2020-03-31

An Actuarial Approach to Personal Injury Protection Severity

Jason Colgrove

Honors Thesis

AN ACTUARIAL APPROACH TO PERSONAL INJURY PROTECTION SEVERITY

By Jason Colgrove

Submitted to Brigham Young University in partial fulfillment

of graduation requirements for University Honors

Statistics Department

Brigham Young University

April 2020

Advisor: Brian Hartman

Honors Coordinator: Del Scott

ABSTRACT

AN ACTUARIAL APPROACH TO PERSONAL INJURY PROTECTION SEVERITY

Jason Colgrove

Statistics Department

Bachelor of Science

Insurance companies examine the risk of financial losses for their policyholders as a way to accurately price insurance policies. Within the automobile insurance sector, the frequency of crashes and the associated liabilities started to increase in late 2013 when it had been on the decline for close to a decade. The purpose of this research focuses on the possible correlated variables that could lead to a better understanding of this change.

To embark on this task, we teamed up with the Society of Actuaries, Casualty Actuarial Society, and the American Property Casualty Insurance Association to obtain data regarding frequency, severity, and loss costs. They have available resources in the insurance sector to inform other people about our findings.

The method for this project primarily focuses on using a random forest model and its associated variable importance plot. This method allowed us to determine which variables are most important for auto coverage.

Our team looked at the coverages of bodily injury, collision, comprehensive, personal injury protection, and property damage. This thesis will focus exclusively on

personal injury protection, especially regarding severity, since that was my main contribution to the project.

TABLE OF CONTENTS

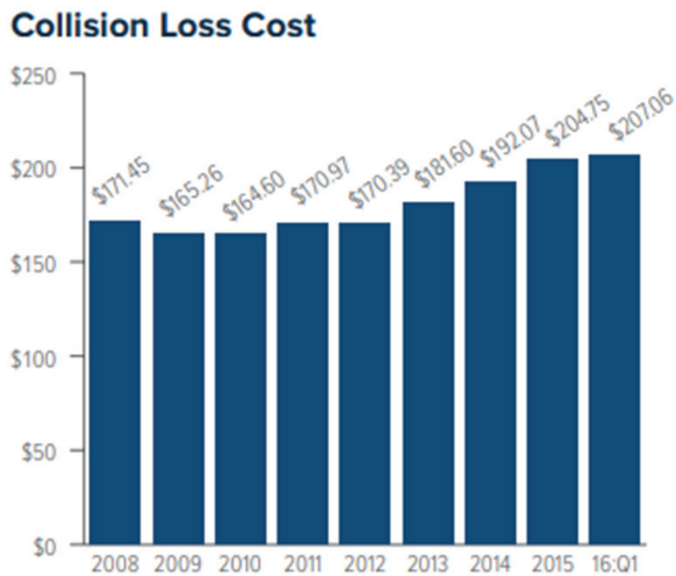
LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Automobile insurance companies rely on data to predict which factors pose the greatest risk for financial loss when using an automobile. Younger and older drivers present a greater risk for auto losses due to inexperience and aging, respectively. Insurance companies track frequency and severity very closely to have the best prediction methods. When aggregated car accident frequencies and costs started to rise in the latter part of 2013, a closer look at these environmental factors was needed to determine this change (Hartwig, 2016). The Insurance Information Institute recorded these changes and expressed them in a graph shown in Figure 1 (Hartwig, 2016). From the years 2008 to 2012, the aggregated loss costs per policy was fairly stable. However, starting in 2013, we can see that there is a steady increase in each sequential year.

Figure 1



Note that the findings from this research do not determine causality; the data collected was not from an experiment, it was observational. Neither is the intent of this research to be used as the basis for pricing insurance. Instead, its purpose is to give an understanding of factors that could likely influence auto losses and it shows an actuarial approach to this problem.

The data was collected from a variety of trusted and reliable public sources. The main sources were the Federal Highway Administration, Department of Transportation, Bureau of Economic Analysis, and the Census Bureau. These organizations had data such as population, square miles of land, amounts of each road type, congestion levels, weather conditions, number of licensed drivers, etc. We used all the data that we could find that had any relevance. The data was then entered into an Excel file where it was organized by state, year, and quarter; this way we could easily isolate any outliers or trends to a specific location or time.

