Running Head: INFLUENCE OF GEOSPATIAL FACTORS ON WILDFIRE OCCURRENCE

Executive Development

The Influence of Geospatial Factors on Wildfire Occurrence in the Black Hills of South Dakota

James E. Strain

South Dakota Department of Agriculture, Division of Wildland Fire Suppression

Rapid City, South Dakota

April 2008

CERTIFICATION STATEMENT

I hereby certify that this paper constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of another.

Signed:____

Abstract

This study used descriptive methodology to explore the problem of describing the geospatial factors that influence wildfire occurrence in the Black Hills of South Dakota. The research purpose analyzed data generated from a Geographical Information System (GIS) using input from Global Positioning System (GPS) units to describe the variables. The research questions were: (a) the type of statistical analysis needed to analyze the geospatial factors; and (b) the influence of the variables of elevation, slope, aspect and road-distance to wildfire occurrence. The procedures involved a nonparametric statistical test. Results showed that elevation, slope, and road-distance are significant factors for occurrence. Recommendations are to apply the results for fire prevention and initial attack strategy at the department level.

Table of Contents

Figure

Appendices

The Influence of Geospatial Factors on Wildfire Occurrence in the Black Hills of South Dakota

Introduction

The Black Hills of South Dakota are located in the southwestern part of the state. A small portion of this mountain range can be found in the northeastern part of Wyoming. This mountainous area is internationally known for the granite rock carvings of Crazy Horse and Mount Rushmore. The Black Hills are considered sacred to the Lakota Native American tribes that live around the region. The densely settled areas found in the valleys adjacent to the intermix of federal and state lands of the mountain areas, combine for complex wildland fire environment with forested areas of ponderosa pine that will burn frequently and severely every fire season with the potential for catastrophic wildland fire that can threaten human life and structures.

Within this setting it should be noted that in the last three years, 2005-2007, the State of South Dakota Division of Wildland Fire Suppression has equipped its agency firefighters with handheld GPS units to accurately locate the point of origin of wildland fires in the Black Hills. In addition, many Geographical Information Systems (GIS) are coming online through internet access to bring accurate spatial data to describe historical wildfire activity.

The time is ready to present the problem of how to describe the geospatial factors that influence wildfire occurrence in the Black Hills of South Dakota. The research purpose will indentify a science based approach based on statistical analysis of descriptive data generated from a GIS database to describe the geospatial factors that influence wildfire occurrence in the Black Hills of South Dakota.

The descriptive research method was used for this research project. The research questions all deal with wildfire occurrence in the Black Hills of South Dakota and can be summarized in the following two points: (a) the type of statistical analysis that would best describe the geospatial factors that influence wildfire occurrence; and (b) the influence of the geospatial factors of site elevation, slope, aspect and road distance on wildfire occurrence.

Background and Significance

The problem of identifying geospatial factors and determining the influence of these explanatory variables on wildfire occurrence in the Black Hills is an important subject area of study. This is because the study area is prone to large scale wildfire activity that will burn across very rugged terrain. These rugged terrain features dissect the landscape through a dense pattern of wildland-urban interface areas. On July 8th of 2007, the Alabaugh Canyon wildfire started in rugged terrain next to a rural subdivision and destroyed over 33 single family residential structures and claimed the life of one civilian, in addition to injuring two firefighters. (National Interagency Fire Center, 2008) (USDA Forest Service, 2008). The Alabaugh fire started on the jurisdiction of the author's fire department, the South Dakota Division of Wildland Fire Suppression.

The State of South Dakota carries out its jurisdictional responsibility to suppress wildland fire on state and privately owned forested lands in the Black Hills through the statutory power vested in the Secretary of Agriculture. That power is delegated to the Division of Wildland Fire Suppression (WFS), the "fire department" for the State of South Dakota, and the agency carries out that mission of wildfire suppression on approximately 283,280 hectares (ha) in the Black Hills in cooperation with local fire departments and federal fire agencies. Therefore the study of the local geospatial factors that influence wildland fire occurrence is very important to WFS given that the terrain varies significantly from flat to very steep, from box canyons and steep mountains to flat agricultural fields and prairies over the area of its jurisdictional responsibility.

