Road Traffic Deaths Attributed to Alcohol Use: How Drinking Habits and Laws Contribute

Kimberly Wendt

4/13/15

**Introduction**

 Traffic accidents occur every day and many times involve fatalities. These accidents could be brought on by a variety of factors, such as inattentiveness to the road, being distracted, falling asleep, or texting. Traffic fatalities may also be spatially related with more fatalities occurring in certain parts of the world. This analysis will look specifically at traffic accidents that are attributable to the use of alcohol. This is important to study because alcohol is a high contributor to traffic accident fatalities. If laws can be changed or created that lessen this risk, it can make everyone a little bit safer.

**Conceptualization**

 Drunk driving has been analyzed many times in recent years. A statistical analysis was performed by Schmitz et al (2014) wherein they analyzed driving-under-the-influence (DUI) offenders and used variables such as age, amount of time having a license, what kinds of licenses, restrictions on licenses, and type of vehicle driving at the time of the offense. They found that people aged 41-50 years, having a license for over 12 years, driving a truck, among other variables were all related to recurring DUIs.

 Traffic fatalities related to alcohol use have also been studied extensively. Chile implemented new alcohol policies in recent years that seem to have reduced traffic fatalities. They lowered the legal BAC limit from 0.8 to 0.3 mg and created more severe penalties for exceeding it, used media campaigns, and provided police with portable devices for testing the BAC (Pemjean 2015). A similar analysis was done in Botswana where aggressive policies, education, and enforcement were put in place and the overall crash rate and fatal crash rates both declined (Sebego 2014). Jiang et al (2015) analyzed per capital alcohol consumption and mortality rates in Australia and found a significant association between the two variables. They also found that the compulsory seat belt legislation and random breath testing may have led to a reduction in traffic crash mortality.

 In this study, I will be looking at a different set of variables and running statistical analyses on these variables with the data coming from countries all over the world. I will see if these variables have any effect on traffic fatalities and if traffic fatalities are spatially autocorrelated.

**Data**

My data comes from the World Health Organization (WHO). They have a Global Health Observatory Data Repository which includes their Global Information System on Alcohol and Health where I found my alcohol-related variables. They have data on levels of consumption, patterns of consumption, harms and consequences, economic aspects, alcohol control policies, prevention, research, treatment, youth and alcohol, and key alcohol indicators relevant to noncommunicable diseases. A weakness of this data is that some countries are missing data. However, most of these countries are so small that they would have made my statistical analysis not as accurate anyway. Much of this data represents the best percentage estimates of WHO as well which may not be as telling as the raw data. The data comes from different sources depending on the surveys conducted in each country. All of the data is from recent years (2010-2012) and for either the entire population or for males only.

*Variables*

 Heavy episodic drinking is defined as the percentage of adults over 15 years of age who have had more than six alcoholic drinks on at least one occasion in the past month. Teenage drinking is defined as the proportion of teenagers from 15 to 19 years old who have consumed any amount of alcohol in the past year. The total alcohol per capita consumption of pure alcohol is defined as both the recorded and unrecorded consumption over a year. The recorded consumption was collected by referring to official statistics such as production, import, export, and sales and taxation data. The unrecorded consumption refers to alcohol which is outside of government control and not taxed. The tourist consumption for areas with high tourist levels was determined and subtracted from the recorded consumption. The minimum age variable is the legal age limit for off-premise sales of alcoholic beverages. The legal BAC limit is the maximum blood alcohol concentration allowed for the general population while driving a vehicle. Income groups are defined by the GNI per capita. Low-income economies have a GNI per capita of less than $1,045, lower-middle-income economies have a GNI per capita between $1,045 and $4,125, upper-middle-income economies have a GNI per capita between $4,125 and $12,746, and high-income economies have a GNI per capita of more than $12,746. Region is defined as World Health Organization regions which include African, Americas, South-East Asia, European, Eastern Mediterranean, and Western Pacific.