Some of the data collected needed to be cleaned and organized. For example, some of the data had values for the whole year instead of values for each of the four quarters of the year. In these instances, we kept the same values for the whole year. Also, some data was missing or incomplete. In these cases, we either did not include that variable or we used interpolation or extrapolation to estimate the values of the missing data. We recognize that these manipulations slightly modified our data; however, these changes were necessary for the integrity of the data and a preliminary step before any analysis.

Once the data was collected and cleaned, our team was prepared to work on specific sectors of the auto insurance industry. From this point on, I will be discussing the work I did for the personal injury protection coverage. Personal injury protection coverage, also known as PIP, is an auto coverage that covers the medical expenses after an auto accident. PIP coverage is available on a state by state basis. Most states have no-fault laws in place which means that regardless of whose fault it was, PIP covers the costs. Some states have a monetary threshold to sue instead of filing a PIP claim, and other states have a verbal threshold that only includes certain types of injuries. In states where PIP is not available, the individuals harmed may be covered under their medical insurance or may sue the driver at-fault.

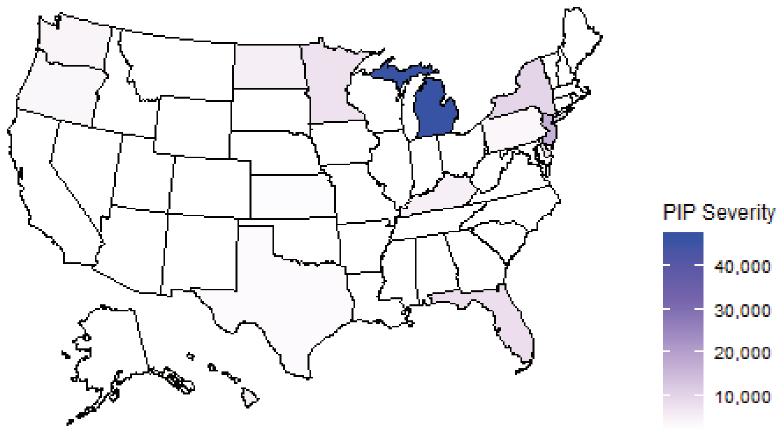
The data contains the response variables of frequency, severity, and loss costs while the explanatory variables were the various factors we collected from the publicly available data. Some of the more important explanatory variables are listed as follows:

- Percentage of licensed male drivers younger than 25
- Percentage of licensed male drivers older than 75
- Percentage of licensed male drivers between 25 and 75, inclusive
- Percentage of licensed female drivers younger than 25
- Percentage of licensed female drivers older than 75
- Percentage of licensed female drivers between 25 and 75, inclusive
- Percentage of road miles on rural roads
- Percentage of road miles on urban roads
- Vehicle miles traveled divided by the land area in miles squared for each state
- Percentage of vehicle miles traveled based on interstate
- Percentage of vehicle miles traveled on other freeways
- Percentage of vehicle miles traveled on principal arterial roads
- Percentage of vehicle miles on minor arterial roads
- Percentage of vehicle miles traveled on major collector roads
- Percentage of vehicle miles traveled on minor collector roads

- Percentage of vehicle miles traveled on local roads
- Total industry GDP per capita
- Vehicle miles traveled over total road miles
- Urban vehicle miles traveled over total urban road miles
- Rural vehicle miles traveled over total rural road miles
- Total road miles divided by the total land area

Since PIP only 18 states where the coverage is offered, states that do not offer PIP were deleted from the data. This allowed us to use only the relevant data. Upon