Another need for this analysis to be carried out at the department level is that there is no "one size fits all" national solution or model that has been discovered when researching the factors that influence wildland fire occurrence. Stephens (2005) noted through his national study of large-scale analysis of forest fire causes on United States Forest Service lands that there was a high geographical diversity of wildfire occurrence across the nation and that local solutions were needed to solve local wildfire management problems.

Furthermore, this research project should show linkages to the local level within the department's initial attack area that will provide direction for the future with regards to more efficient initial attack strategies and guidance to future prevention efforts. This will be increasingly important in a wildland setting where the ponderosa pine forest in the Black Hills before European settlement burned on average every 16 years. (Brown & Sieg, 1996)

The research purpose of identifying the geospatial factors that influence wildland fire occurrence fits well with the first year Executive Development course objectives of Change Management and Service Quality. The effort involved in the validation of the database used in this applied research project shows the need to be able to manage change in a manner that is beneficial to the fire service and to the public it serves. A case in point exists in the gathering of the data and the validation of the dataset for this project. When Global Positioning System (GPS) units were first deployed in the field in the study area, there was no standardization of a process for x-y coordinate reporting by firefighters or dispatchers. This resulted in erroneous location data being reported into dispatch and recorded into the archival fire records. However once

Standard Operating Procedures were developed and implemented, data collection was more accurate and a higher level of confidence could be placed into the veracity of the data. Hopefully, any fire service officer or researcher who will reference this study will understand the need to manage change in order to maintain the service quality when new technology is deployed in the field.

The objective of this applied research project is primarily congruent with United States Fire Administration Operational Objective Number Five: To respond appropriately in a timely manner to emerging issues. The collection and analysis of accurate GPS data on wildfire point of origins has just started nationwide with the wildland fire agencies, and combined with the recent advent of user-friendly online internet based GIS databases, perhaps this study can show the fire service another way to manage the issue of spatial data analysis in an inexpensive and simple manner to benefit their fire department as it relates to wildfire occurrence.

Literature Review

The literature review of wildfire occurrence shows a range from international in scope, to a national or federal perspective for the USA, to that of smaller scale studies that provide a more in-depth look at a state or region.

Statistical Analysis

The review of the first research question which seeks to determine the type of statistical analysis that would best describe the geospatial factors that influence wildfire occurrence will start with *Fire Data Analysis Handbook, 2nd Edition* (FEMA, 2004). A template in this handbook is provided for a researcher to follow a four-stage process of: (a) collection of the data; (b) summarizing the information from the data; (c) understanding of the processed information from analyzing the data; and (d) allowing the utilization of information so the appropriate decision-making process is carried out. This research methodology is evident in the seminal research done by the United States Forest Service on the analysis of forest fire causation in Pennsylvania (Haines, Main, & McNamara, 1978). In that study, the focus was to present just enough statistical analysis for management planning even though the researchers could have presented more complex statistical analysis in their research report. As the research of fire data analysis moved into the decade of the 1980’s, multivariate regression analysis become more commonly used as an analytical method of research into the occurrence of wildfire, as one study was done in 1984 that covered 27 states (Donoghue & Main, 1985).

Later research typically involved more sophisticated statistical analysis aided by more convenient access to more powerful computer systems and software. This type of analysis moved away from the philosophy of Haines and the other pioneer researchers where just enough statistical analysis was done for the field level to understand and implement a decision making process to more complex statistical models to explain wildland fire occurrence. The latter trend was best typified by Stephens (2005) in a research paper that presented three types of statistical analysis working towards a model of multi-modal inference: analysis of variance, regression analysis, and time series analysis. Research at that time also moved in the direction of building mathematical models such as complex production functions borrowed from the field of econometrics and adapted for use in predicting wildfire events in the state of Florida (Mercer & Prestemon, 2005).