|  |
| --- |
| Conceptualization of the Variables |
|  | Variable | Definition | Hypothesized Effect (+/-) |
| Dependent | Traffic | The percentage of traffic fatalities attributable to alcohol use |  |
| Independent  | Heavy Drinking | The percentage of the population who participates in heavy drinking at least once a month | + |
| Independent | Teenage Drinking | The percentage of teenagers from 15-19 years old who drink at least once a month | + |
| Independent | Consumption per Capita | The average number of liters of pure alcohol drunk by a person in the past three years | + |
| Independent | Minimum Age | The minimum legal drinking age of the country | - |
| Independent | BAC Limit | The legal Blood Alcohol Content limit for driving a vehicle | + |
| Independent | Income | The income group to which a country belongs | + for High, Upper-middle- for Lower-middle, Low |
| Independent | Region | The region of the world where the country is located | + for Americas, Eastern Mediterranean, Europe- for Africa, South-East Asia, Western Pacific |

**Methods**

After downloading this data from the WHO, I cleaned it up to just show the variables in which I was interested. Income level and region are both nominal variables, so I created dummy variables for both. To analyze this data, I decided to begin by computing an ordinary least squares regression. Then I ran a spatial error model because of my spatial autocorrelation results. Finally, I ran some spatial statistics on my data.

**Analysis**

 The ANOVA tables breaks down variance in the dependent variable into regression (the variance explained by the independent variables or the model) and residual (the variance not explained by the independent variables or the error). The significance value of .000 in the Analysis of Variance test shows that the results are significant and likely not due to random chance, which means that there is a linear relationship between the variables in my model. Total degrees of freedom is calculated by N-1 and the regression degrees of freedom is calculated by the number of coefficients minus 1.

|  |
| --- |
| **ANOVAa** |
| Model | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 12475.769 | 13 | 959.675 | 9.019 | .000b |
| Residual | 9257.586 | 87 | 106.409 |  |  |
| Total | 21733.355 | 100 |  |  |  |
| a. Dependent Variable: Traffic |
| b. Predictors: (Constant), Western Pacific, Lower-middle-income, Consumption, Eastern Mediterranean, BAC limit, Minimum Age, South-East Asia, Upper-middle-income, Africa, Heavy, Low-income, Europe, Teenage |

|  |
| --- |
| **Correlations** |
|  | Traffic | Heavy  | Teenage  | Consumption  | Minimum Age | BAC limit |
| Pearson Correlation | Traffic  | 1.000 | .474 | .111 | .559 | .075 | -.133 |
| Heavy  | .474 | 1.000 | .586 | .607 | -.161 | -.237 |
| Teenage | .111 | .586 | 1.000 | .513 | -.263 | -.325 |
| Consumption | .559 | .607 | .513 | 1.000 | -.118 | -.291 |
| Minimum Age | .075 | -.161 | -.263 | -.118 | 1.000 | .162 |
| BAC limit | -.133 | -.237 | -.325 | -.291 | .162 | 1.000 |
| Low-income | -.238 | -.328 | -.422 | -.441 | -.064 | .183 |
| Lower-middle-income | .147 | -.171 | -.298 | -.057 | .220 | .008 |
| Upper-middle-income | .228 | -.036 | .048 | .109 | .018 | -.188 |
| Africa | -.043 | -.358 | -.612 | -.296 | .082 | .348 |
| Eastern Mediterranean | -.043 | -.127 | -.252 | -.151 | .003 | -.025 |
| Europe | .116 | .567 | .745 | .486 | -.330 | -.423 |
| South-East Asia | -.091 | -.244 | -.203 | -.201 | .250 | .138 |
| Western Pacific | -.111 | -.078 | -.069 | -.046 | .210 | .084 |
| Sig. (1-tailed) | Traffic  | . | .000 | .134 | .000 | .227 | .092 |
| Heavy  | .000 | . | .000 | .000 | .054 | .008 |
| Teenage | .134 | .000 | . | .000 | .004 | .000 |
| Consumption  | .000 | .000 | .000 | . | .121 | .002 |
| Minimum Age | .227 | .054 | .004 | .121 | . | .053 |
| BAC limit | .092 | .008 | .000 | .002 | .053 | . |
| Low-income | .008 | .000 | .000 | .000 | .264 | .033 |
| Lower-middle-income | .071 | .044 | .001 | .287 | .013 | .470 |
| Upper-middle-income | .011 | .360 | .317 | .139 | .429 | .030 |
| Africa | .334 | .000 | .000 | .001 | .207 | .000 |
| Eastern Mediterranean | .336 | .103 | .006 | .066 | .489 | .403 |
| Europe | .124 | .000 | .000 | .000 | .000 | .000 |
| South-East Asia | .182 | .007 | .021 | .022 | .006 | .084 |
| Western Pacific | .135 | .219 | .245 | .322 | .018 | .203 |