Figure 2
Personal Injury Protection Severity

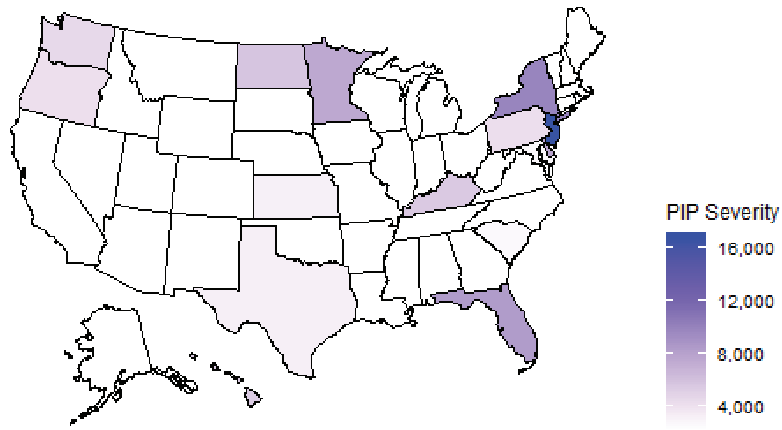


the initial examination of the data, it was discovered that Michigan was a clear outlier; Michigan was only an outlier for severity and not for frequency. Figure 2 shows Michigan when compared to the other states regarding severity.

Michigan is an outlier because PIP "will pay all reasonably necessary medical expenses with no maximum limit", including "up to 85% of the income you would have earned if you had not been hurt, for up to three years" (The State of Michigan, 2019). Because Michigan has no limit for PIP related losses, auto losses involving PIP coverage can potentially be much more expensive. To fix this outlier, we deleted Michigan from our data. Figure 3 shows the severity of PIP by state without Michigan. As seen on the graphs, the top of the scale drops from around 40,000 in Figure 1 to 16,000 in Figure 2—a significant difference. After excluding Michigan, there were no other outliers that needed to be fixed.

Another important part of preparing the data was to use the log of the severity over time instead of the severity because we need to account for inflation. It is important to note that this was not simply the log of severity because it has to account for time instead of minimizing the severity. Loss costs were then recalculated from the new severity values; which again, accounted for time. These values were calculated by fitting the log of the data to a linear model with time as the explanatory variable. The standard residuals of the linear model were then calculated; once the residuals were calculated, inflation did not have any effect over time. Note that from here on out, the severity will refer to these new calculated values.

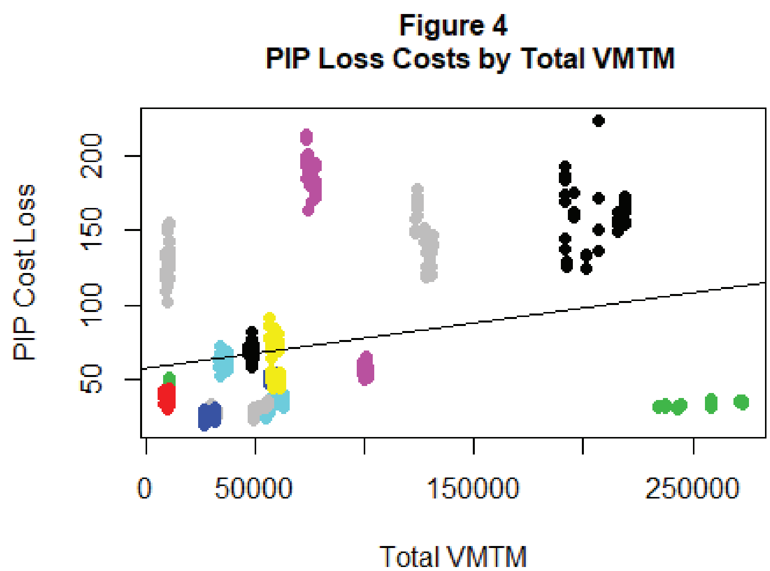
Figure 3
Personal Injury Protection Severity (without Michigan)



After these initial steps, the data needed to be fit to a model to find the most important variables. I used a random forest model to fit the data and find the most valuable variables. However, this was not my initial approach.

My first approach was to use a regression model for the analysis. I decided to fit a linear model for the top 5 variables against the loss costs using AIC— this way the model was not overfit. From this analysis, I found the top 5 variables: the state, licensed female drivers between the ages of 20 and 24, licensed female drivers between the ages of 80 and 84, vehicle miles traveled total, and the quarter (three-month period). To verify my analysis, I ran a simple linear regression with each of the variables against loss costs. I found that although the state was an important variable, it did not give us any useful information because of the differences in state laws regarding PIP— it merely told us those differences. Furthermore, since the state is a nonbinary categorical variable, each state would be treated as a different variable. This creates a problem because it would add 17 more variables to the regression model. A regression model is hard to be accurate when there are a lot of variables because it can easily become overfit. Also, it was found that the other variables, though linear overall, were mostly clumped in the lower parts of the graph.