The latest trend in the statistical analysis of wildfire occurrence can be found in research done on the Mark Twain National Forest in Missouri. Researchers in that study analyzed a dataset from 1986-2002 using logistic regression models and classification and regression tree analysis (Brosofske, Cleland, & Saunders, 2007). The analytical method of classification and regression tree analysis (CART) was first postulated in a book published in 1984 by a group of researchers from the Universities of California-Berkeley and Stanford (Breiman, Friedman, Olshen, & Stone, 1984). Although new to fire data analysis, CART has been used extensively in the medical research field (Lewis, 2000). Furthermore, marine biologists and ecologists that have researched subjects such as the geospatial factors of coral reefs have used CART to explain the variation of location of different soft coral taxa found in the Great Barrier Reef of Australia. (De'ath & Fabricius, 2000).

However, further review of the literature does show a concern that classification tree analysis can be used as a tool for careless data-mining resulting in analysis that will infer causation or show correlation between the response variable and the explanatory variables when none exists. This can lead to the fallacy of Post-hoc analysis when data-dredging leads to conclusions that cannot be tested because they were not posited as “a priori” research questions at the start of the research investigation. (Anderson & Burnham, 2002). In addition, Gay and Airasian (2003, p. 471) warn against a “fishing expedition” that is just looking for any spurious cause and effect relationship from the analysis of data.

When analyzing data for spatial analysis regarding the point of origin of a wildfire, understanding location error is important. An example of location error is the 1.6 km location error for point of origin that is accepted into the Lake States Fire Database that was used for the Upper Midwest study (Cardille, et al., 2001). Other research studies do not speak to location error, or are studies that are looking at aggregate fire size without regards to the wildfire point of origin.

The issue of the distribution of wildfire area burned data is an important issue to consider for statistical analysis. This data tends to be heavily right-skewed due to the average nationwide initial-attack efficiency of containing approximately 97% to 99% of wildfires at small sizes (< 0.1 ha) as opposed to larger sizes (Stephens & Ruth, 2005). Solutions to transform this type of data to a more normal distribution were varied: a national study of USFS data used a $\log_{10}(x + 1)$ transformation (Stephens, 2005) ; the California data required a square-root transformation (Syphard, et al., 2007); the upper Midwest research based each regression analysis on three different distributions, (Binominal, Poisson, Negative Binomal) (Cardille, Ventura, & Turner, 2001); the Missouri study used a nonparametric CART analysis and transformed the fire data for regression analsyis (Brosofske, et al., 2007); and the Kentucky study utilized the nonparameteric Kruskal-Wallis test for analysis based on their assumption of non-normality of the data (Maingi & Henry, 2007).

Geospatial Factors

The research questions of wildfire occurrence with regards to the geospatial factors of road distance, elevation, slope and aspect in the Black Hills of South Dakota will have to be reviewed outside of the study area; because a search of the literature shows that no research on those geospatial factors has been done within the study area. The study of the influence of geospatial variables falls under the major variable of topography, which is one of three major variables of fuel, weather, and topography that have impact on wildfire behavior. Of those three variables, considerable research has been shown that all three are interrelated to a significant degree, but weather is the most important variable in determining wildfire behavior (Pyne, Andrews, & and Laven, 1996).

The research question of the geospatial factor of road distance and wildfire occurrence has been studied throughout the United States. Some surrogate variables for this geospatial factor in other research studies have been labeled road density, and housing density (based on the assumption that there are roads leading into housing areas). Research done on the Mark Twain National Forest (MTNF) showed that distance to roads had a significant value of importance in wildfire occurrence with regards to human factors. But this variable in that study did not rank in importance as some of the ecological variables such as ecosystem fire resistance and ecosystem subsection in explaining the variation in wildfire occurrence on the MTNF. (Brosofske, et al., 2007). Studies of wildfire risk factors in northern Florida found that the wildland-urban interface was identified as a variable that was statistically insignificant in explaining the risk factors from wildfire occurrence next to housing areas, but the researchers did point out the need for future research in this area (Prestemon, Pye, Butry, Holmes, & Mercer, 2002). Research into wildfire occurrence in California showed no explanatory significance with a distance to road variable (Syphard, et al., 2007).

