)	-0.1542	0.1327	-1.161	0.2455

(Dispersion parameter for gaussian family taken to be 32.08896)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261814 on 8159 degrees of freedom

	Estimate	Std. Error	t value	Pr(>|t|)
log(BLUEBOOK)	-0.9649	0.09934	-9.713	3.485e-22
(Intercept)	7.955	0.9453	8.415	4.588e-17

(Dispersion parameter for gaussian family taken to be 32.03687)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261389 on 8159 degrees of freedom

3.1.4 Mallow's C_p

```
Reordering variables and trying again:
```

```
[1] 9
```

	Estimate	Std. Error	t value	Pr(>|t|)
BLUEBOOK	-6.701e-05	7.448e-06	-8.997	2.835e-19
(Intercept)	-0.1542	0.1327	-1.161	0.2455

(Dispersion parameter for gaussian family taken to be 32.08896)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261814 on 8159 degrees of freedom

	Estimate	Std. Error	t value	Pr(>|t|)
log(BLUEBOOK)	-0.9649	0.09934	-9.713	3.485e-22
(Intercept)	7.955	0.9453	8.415	4.588e-17

(Dispersion parameter for gaussian family taken to be 32.03687)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261389 on 8159 degrees of freedom

3.2 Binary Logistic Regression

3.2.1 Full Model

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.3863	0.06121	6.311	2.764e-10

	Estimate	Std. Error	z value	Pr(>|z|)
AGE	-0.001007	0.004019	-0.2506	0.8021
HOMEKIDS	0.04958	0.03712	1.336	0.1817
YOJ	-0.01106	0.008582	-1.289	0.1975
INCOME	-3.423e-06	1.082e-06	-3.165	0.001553
PARENT1	0.3822	0.1096	3.488	0.0004873
HOME_VAL	-1.305e-06	3.419e-07	-3.818	0.0001346
MSTATUS	-0.4937	0.08357	-5.908	3.465e-09
SEX	0.07685	0.1002	0.7674	0.4429
TRAVTIME	0.01457	0.001883	7.736	1.029e-14
CAR_USE	0.7564	0.09172	8.248	1.615e-16
BLUEBOOK	-2.082e-05	5.261e-06	-3.958	7.567e-05
TIF	-0.05546	0.007344	-7.552	4.277e-14
OLDCLAIM	-1.39e-05	3.91e-06	-3.554	0.0003788
CLM_FREQ	0.1959	0.02854	6.863	6.72e-12
REVOKED	0.8874	0.09133	9.717	2.555e-22
MVR_PTS	0.1133	0.01361	8.323	8.562e-17
CAR_AGE	-0.0007267	0.007548	-0.09627	0.9233
URBANICITY	2.39	0.1128	21.18	1.432e-99
HSDropout	0.1675	0.214	0.7831	0.4336
HS	0.1852	0.1962	0.9438	0.3453
Bachelors	-0.2138	0.1796	-1.19	0.2339
Masters	-0.1227	0.1527	-0.8036	0.4217
Minivan	-0.7687	0.1112	-6.913	4.734e-12
Panel_Truck	-0.2081	0.2002	-1.04	0.2985
Pickup	-0.2147	0.1165	-1.844	0.0652
Sports_Car	0.2568	0.09813	2.616	0.008885
Van	-0.1501	0.1588	-0.9451	0.3446
Blank_Job	-0.3106	0.1856	-1.674	0.09415
Professional	-0.1488	0.1198	-1.243	0.214
Clerical	0.1001	0.1067	0.9379	0.3483
Doctor	-0.7567	0.2877	-2.63	0.008529
Lawyer	-0.2057	0.1875	-1.097	0.2725
Manager	-0.8681	0.1396	-6.217	5.068e-10
Home_Maker	-0.07818	0.1526	-0.5125	0.6083
Student	-0.09478	0.1289	-0.7356	0.462
(Intercept)	-2.753	0.3458	-7.959	1.728e-15

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7298 on 8124 degrees of freedom

Quitting from lines 345-355 (HW4_Master.Rmd) Error in vif.default(glmflagfull) : there are aliased coefficients in the model Calls: ... withVisible -> eval -> eval -> pander -> vif -> vif.default In addition: Warning messages: 1: Removed 1879 rows containing non-finite values (stat_boxplot). 2: Removed 909 rows containing non-finite values (stat_density). 3: Removed 1879 rows containing non-finite values (stat_boxplot). 4: In leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in = force.in, : 3 linear dependencies found 5: In leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in = force.in, : 3 linear dependencies found