Figure 4 illustrates this problem using the example of total vehicle miles traveled in miles against the loss costs. The different colors and groupings represent the different states. Some of the colors were repeated; if the colors are far apart like some of the grays and purples, then they are from different states. Each of the points represents the 17 states for each quarter of the year between 2010 and 2018; this makes for a total of 612 points on the graph. The problem with Figure 4 is that though it appears linear overall, the data is very grouped around 50 for loss costs and 50,000 for VMTM.



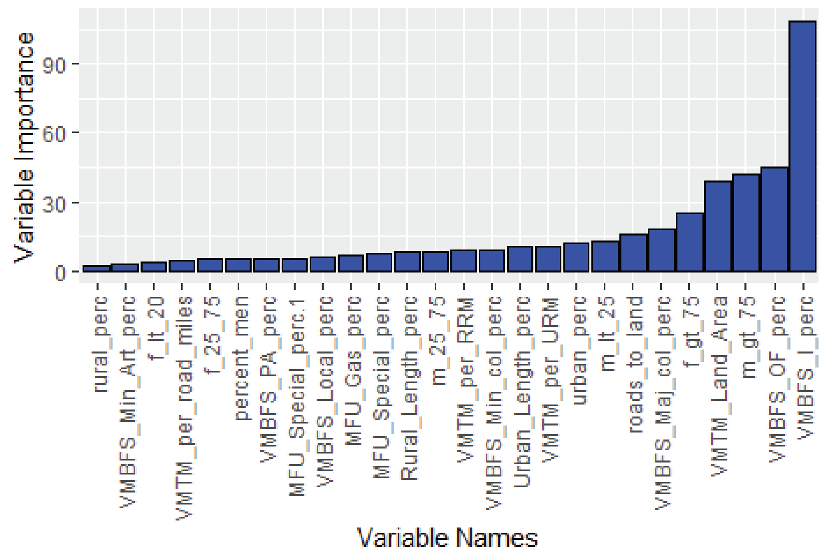
Outside that scope, the regression line fails to estimate those values accurately. Thus, this linear regression does not accurately reflect the data.

Furthermore, there are only 18 states that have PIP. Michigan was already excluded and throwing out a few more states would not be a viable option because there would be little data to work with. The more I analyzed this model, the more I realized that linear regression was not the best method. It assumes that there exists a linear relationship between the data, which was not always the case. Furthermore, there were over 100 variables which made collinearity a big problem as well, especially related to age. Hence, I decided to forgo this model in pursuit of a better model.

After I attempted using a linear model, I decided to use a random forest model. This is the model that I ended up using for my analysis. This model accounts for some of the non-linear relationships, the standard residuals of the frequency, severity, and loss costs were used.

For the random forest model, the data were divided into a training dataset and a test dataset as a way to improve its accuracy and to verify the results. The random forest model I used had bootstrap samples, grew 50 trees, used 7 variables for each tree, and had a node size of 25.

Figure 5
Severity Variable Importance Plot



The results of the random forest model were fairly consistent with what was expected. Figure 5 shows the variable importance plot for severity. As shown in Figure 5, the percent of interstate roads is the most important factor in determining PIP severity. The next biggest factors were the percent of other freeways, the percent of men greater than 75 years old, the percent of vehicle miles traveled in millions divided by the land area, and the percent of females greater than 75 years old.