However, research on wildfire occurrence in the forested areas of Minnesota, Wisconsin and Michigan indicate that an increase in human access and activity showed a positive correlation to an increase in wildfire occurrence (Cardille, et al., 2001). Stephens and Ruth (2005) in a nationwide study observed a trend that human-caused fires on federal lands commonly occur near transportation corridors. Research on a dataset of wildland fire occurrence from 1985-2002 in the Appalachian mountains of eastern Kentucky indicated that a distance to roads explanatory variable explained 54% of the total variance observed in the density of wildfire that occurred (Maingi & Henry, 2007).

The remaining research questions of the influence on wildfire occurrence of the geospatial factors of elevation, slope, and aspect shows a varying amount of research that has been accomplished to date. Furman (1978) tested for various climatic and fuel explanatory variables at different aspects and elevation over 1000 meter (m) elevation difference on a mountain range in the western U.S. In that study, no stand-alone influence was found for either elevation or aspect, but rather a combination of aspect and elevation was better explanatory variable for wildfire ignition. Research on climate change and its influence on large-scale wildfire frequency across the western U.S. observed that the northern Rockies of Montana and Idaho showed a significant increase of wildfire occurrence at mid-elevation sites since the mid-1980's (Westerling, Hidalgo, Cayan, & Swetnam, 2006). An extensive study on tree rings in a ponderosa pine forest in Colorado showed more wildfire occurrence in both in hectares burned and increased frequency at lower elevations than at higher elevations (Veblen, Kitzberger, & Donnegan, 2000). But contrasting research findings were found with respect to elevation and slope on fire occurrence on both the MTNF study area in Missouri and the eastern Kentucky study area. On the MTNF, elevation and slope were of low importance in the regression models used in the study (Brosofske, et al. 2007). But in the forested areas of eastern Kentucky, there was observed a significant relationship for slope and elevation as explanatory variables for wildfire occurrence (Maingi & Henry, 2007). However both Brosofske et al. and Maingi and Henry observed in both studies in 2007 that the variable of aspect showed no statistical power in explaining the response of wildfire occurrence in both study areas.

In summarization of the literature review, it appears that many statistical analyses have been used over the years with regard to wildfire occurrence data. The process outlined in the US Fire Administration's second edition handbook on data analysis outlines a four stage process for analyzing fire data. Since the advent of more accessible and faster computing power and sophisticated software packages, the complexity of analysis has been more complex and more use of multi-modal inference is noted in the literature. CART analysis has started to demonstrate promise as a tool used in recent years for wildfire data analysis. Reviewing the literature with regard to the research questions of the geospatial factors of road location, elevation, slope and aspect in the Black Hills of South Dakota, a search of the literature has no information on that particular study area. When reviewing the body of literature in other parts of the USA, it appears that are more local influences on wildfire occurrence in specific study areas that what has been observed on a coarse aggregate scale at the national level. This has led to conflicting observations on the geospatial variables of road location, elevation and slope in other study areas and no national trend or pattern has been observed by researchers on the subject of these explanatory variables. It does appear from the limited amount of research on the subject that aspect has been observed so far to play no significant role in wildfire occurrence, unless it is combined with elevation or slope.

Procedures

The methodology used in studying the research questions involved the following procedures: (a) delineation of the study area; (b) framing the null (H_0) and alternative hypotheses (H_1) under which the research questions will be tested; (c) constructing and validating a database to produce the datasets used for analysis; and (d) choosing and conducting the correct statistical analysis to test the selected H_0 .

The general location of the study area is shown on page 15 (figure 1). The study area comprises all of the state and private land within the initial attack area within and adjacent to the Black Hills Forest Fire Protection District which is the jurisdictional area for the (WFS).

The next step of the procedure was to construct the null (H_0) and alternative hypotheses (H_1) under which the research questions will be tested. For this applied research project, independent variables are called "*explanatory*" variables, and dependent variables are called "*response*" variables. The H_0 for all testing of the geospatial factors or variables and their influence on wildfire occurrence is that there is no statistically significant relationship between the explanatory variables of elevation, aspect, slope, or road distance and any of the response variables of ha-burned or frequency of fire occurrence.