	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	314.9	113.1	2.783	0.005395
AGE	5.265	7.064	0.7454	0.456
HOMEKIDS	77.49	65.35	1.186	0.2357
YOJ	-3.987	15.08	-0.2643	0.7915
INCOME	-0.004414	0.001805	-2.445	0.0145
PARENT1	576.7	202	2.855	0.004317
HOME_VAL	-0.0005531	0.0005906	-0.9365	0.349
MSTATUS	-570	144.8	-3.935	8.385e-05
SEX	340.5	162.1	2.1	0.03575
TRAVTIME	11.94	3.222	3.706	0.0002121
CAR_USE	779.4	164.4	4.741	2.163e-06
BLUEBOOK	0.01441	0.008619	1.671	0.09467
TIF	-48.18	12.18	-3.956	7.681e-05
OLDCLAIM	-0.01057	0.007435	-1.422	0.1551
CLM_FREQ	141.5	55.02	2.572	0.01012
REVOKED	549.6	173.5	3.168	0.001542
MVR_PTS	175.3	25.92	6.764	1.437e-11
CAR_AGE	-26.94	12.79	-2.105	0.0353
URBANICITY	1664	139.4	11.94	1.354e-32
HSDropout	-287.2	356	-0.8067	0.4199
HS	-376	322.4	-1.166	0.2435
Bachelors	-546.2	288.8	-1.891	0.05862
Masters	-262.9	254	-1.035	0.3007
Minivan	-753.6	179.2	-4.205	2.637e-05
Panel_Truck	-489.2	340.5	-1.437	0.1508
Pickup	-378.2	199	-1.9	0.05742
Sports_Car	269.2	178.9	1.505	0.1324
Van	-239.1	267	-0.8957	0.3704
Blank_Job	-507.6	321.8	-1.577	0.1148
Professional	-51.82	213	-0.2433	0.8078

	Estimate	Std. Error	t value	Pr(>|t|)
Clerical	21.64	192.6	0.1124	0.9105
Doctor	-1009	435.2	-2.319	0.02042
Lawyer	-277	310.9	-0.891	0.3729
Manager	-987.4	234.3	-4.214	2.531e-05
Home_Maker	-155.1	268.5	-0.5776	0.5636
Student	-221.4	233.4	-0.9488	0.3428
(Intercept)	367.5	571.3	0.6434	0.52

Table 21: Fitting linear model: TARGET_AMT ~ . - RED_CAR - Blue_Collar

Observations	Residual Std. Error	R^2	Adjusted R^2
8161	4544	0.07103	0.06691

3.2.2 Reduced Model

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.4204	0.05507	7.634	2.268e-14
INCOME	-3.896e-06	9.882e-07	-3.942	8.077e-05
PARENT1	0.4721	0.09399	5.023	5.094e-07
HOME_VAL	-1.322e-06	3.307e-07	-3.997	6.423e-05
MSTATUS	-0.4652	0.07877	-5.906	3.499e-09
TRAVTIME	0.01453	0.001878	7.736	1.025e-14
CAR_USE	0.7862	0.07541	10.43	1.883e-25
BLUEBOOK	-2.308e-05	4.701e-06	-4.91	9.095e-07
TIF	-0.0548	0.00733	-7.477	7.619e-14
OLDCLAIM	-1.384e-05	3.904e-06	-3.546	0.0003919
CLM_FREQ	0.1954	0.02847	6.864	6.687e-12
REVOKED	0.8923	0.09113	9.792	1.219e-22
MVR_PTS	0.1141	0.01356	8.416	3.898e-17
URBANICITY	2.374	0.1124	21.12	4.983e-99
HS	-0.02095	0.09218	-0.2273	0.8202
Bachelors	-0.4711	0.09922	-4.748	2.052e-06
Masters	-0.5032	0.1144	-4.399	1.086e-05
PhD	-0.5329	0.1542	-3.456	0.0005476
Minivan	-0.7176	0.08522	-8.42	3.761e-17
Panel_Truck	-0.1171	0.1493	-0.7846	0.4327
Pickup	-0.1728	0.09178	-1.883	0.05975
Sports_Car	0.2563	0.09772	2.622	0.008735
Van	-0.08019	0.119	-0.6741	0.5003
Blank_Job	-0.06311	0.1436	-0.4396	0.6602
Manager	-0.7437	0.1082	-6.875	6.198e-12
(Intercept)	-2.666	0.1688	-15.8	3.083e-56

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7313 on 8135 degrees of freedom