Figure 6
Added-Variable Plot: Interstate Percantage

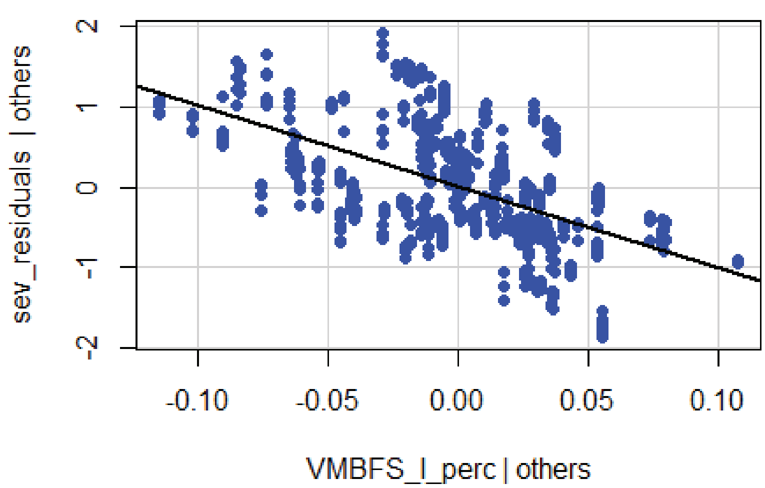


Figure 6 shows an added-variable plot between the interstate percentage and the severity. There is a clear negative correlation which is calculated to be -0.72.

This indicates that as the percentage of interstates increases, the severity of PIP decreases.

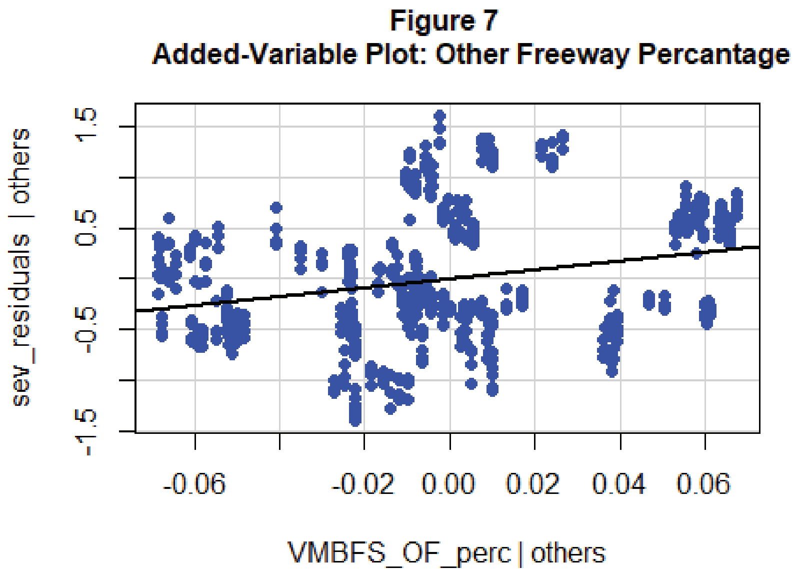


Figure 7 shows the positive correlation between the percent of other freeways and the severity residuals. This shows that as the percent of other freeways decreases, so does the severity. Interestingly, the interstate is positively correlated when the other freeways are negatively correlated. The correlation between the interstate percentage variable and the other freeways variable is -0.19. Based on this information, it appears that the interstate percentage variable and the other freeways percentage variable are not strongly correlated. Hence, other freeways is another important variable at play.

CONCLUSION

The conclusion of the results indicates that road type is a key factor for PIP severity. The most important two variables were interstate roads and other freeways. The negative correlation of the interstate percentage variable and severity suggests that interstates have a smaller chance of having severe accidents that involve PIP coverage. However, the data is not conclusive and does not show causality; this is not suggesting that you drive on the interstate more because the data suggests that it is safer. Furthermore, drivers over the age of 75 years of age pose a greater risk for PIP severity.

For convenience, I fit the top 5 variables in a linear model. Note that this linear model is much better than the one mentioned earlier (the one I attempted before using a random forest model) because these variables are more accurate. The parameters are given in Table 1.

Parameter	Table 1: Severity Linear Model			
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	135.65	1024.41	1.109	0.268
VMBFS Interstate Percentage	-17,623.58	192541	-9.153	< 0.0001
VMBFS Other Freeways Percentage	14786.13	2704.64	5.467	< 0.0001
Percentage of Men Older Than 75	97488.36	22955.05	4.247	< 0.0001
VMBFS Land Area	412.29	45.56	9.048	< 0.0001
Percentage of Women Older Than 75	-1781.03	20555.44	-0.087	0.931

In future studies, it would be recommended to account for the differences in each state. Methods like spatial statistics or a linear model by state, like the linear model used to account for inflation, could be a solid option to give more accuracy to the model.