The H_1 for all testing of the geospatial factors or variables and their influence on wildfire occurrence is that there is a statistically significant relationship between the explanatory variables of elevation, aspect, slope, or road distance and any of the response variables of ha-burned or frequency of fire occurrence.

The H_0 is the expected outcome for all statistical analysis in this study design and the H_0 will only be rejected in favor of the H_1 when it can be statistically proven that the observed relationship from the sample can be expected to be observed at least 95% of the time, or conversely, the H_1 will be accepted when the probability level is less than 5 percent ($\alpha < 5\%$).

Database

The database used for this analysis was gathered from the initial attack fire records maintained by the Northern Great Plains Interagency Dispatch Center in Rapid City, SD. The database used for this study contained the record set of calendar years (CY) 1994-2007. However in the process of validating the data, errors were discovered in how the location coordinates of latitude and longitude (x-y coordinates) of the wildfire location were recorded in the database. It was apparent that prior CY 2005 many x-y coordinates were derived from the reported Public Land Survey (PLS) Township, Range, Section (TRS) location reported by the initial attack incident commander. These TRS locations were accurate from 2.5 acres (1.01 ha) to 40 acres (16.6 ha) but when converted to x-y coordinates of decimal degrees (e.g., -104.3659, 44.2036) the algorithm used for the transformation would return a value from the inputted TRS location to a x-y coordinate in the exact middle of the TRS location. This resulted in errors of location that could range from 0-800 meters. Therefore, any distance error would affect the reliability of data to reflect an accurate location of the fire. In some cases, the transformed x-y coordinates changed the jurisdiction of the wildfire location by placing the wildfire point of origin in a different landownership (e.g., federal vs. private). Such location error would change the geospatial variables of elevation, slope, aspect and distance to roads in mountainous terrain.

As a result of this transformation of the location data, no assumptions could be made as to the overall reliability of the x-y coordinate data until CY2005. Then standard operating procedures combined with the availability of handheld GPS units for use by the initial attack incident commanders allowed for accurate reporting of x-y coordinates from the field without having to transform the wildfire location from TRS data. This allowed the use of WFS archival fire records from CY2005 to CY2007 for the study. These fire records were enhanced with the inclusion of geospatial data generated from an internet-based GIS database called the *Graphical Locater* created and maintained by Environmental Statistics Group (ESG) at Montana State University, Bozeman, MT. The Graphical Locater will return values from inputted x-y coordinates for geospatial variables such elevation, slope, aspect, roughness of terrain, and nearest landmarks compiled from United States Geological Survey databases. The Graphical Locater will also create 2 km resolution map showing the nearest road and water features from the inputted x-y coordinate data (Environmental Statistics Group, Montana State University, 2005).

The following data was collected from each record in the CY2005-2007 database and merged with the geospatial data from the Graphical Locater and maintained in an electronic spreadsheet. The database may be viewed and downloaded at the author's webpage at http://www.state.sd.us/doa/wfs/chieffmo.htm for use by other researchers. The lists of variables are as follows:

1. Ha-burned. A response variable that is a measurement of converted acres to ha and was created by multiplying the "acres burned" observation by .40469 and rounding out to 2 decimal points.
2. Frequency. The count data for each individual record. A response variable.
3. Longitude. This is the decimal degree observation of Longitude that was converted from the degrees/minutes/seconds (DMS) longitude report to dispatchers by the initial attack incident commander. The recreation grade handheld GPS units used for the measurement report are subject to a location error on average of ± 10 m (ESRI, 2004).
4. Latitude. Derived in same manner as Longitude and subject to same error.
5. Elevation. An explanatory variable that is derived from inputting the x-y coordinates of Latitude and Longitude into the Graphical Locater. The returned value of elevation is measured in meters above Mean Sea Level and subject to an average vertical location error of ± 5.5 m.