	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	383.4	102	3.76	0.000171
INCOME	-0.005324	0.001344	-3.961	7.518e-05
PARENT1	663	176.1	3.764	0.0001683
MSTATUS	-588.3	119.1	-4.939	8.02e-07
SEX	322.3	148.2	2.175	0.02968
TRAVTIME	12.02	3.216	3.737	0.0001874
CAR_USE	759.5	127.7	5.947	2.847e-09
BLUEBOOK	0.01334	0.007256	1.839	0.06595
TIF	-46.9	12.16	-3.858	0.0001151
CLM_FREQ	107.1	48.78	2.196	0.02815
REVOKED	451.9	154.8	2.92	0.003508
MVR_PTS	173.8	25.77	6.744	1.645e-11
CAR_AGE	-31.48	10.67	-2.95	0.003191
URBANICITY	1635	137.3	11.91	1.909e-32
HS	55.37	126.8	0.4365	0.6625
Minivan	-424.4	148	-2.867	0.00415
Sports_Car	590.8	226.2	2.612	0.009009
SUV	321.6	190.4	1.69	0.09116
Blank_Job	-174.7	237.1	-0.7366	0.4614
Manager	-856.5	163.8	-5.228	1.753e-07
(Intercept)	-181.5	308.5	-0.5883	0.5563

Table 25: Fitting linear model: TARGET_AMT ~ KIDSDRIV + INCOME + PARENT1 + MSTATUS + SEX + TRAVTIME + CAR_USE + BLUEBOOK + TIF + CLM_FREQ + REVOKED + MVR_PTS + CAR_AGE + URBANICITY + HS + Minivan + Sports_Car + SUV + Blank_Job + Manager

Observations	Residual Std. Error	R^2	Adjusted R^2
8161	4544	0.06932	0.06703

3.2.3 Best Subsets

`Reordering variables and trying again:`

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.4066	0.05428	7.491	6.828e-14
INCOME	-5.075e-06	9.47e-07	-5.359	8.389e-08
PARENT1	0.4864	0.09237	5.266	1.396e-07
HOME_VAL	-1.339e-06	3.375e-07	-3.968	7.238e-05
MSTATUS	-0.4382	0.07845	-5.586	2.323e-08
TRAVTIME	0.01479	0.001851	7.988	1.377e-15
CAR_USE	0.8208	0.06639	12.36	4.07e-35
BLUEBOOK	-3.212e-05	4.623e-06	-6.948	3.703e-12
TIF	-0.05361	0.007238	-7.406	1.301e-13
OLDCLAIM	-1.411e-05	3.848e-06	-3.668	0.0002447
CLM_FREQ	0.2015	0.02814	7.161	7.983e-13
REVOKED	0.9076	0.0898	10.11	5.142e-24
MVR_PTS	0.1217	0.01344	9.06	1.303e-19
URBANICITY	2.273	0.1118	20.33	7.698e-92
HSDropout	0.5516	0.09114	6.052	1.426e-09

	Estimate	Std. Error	z value	Pr(>|z|)
HS	0.5037	0.07234	6.963	3.325e-12
Panel_Truck	0.189	0.1314	1.438	0.1504
Van	0.1749	0.1062	1.647	0.09959
Student	-0.05772	0.1072	-0.5382	0.5904
(Intercept)	-3.234	0.1699	-19.04	8.633e-81

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7471 on 8141 degrees of freedom

3.2.4 Asher's Models

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.3863	0.06121	6.311	2.764e-10
AGE	-0.001007	0.004019	-0.2506	0.8021
HOMEKIDS	0.04958	0.03712	1.336	0.1817
YOJ	-0.01106	0.008582	-1.289	0.1975
INCOME	-3.423e-06	1.082e-06	-3.165	0.001553
PARENT1	0.3822	0.1096	3.488	0.0004873
HOME_VAL	-1.305e-06	3.419e-07	-3.818	0.0001346
MSTATUS	-0.4937	0.08357	-5.908	3.465e-09
SEX	0.07685	0.1002	0.7674	0.4429
TRAVTIME	0.01457	0.001883	7.736	1.029e-14
CAR_USE	0.7564	0.09172	8.248	1.615e-16
BLUEBOOK	-2.082e-05	5.261e-06	-3.958	7.567e-05
TIF	-0.05546	0.007344	-7.552	4.277e-14