WORKS CITED

Hartwig, P. R., Lynch, J., & Weisbart, S. (2016, October). More Accidents, Larger Claims Drive Costs Higher. Insurance Information Institute. Retrieved from Insurance Information Institute: https://www.iii.org/sites/default/files/docs/pdf/auto_rates_wp_092716-62.pdf

The State of Michigan (2019, October). *The State of Michigan.* Retrieved from Brief Explanation of Michigan No-Fault Insurance: https://www.michigan.gov/documents/cis_ofis_ip202_25083_7.pdf

APPENDIX

Table 2 contains the original data that was collected. The table has been transposed to show the data more clearly. The rows represent the different variables and the columns represent the different observations.

Table 3 contains the data after it was manipulated and changed. The table has been transposed to show the data more clearly. It only contains the data regarding PIP severity. This is the data that was used in the analysis.

Table 2
Transposed Original Data: First 5 Observations

	1	2	3	4	5
ID	253	254	255	256	257
State.Name	Delaware	Delaware	Delaware	Delaware	Delaware
State.Abb	DE	DE	DE	DE	DE
Year	2010	2010	2010	2010	2011
Quarter	1	2	3	4	1
BI_Frequency	0.010656	0.010365	0.01053	0.011661	0.011766
BI_LossCost	171.1009	182.0824	179.7748	184.8688	196.5725
BI_Severity	16056.12	17566.35	17072.98	15853.22	16706.34
Coll_Frequency	0.076804	0.063678	0.063113	0.06157	0.07084
Coll_LossCost	201.8231	169.985	170.2478	176.3348	194.9278
Coll_Severity	2627.785	2669.46	2697.488	2863.984	2751.674
Comp_Frequency	0.050886	0.057621	0.055752	0.048755	0.049027
Comp_LossCost	49.69741	52.1136	55.6736	67.09086	47.98465
Comp_Severity	976.6437	904.4185	998.5871	1376.088	978.7312

PD_Frequency	0.039668	0.036864	0.039072	0.038513	0.040426
PD_LossCost	111.3561	107.1765	115.6939	110.6864	119.2226
PD_Severity	2807.201	2907.388	2961.03	2874.034	2949.146
PIP_Frequency	0.018329	0.015627	0.015339	0.016612	0.019561
PIP_LossCost	132.7397	123.3819	117.2347	119.376	135.4333
PIP_Severity	7242.178	7895.613	7642.947	7186.173	6923.523
L_M_D_Age_20	16225	16225	16225	16225	17244
L_M_D_Age20_24	29011	29011	29011	29011	29794
L_M_D_Age25_29	30474	30474	30474	30474	31847
L_M_D_Age30_34	28703	28703	28703	28703	30137
L_M_D_Age35_39	27511	27511	27511	27511	27791
L_M_D_Age40_44	30211	30211	30211	30211	30956
L_M_D_Age45_49	32780	32780	32780	32780	33243
L_M_D_Age50_54	31720	31720	31720	31720	32442
L_M_D_Age55_59	27617	27617	27617	27617	28459
L_M_D_Age60_64	25643	25643	25643	25643	26544
L_M_D_Age65_69	20180	20180	20180	20180	20682
L_M_D_Age70_74	14782	14782	14782	14782	15409
L_M_D_Age75_79	11038	11038	11038	11038	11391
L_M_D_Age80_84	7805	7805	7805	7805	8204
L_M_D_Age85__	5323	5323	5323	5323	5836
L_M_D_TOTAL	339023	339023	339023	339023	349979
L_F_D_Age_20	17158	17158	17158	17158	18255
L_F_D_Age20_24	28997	28997	28997	28997	29597
L_F_D_Age25_29	31142	31142	31142	31142	32146
L_F_D_Age30_34	29657	29657	29657	29657	31090
L_F_D_Age35_39	28610	28610	28610	28610	28816
L_F_D_Age40_44	31577	31577	31577	31577	32046
L_F_D_Age45_49	34000	34000	34000	34000	34440
L_F_D_Age50_54	33607	33607	33607	33607	34321
L_F_D_Age55_59	30191	30191	30191	30191	30824
L_F_D_Age60_64	27927	27927	27927	27927	28887
L_F_D_Age65_69	21672	21672	21672	21672	22292
L_F_D_Age70_74	15634	15634	15634	15634	16239
L_F_D_Age75_79	11664	11664	11664	11664	12004