The Pearson Correlation values at the top of the chart show the correlations between all of my variables. Below the correlation values are the respective significance values. Heavy drinking has a .474 correlation coefficient with traffic fatalities. This means there is a significant, moderate, positive correlation between the two variables. Total consumption per capita has a correlation coefficient of .559 which is also a significant, stronger, positive correlation with traffic fatalities. My dummy variables for income group are significant as well. The low-income group has a negative correlation with traffic fatalities while the lower-middle and upper-middle groups have a positive correlation with traffic fatalities. My other independent variables do not have significant correlations with my dependent variable.

|  |
| --- |
| **Model Summaryb** |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | Durbin-Watson |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .758a | .574 | .510 | 10.3155 | .574 | 9.019 | 13 | 87 | .000 | 2.203 |
| a. Predictors: (Constant), Western Pacific, Lower-middle-income, Consumption, Eastern Mediterranean, BAC limit, Minimum Age, South-East Asia, Upper-middle-income, Africa, Heavy, Low-income, Europe, Teenage |
| b. Dependent Variable: Traffic |

R is the correlation coefficient for all of the variables included in the model. R is .758 in my model, which means that there is a strong relationship between my variables. R Square is the coefficient of determination and it shows how much of the variability in the dependent variable can be accounted for by the independent variables in the model. In this case, my model explains 57.4% of the variability in traffic fatalities. The other 42.6% of variability may be due to other variables not included or to random chance. The standard error of the estimate, 10.3155 measures the dispersion in the predicted scores, or the standard deviation of the residuals. The Durbin-Watson statistic is close to 2 (2.203), which means that there is no correlation between the residuals.

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | Correlations | Collinearity Statistics |
| B | Std. Error | Beta | Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | -32.079 | 21.719 |  | -1.477 | .143 | -75.247 | 11.089 |  |  |  |  |  |
| Heavy | .501 | .102 | .499 | 4.928 | .000 | .299 | .703 | .474 | .467 | .345 | .479 | 2.090 |
| Teenage | -.029 | .113 | -.038 | -.260 | .796 | -.253 | .195 | .111 | -.028 | -.018 | .226 | 4.432 |
| Consumption | 1.237 | .257 | .475 | 4.816 | .000 | .727 | 1.748 | .559 | .459 | .337 | .503 | 1.989 |
| Minimum Age | .980 | 1.112 | .069 | .881 | .381 | -1.230 | 3.190 | .075 | .094 | .062 | .790 | 1.266 |
| BAC limit | 39.899 | 57.882 | .059 | .689 | .492 | -75.148 | 154.946 | -.133 | .074 | .048 | .677 | 1.477 |
| Low-income | 10.203 | 4.637 | .260 | 2.200 | .030 | .986 | 19.420 | -.238 | .230 | .154 | .350 | 2.857 |
| Lower-middle-income | 14.330 | 3.672 | .410 | 3.903 | .000 | 7.031 | 21.628 | .147 | .386 | .273 | .444 | 2.250 |
| Upper-middle-income | 13.866 | 3.243 | .428 | 4.276 | .000 | 7.420 | 20.312 | .228 | .417 | .299 | .489 | 2.043 |
| Africa | 2.791 | 4.111 | .083 | .679 | .499 | -5.379 | 10.962 | -.043 | .073 | .048 | .326 | 3.066 |
| Eastern Mediterranean | 6.600 | 11.815 | .045 | .559 | .578 | -16.883 | 30.084 | -.043 | .060 | .039 | .770 | 1.299 |
| Europe | -2.901 | 3.815 | -.096 | -.760 | .449 | -10.484 | 4.682 | .116 | -.081 | -.053 | .305 | 3.275 |
| South-East Asia | 2.015 | 6.406 | .027 | .315 | .754 | -10.717 | 14.747 | -.091 | .034 | .022 | .675 | 1.481 |
| Western Pacific | -1.018 | 4.393 | -.020 | -.232 | .817 | -9.749 | 7.713 | -.111 | -.025 | -.016 | .673 | 1.487 |
| a. Dependent Variable: Traffic |