	Estimate	Std. Error	z value	Pr(>|z|)
OLDCLAIM	-1.39e-05	3.91e-06	-3.554	0.0003788
CLM_FREQ	0.1959	0.02854	6.863	6.72e-12
REVOKED	0.8874	0.09133	9.717	2.555e-22
MVR_PTS	0.1133	0.01361	8.323	8.562e-17
CAR_AGE	-0.0007267	0.007548	-0.09627	0.9233
URBANICITY	2.39	0.1128	21.18	1.432e-99
HSDropout	0.1675	0.214	0.7831	0.4336
HS	0.1852	0.1962	0.9438	0.3453
Bachelors	-0.2138	0.1796	-1.19	0.2339
Masters	-0.1227	0.1527	-0.8036	0.4217
Minivan	-0.7687	0.1112	-6.913	4.734e-12
Panel_Truck	-0.2081	0.2002	-1.04	0.2985
Pickup	-0.2147	0.1165	-1.844	0.0652
Sports_Car	0.2568	0.09813	2.616	0.008885
Van	-0.1501	0.1588	-0.9451	0.3446
Blank_Job	-0.3106	0.1856	-1.674	0.09415
Professional	-0.1488	0.1198	-1.243	0.214
Clerical	0.1001	0.1067	0.9379	0.3483
Doctor	-0.7567	0.2877	-2.63	0.008529
Lawyer	-0.2057	0.1875	-1.097	0.2725
Manager	-0.8681	0.1396	-6.217	5.068e-10
Home_Maker	-0.07818	0.1526	-0.5125	0.6083
Student	-0.09478	0.1289	-0.7356	0.462
(Intercept)	-2.753	0.3458	-7.959	1.728e-15

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7298 on 8124 degrees of freedom

Quitting from lines 406-415 (HW4_Master.Rmd) Error in vif.default(glmflag) : there are aliased coefficients in the model Calls: … withVisible -> eval -> eval -> pander -> vif -> vif.default In addition: Warning messages: 1: In predict.lm(object, newdata, se.fit, scale = 1, type = ifelse(type == : prediction from a rank-deficient fit may be misleading 2: In predict.lm(object, newdata, se.fit, scale = 1, type = ifelse(type == : prediction from a rank-deficient fit may be misleading 3: In predict.lm(object, newdata, se.fit, scale = 1, type = ifelse(type == : prediction from a rank-deficient fit may be misleading 4: In predict.lm(object, newdata, se.fit, scale = 1, type = ifelse(type == : prediction from a rank-deficient fit may be misleading 5: In leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in = force.in, : 2 linear dependencies found ——————————————————————————————— Estimate Std. Error z value Pr(>|z|) ——————— ———- ———— ——— ———- **KIDSDRIV** 0.4204 0.05507 7.634 2.268e-14

INCOME -3.896e-06 9.882e-07 -3.942 8.077e-05

PARENT1 0.4721 0.09399 5.023 5.094e-07

HOME_VAL -1.322e-06 3.307e-07 -3.997 6.423e-05

MSTATUS -0.4652 0.07877 -5.906 3.499e-09

TRAVTIME 0.01453 0.001878 7.736 1.025e-14

CAR_USE 0.7862 0.07541 10.43 1.883e-25

BLUEBOOK -2.308e-05 4.701e-06 -4.91 9.095e-07

```
 **TIF**        -0.0548      0.00733     -7.477   7.619e-14
```

OLDCLAIM -1.384e-05 3.904e-06 -3.546 0.0003919

CLM_FREQ 0.1954 0.02847 6.864 6.687e-12

REVOKED 0.8923 0.09113 9.792 1.219e-22

MVR_PTS 0.1141 0.01356 8.416 3.898e-17

URBANICITY 2.374 0.1124 21.12 4.983e-99

```
 **HS**         -0.02095     0.09218     -0.2273    0.8202
```

Bachelors -0.4711 0.09922 -4.748 2.052e-06

Masters -0.5032 0.1144 -4.399 1.086e-05

```
 **PhD**        -0.5329      0.1542     -3.456   0.0005476
```

Minivan -0.7176 0.08522 -8.42 3.761e-17

Panel_Truck -0.1171 0.1493 -0.7846 0.4327

Pickup -0.1728 0.09178 -1.883 0.05975

Sports_Car 0.2563 0.09772 2.622 0.008735

```
 **Van**        -0.08019     0.119      -0.6741    0.5003
```

Blank_Job -0.06311 0.1436 -0.4396 0.6602

Manager -0.7437 0.1082 -6.875 6.198e-12

3.3 (Intercept) -2.666 0.1688 -15.8 3.083e-56

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7313 on 8135 degrees of freedom

```
Call:
roc.formula(formula = class ~ logit, data = predflag, auc = TRUE,     plot = TRUE, smooth = TRUE)

Data: logit in 6008 controls (class 0) < 2153 cases (class 1).
Smoothing: binormal
Area under the curve: 0.8084
```

4 Model Selection & Prediction

Appendix A: Index-wise Results from Predictive Model

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Appendix B: R Code