L_F_D_Age80_84	8208	8208	8208	8208	8688
L_F_D_Age85__	5969	5969	5969	5969	6485
L_F_D_TOTAL	356013	356013	356013	356013	366130
RRM_County	0	0	0	0	0
RRM_Town__Township__Municipal	146.58	146.58	146.58	146.58	146.58
RRM_Other_Jurisdiction	46.35	46.35	46.35	46.35	46.35
RRM_Federal_Agency	71.91	71.91	71.91	71.91	71.91
RRM_Total	3360.85	3360.85	3360.85	3360.85	3360.85
URM_State_Highway_Agency	2278.61	2278.61	2278.61	2278.61	2278.61
URM_County	0	0	0	0	0
URM_Town__Township__Municipal	647.07	647.07	647.07	647.07	647.07
URM_Other_Jurisdiction	20.9	20.9	20.9	20.9	20.9
URM_Federal_Agency	50.16	50.16	50.16	50.16	50.16
URM_Total	2996.74	2996.74	2996.74	2996.74	2996.74
Total_Road_Miles	6357.59	6357.59	6357.59	6357.59	6357.59
Medicinal	No	No	No	No	No
Decriminalized	No	No	No	No	No
Legal_Status	Fully Illegal	Fully Illegal	Fully Illegal	Fully Illegal	Fully Illegal
VMTM_Rural	2779	2779	2779	2779	2902
VMTM_Urban	6169	6169	6169	6169	6127
VMTM_Total	8948	8948	8948	8948	9029
Trfc_Rural_Interstate	0	0	0	0	0
Trfc_Rural_Freeway	0	0	0	0	0
Trfc_Rural_Other_Arterial	6642.126	6642.126	6642.126	6642.126	6642.126
Trfc_Rural_All_Arterials	6642.126	6642.126	6642.126	6642.126	6642.126
Trfc_Urban_Interstate	13647.19	13647.19	13647.19	13647.19	13647.19
Trfc_Urban_Freeway	10063.21	10063.21	10063.21	10063.21	10063.21
Trfc_Urban_Other_Arterial	7564.312	7564.312	7564.312	7564.312	7564.312
Trfc_Urban_All_Arterials	8875.171	8875.171	8875.171	8875.171	8875.171
Trfc_All_Arterials	7993.296	7993.296	7993.296	7993.296	7993.296
MFU_Gasoline_Gasohol	454546	454546	454546	454546	438678
MFU_Special_Fuel	61039	61039	61039	61039	61119

MFU_Total_Fuel	515585	515585	515585	515585	499797
Personal_Income	35602612	36312996	37195824	37717280	39347924
Total_Land_Area	1954	1954	1954	1954	1954
All_Industry_Total_GDP	57422.7	56716.6	57406.1	58244.4	59077
Rural_Interstate_Speed_Limt__mph	65	65	65	65	65
Urban_Interstate_Speed_Limit__mp	55	55	55	55	55
Other_Limited_Access_Roads_speed	65	65	65	65	65
Other_Roads__mph_	55	55	55	55	55
Public_Road_Length_Rural	3346.33	3346.33	3346.33	3346.33	3361
Public_Road_Length_Urban	2990.87	2990.87	2990.87	2990.87	2997
Public_Road_Length_Total	6337.2	6337.2	6337.2	6337.2	6358
VMBFS_Interstate	1191	1191	1191	1191	1263
VMBFS_OtherFreeways	440	440	440	440	491
VMBFS_Principal_Arterial	3186	3186	3186	3186	3269
VMBFS_Minor_Arterial	1281	1281	1281	1281	1305
VMBFS_Major_Collector	1325	1325	1325	1325	1332
VMBFS_Minor_Collector	106	106	106	106	108
VMBFS_Local	1419	1419	1419	1419	1260
VMBFS_Total	8948	8948	8948	8948	9028
Year_and_Quarter	2010	2010.25	2010.5	2010.75	2011
m_lt_25	0.13343	0.13343	0.13343	0.13343	0.134402
m_gt_75	0.071281	0.071281	0.071281	0.071281	0.072664
m_25_75	0.795288	0.795288	0.795288	0.795288	0.792933
f_lt_20	0.129644	0.129644	0.129644	0.129644	0.130697
f_gt_75	0.072584	0.072584	0.072584	0.072584	0.074228
f_25_75	0.797771	0.797771	0.797771	0.797771	0.795076
rural_perc	0.528636	0.528636	0.528636	0.528636	0.528636
urban_perc	0.471364	0.471364	0.471364	0.471364	0.471364
MFU_Gas_perc	0.881612	0.881612	0.881612	0.881612	0.877712