 The constant is also known as the Y intercept and is the predicted value of the dependent variable when all of the other variables are 0. This value is -32 in my model which does not mean much since the percentage of alcohol-attributable traffic fatalities cannot be below 0. The unstandardized coefficients show the impact of each independent variable on the dependent variable. For example, for every 1% increase in heavy drinkers, the traffic fatality rate goes up by 0.5%. For every 1 liter increase in pure alcohol drunk per capita, the traffic fatality rate goes up by 1.24%. For every increase of 1 year as the minimum legal drinking age, the traffic fatality rate goes up by almost 1%. Based on this table, my regression line equation would be:

Traffic fatalities = -32.079 + .501(heavy) - .029(teenage) + 1.237(consumption) + .980(minimum age) + 39.899(BAC limit) + 10.203(low-income) + 14.330(lower-middle-income) + 13.866(upper-middle-income) + 2.791(Africa) + 6.600(Eastern Mediterranean) - 2.901(Europe) + 2.015(South-East Asia) - 1.018(Western Pacific)

The standardized coefficients, or beta, show the coefficients you would obtain if you standardized all of the variables first before running the regression. By looking at these numbers, we can compare the coefficients for the different independent variables and see which variables have the greatest impact on the dependent variable. Heavy drinking, total consumption, lower-middle income, and upper-middle income have the most impact on the number of traffic fatalities attributed to alcohol use.

The significance values show us which coefficients are not due to random chance. I found that heavy drinking, total consumption per capita, and income group are all significant while teenage drinking, minimum age, BAC limit, and region are not significant. The collinearity statistics in the table show us if there is multicollinearity within our data. A tolerance of 0.2 or lower may be because of multicollinearity. Although some of my tolerance values are low (such as for teenage drinking and the African and European regions), none are below 0.2.



This graph plots the frequency of the standardized residuals of my dependent variable, traffic fatalities attributed to alcohol use. This graph shows a relatively symmetrical normal curve with just a slight positive skew. This means my model under-predicted a couple of the data points.

 Next, I ran an ordinary least squares regression in GeoDa to see if there was any spatial autocorrelation present in my data. I used a Queen-Based Contiguity weight so countries that were sharing a side or just a point were included in the spatial analysis. It shows that the mean of alcohol-attributable traffic fatalities is 10.89% and the standard deviation is 13.93%. I did not run all of the independent variables in GeoDa, so my R-squared is slightly different.

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : shapefiledata

Dependent Variable : Alcohol\_Tr Number of Observations: 246

Mean dependent var : 10.8951 Number of Variables : 5

S.D. dependent var : 13.9257 Degrees of Freedom : 241

R-squared : 0.589866 F-statistic : 86.6532

Adjusted R-squared : 0.583059 Prob(F-statistic) : 1.4013e-045

Sum squared residual: 19565.7 Log likelihood : -887.332

Sigma-square : 81.1856 Akaike info criterion : 1784.66

S.E. of regression : 9.01031 Schwarz criterion : 1802.19

Sigma-square ML : 79.5355

S.E of regression ML: 8.91827

-----------------------------------------------------------------------------------------------------

 Variable Coefficient Std.Error t-Statistic Probability

-----------------------------------------------------------------------------------------------------

 CONSTANT 1.169421 0.912543 1.281497 0.20125

 Heavy 0.2943934 0.06622252 4.445518 0.00001

 Teenage -0.1028357 0.03546624 -2.899537 0.00408

 Consumption 1.409268 0.1507741 9.346885 0.00000

 BAC Limit 8.590454 19.9793 0.4299678 0.66760

-----------------------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 6.608883

TEST ON NORMALITY OF ERRORS

TEST DF VALUE PROB

Jarque-Bera 2 349.8284 0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST DF VALUE PROB

Breusch-Pagan test 4 49.2432 0.00000

Koenker-Bassett test 4 14.4171 0.00608

SPECIFICATION ROBUST TEST

TEST DF VALUE PROB

White 14 53.0115 0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : shapefilequeen.gal

 (row-standardized weights)

TEST MI/DF VALUE PROB

Moran's I (error) 0.3908 8.5243 0.00000

Lagrange Multiplier (lag) 1 50.1717 0.00000

Robust LM (lag) 1 4.3238 0.03758

Lagrange Multiplier (error) 1 68.3555 0.00000

Robust LM (error) 1 22.5076 0.00000

Lagrange Multiplier (SARMA) 2 72.6793 0.00000

========================== END OF REPORT =========================

Again, we can see that heavy drinking and total consumption are both significant and the BAC limit is insignificant. However, in this regression output, teenage drinking is now significant and it was not significant before with the other independent variables added in. All of the independent variables have the same direction relationship as they had in the previous model. The multicollinearity condition number is under 30, meaning that there is no multicollinearity present among my independent variables. The Jarque-Bera test is statistically significant which means my model may be biased. The Breusch-Pagan test, Koenker-Bassett test, and White test are all significant which indicates heteroskedasticity or a relationship between some of my independent variables. If running this model in the future, I should try to eliminate one or more variables.

The Moran’s I significance value is 0.00 which means that there is a 99.9% chance that my data has some amount of spatial autocorrelation of the residuals. Because both of the Lagrange Multipliers also have a significance of 0.00, I looked at the Robust LM significance values to determine which spatial model to run. Since the Robust LM (error) also has a significance of 0.00, I ran a spatial error model.

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : shapefiledata

Spatial Weight : shapefilequeen.gal

Dependent Variable : Alcohol\_Tr Number of Observations: 246

Mean dependent var : 10.895122 Number of Variables : 5

S.D. dependent var : 13.925723 Degrees of Freedom : 241

Lag coeff. (Lambda) : 0.530462

R-squared : 0.703538 R-squared (BUSE) : -

Sq. Correlation : - Log likelihood : -857.256296

Sigma-square : 57.4916 Akaike info criterion : 1724.51

S.E of regression : 7.58232 Schwarz criterion : 1742.04

-----------------------------------------------------------------------

 Variable Coefficient Std.Error z-value Probability

-----------------------------------------------------------------------

 CONSTANT 0.3360357 0.8443355 0.3979884 0.69064

 Heavy 0.3442109 0.05310396 6.481831 0.00000

 Teenage -0.08369535 0.03176218 -2.635063 0.00841

 Consumption 1.27194 0.147529 8.621625 0.00000

 BAC Limit -21.12373 17.6602 -1.196121 0.23165

 LAMBDA 0.5304618 0.05967563 8.889086 0.00000

-----------------------------------------------------------------------

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST DF VALUE PROB

Breusch-Pagan test 4 49.0652 0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : shapefilequeen.gal

TEST DF VALUE PROB

Likelihood Ratio Test 1 60.1513 0.00000

========================== END OF REPORT =========================

 Since the Log likelihood, Akaike info criterion, and Schwarz criterion are all closer to 0 in this spatial error model compared to the ordinary least squares regression, we can say that this spatial error model explains more of the data. Heavy drinking and BAC limit became more significant in this model while teenage drinking became less significant. An additional variable, Lambda, has been added to my model. This variable, or error term, is significant with a positive correlation and helped to improve the fit of my model. The Breusch-Pagan test remained significant, suggesting the presence of heterskedasticity. The Likelihood Ratio Test is also significant which means that the spatial effects are not completely gone in this model.