MFU_Special_perc	0.118388	0.118388	0.118388	0.118388	0.122288
VMTM_Land_Area	4.579324	4.579324	4.579324	4.579324	4.620778
Rural_Length_perc	0.528046	0.528046	0.528046	0.528046	0.528625
Urban_Length_perc	0.471954	0.471954	0.471954	0.471954	0.471375
VMBFS_I_perc	0.133102	0.133102	0.133102	0.133102	0.139898
VMBFS_OF_perc	0.049173	0.049173	0.049173	0.049173	0.054386
VMBFS_PA_perc	0.356057	0.356057	0.356057	0.356057	0.362096
VMBFS_Min_Art_perc	0.14316	0.14316	0.14316	0.14316	0.14455
VMBFS_Maj_col_perc	0.148078	0.148078	0.148078	0.148078	0.147541
VMBFS_Min_col_perc	0.011846	0.011846	0.011846	0.011846	0.011963
VMBFS_Local_perc	0.158583	0.158583	0.158583	0.158583	0.139566
VMTM_per_road_miles	1.407452	1.407452	1.407452	1.407452	1.420192
VMTM_per_URM	2.05857	2.05857	2.05857	2.05857	2.044555
VMTM_per_RRM	1.209374	1.209374	1.209374	1.209374	1.158115
percent_men	0.487778	0.487778	0.487778	0.487778	0.488723
roads_to_land	3.253628	3.253628	3.253628	3.253628	3.253628

Table 3
Transposed Revised Data: First 5 Observations

	1	2	3	4	5
sev_residuals	0.832783	0.9833	0.918894	0.801961	0.729145
m_lt_25	0.13343	0.13343	0.13343	0.13343	0.134402
m_gt_75	0.071281	0.071281	0.071281	0.071281	0.072664
m_25_75	0.795288	0.795288	0.795288	0.795288	0.792933
f_lt_20	0.129644	0.129644	0.129644	0.129644	0.130697
f_gt_75	0.072584	0.072584	0.072584	0.072584	0.074228
f_25_75	0.797771	0.797771	0.797771	0.797771	0.795076
rural_perc	0.528636	0.528636	0.528636	0.528636	0.528636
urban_perc	0.471364	0.471364	0.471364	0.471364	0.471364
MFU_Gas_perc	0.881612	0.881612	0.881612	0.881612	0.877712
MFU_Special_perc	0.118388	0.118388	0.118388	0.118388	0.122288
MFU_Special_perc.1	0.118388	0.118388	0.118388	0.118388	0.122288
VMTM_Land_Area	4.579324	4.579324	4.579324	4.579324	4.620778
Rural_Length_perc	0.528046	0.528046	0.528046	0.528046	0.528625
Urban_Length_perc	0.471954	0.471954	0.471954	0.471954	0.471375
VMBFS_I_perc	0.133102	0.133102	0.133102	0.133102	0.139898
VMBFS_OF_perc	0.049173	0.049173	0.049173	0.049173	0.054386
VMBFS_PA_perc	0.356057	0.356057	0.356057	0.356057	0.362096
VMBFS_Min_Art_perc	0.14316	0.14316	0.14316	0.14316	0.14455
VMBFS_Maj_col_perc	0.148078	0.148078	0.148078	0.148078	0.147541
VMBFS_Min_col_perc	0.011846	0.011846	0.011846	0.011846	0.011963
VMBFS_Local_perc	0.158583	0.158583	0.158583	0.158583	0.139566
VMTM_per_road_miles	1.407452	1.407452	1.407452	1.407452	1.420192
VMTM_per_URM	2.05857	2.05857	2.05857	2.05857	2.044555
VMTM_per_RRM	1.209374	1.209374	1.209374	1.209374	1.158115
percent_men	0.487778	0.487778	0.487778	0.487778	0.488723
roads_to_land	3.253628	3.253628	3.253628	3.253628	3.253628