 Next, I ran some spatial statistics on my model. First, I graphed my dependent variable against its spatially lagged variable. This shows that I have positive spatial autocorrelation which means that locations with high alcohol-attributable traffic fatalities are close to other similar locations. This is also true in the reverse. Places with low alcohol-attributable traffic fatalities are next to each other as well.



Next, I looked at local spatial autocorrelation. The map on the left shows places where there is significant spatial autocorrelation. The darker the color, the more significant it is. The map on the right shows the specific type of spatial autocorrelation. Red countries are countries with high alcohol-attributable traffic fatality rates that are next to similar countries. Pink countries are countries with high alcohol-attributable traffic fatality rates that are next to countries with low rates. Dark blue countries are countries with low alcohol-attributable traffic fatality rates that are next to similar countries. Light blue countries are countries with low alcohol-attributable traffic fatality rates that are next to countries with high rates.

**Conclusion**

 According to the statistical analyses I ran, alcohol-attributable traffic fatalities are most impacted by the percentage of heavy drinkers in the country, the total consumption of alcohol per capita, and the income group to which a country belongs. The higher the percentage of heavy drinkers and the higher the consumption of alcohol per capita, the higher the percentage of traffic fatalities attributed to alcohol use. Countries in the lower-middle-income or upper-middle-income groups also indicate a higher percentage of traffic fatalities attributed to alcohol use. The percentage of teenage drinkers was found to be insignificant when ran with all of my variables, but significant when ran with only a couple of the other variables. This indicates that it may be correlated with another of my independent variables and more tests would need to be ran, in order to see if teenage drinking has an effect on the percentage of alcohol-attributable traffic fatalities. The other variables I ran were not significant, but their effects in my model are listed in the table below.

There is also positive spatial autocorrelation between traffic fatalities in different countries. Countries with high rates of alcohol-attributable traffic fatalities are close to similar countries. The same is true for countries with low rates of alcohol-attributable traffic fatalities. Countries in Asia, Eastern Europe, Southern Africa, and Northern South America tend to have high rates of traffic fatalities attributed to alcohol use. Northern African countries have lower rates of alcohol-attributable traffic fatalities.

|  |
| --- |
| Variable Results |
|  | Variable | Hypothesized Effect (+/-) | Actual Effect (+/-) |
| Dependent | Traffic |  |  |
| Independent  | Heavy Drinking | + | + |
| Independent | Teenage Drinking | + | - |
| Independent | Consumption per Capita | + | + |
| Independent | Minimum Age | - | + |
| Independent | BAC Limit | + | + |
| Independent | Income | + for Upper-middle- for Lower-middle, Low | +  |
| Independent | Region | + for Eastern Mediterranean, Europe- for Africa, South-East Asia, Western Pacific | + for Africa, Eastern Mediterranean, South-East Asia- for Europe, Western Pacific |

**References**

Jiang, H., Livingston, M., & Room, R. (2015). Alcohol Consumption and Fatal Injuries in Australia Before and After Major Traffic Safety Initiatives: A Time Series Analysis. Alcoholism: Clinical & Experimental Research, 39(1), 175-183. doi:10.1111/acer.12609

Pemjean, A., Ramirez, J., & Zitko, P. (2015). New alcohol policies appear to have reduced traffic casualties in Chile. Addiction, 110(2), 369-370. doi:10.1111/add.12790

Schmitz, A. R., Goldim, J. R., Guimarães, L. P., Lopes, F. M., Kessler, F., Sousa, T., & ... Pechansky, F. (2014). Factors associated with recurrence of alcohol-related traffic violations in southern Brazil. Revista Brasileira De Psiquiatria, 36(3), 199-205. doi:10.1590/1516-4446-2013-1128

Sebego, M., Naumann, R. B., Rudd, R. A., Voetsch, K., Dellinger, A. M., & Ndlovu, C. (2014). The impact of alcohol and road traffic policies on crash rates in Botswana, 2004-2011: a time-series analysis. Accident; Analysis And Prevention, 7033-39. doi:10.1016/j.aap.2014.02.017