6. Slope. The measure in percent slope of the line that is through the elevation point and measured at its maximum slope. An explanatory variable.
7. Aspect. The direction the slope is facing on the earth's surface. Based on the degrees of the compass and reported from the Graphical Locater in degrees from 0-360 with one decimal point. An explanatory variable.
8. Road-Dist. The straight-line distance to the nearest road feature from the x-y coordinate input as marked on the 2km-20km resolution maps available on the Graphical Locater and rounded off to the nearest 5 meter whole measurement. An explanatory variable.
9. HWC. Human Wildfire Cause. Nominal variable that is categorized into 10 casual factors: (a) arson, (b) campfire, (c) children, (d) debris burning, (e) equipment use, (f) fireworks, (g) railroads, (h) power lines, (i) smoking, and (j) miscellaneous. This nominal variable is used only for analysis of human-caused fires. HWC is also used as term to describe one of the two datasets used in the study.
10. Seasonal. Nominal variable that is classified into 4 categories: (a) Winter, (Dec/Jan/Feb); (b) Spring (Mar/Apr/May); (c) Summer, (Jun/Jul/Aug); and (d) Fall, (Sep/Oct/Nov). This nominal variable is used only for analysis of human-caused fires.
11. Month. Classified into 5 categories: May, June, July, August, and September. This nominal variable is used only for the analysis of lightning-caused wildfires.
12. District. Nominal Variable that is categorized into the four districts of the WFS jurisdictional area: Rapid City, Hot Springs, Lead, and Custer State Park.

Statistical Methods

The last research question to answer under this procedure is that of the proper statistical analysis to use when testing the H_0. Given that the CY2005-2007 database has a sample size of *n* = 328 and it was divided into two subsamples: the Lightning dataset (*n* = 157) and the human-caused dataset (*n* = 171); some statistical assumptions have to be tested as to the distribution of the data relative to sample size. Therefore the following statistical tests were selected and performed on the subsamples of the lightning and human-caused datasets.

1. Kolmogorov-Smirnov Comparison of Two Data Sets to check for a normal distribution and generate descriptive statistics of the mean, mode and median. Although the Kolmogorov-Smirnov test does not test for a normal distribution, the statistics generated by the calculator does give information as the probability of a normal distribution and if it can be log-transformed to a normal distribution. (Kirkman, 1996). Also histograms of the sample data were generated to visually compare the measures of central tendency by inputting the data into an internet-based graphing spreadsheet function to see if log-normal or square-root transformations would bring the distribution of the dataset back to the "bell" curve. (McDonald, 2008). Additionally, the spreadsheet function had a calculator that determined the skewness of the dataset.
2. If the assumptions of normality could be made for the sample distribution, then One-Way ANOVA (Model I) test will be conducted on both subsamples using a balanced design based on equal samples to test the H_0 that there is no relationship between the explanatory variables of elevation, slope, road distance and the response variable of ha-burned based on the comparison of means by group. Given that aspect is a circular variable (0-360), then an unbalanced design will have to be used for that variable. A planned comparison of the means

would be carried out in to determine the source of significant difference, but no more than two means will be compared in order to maintain the integrity of the study design. The means to be tested would be the lowest (0-25%) and highest treatment (75-100%) level based on the literature discussion, and between North and South aspects, for the aspect test.

3. The Kruskal-Wallis will be conducted as the alternative test if the sample distribution is not normal. This test will be used on the same subsets of data planned for the One-Way ANOVA test.
4. An Exact Contingency Table analysis will be used to test the H_0 that the variables of elevation, slope, aspect and road distance have no relationship to the frequency of wildfire as categorized by the nominal response variables of HWC, Seasonal, Month or District.
5. A T-Test for Independent Samples will be used to test the equality of means between the geospatial variables of the Lightning and HWC datasets if the sample data can be assumed to be normal. This was used to test the H_0 that no significant difference exists between the means of following measurement variables: elevation, slope, aspect, and road distance. An example of the H_0 would be there is no difference between the average distances that a human-caused fire starts from a road as opposed to a lightning-caused wildfire. It is planned in this study design to compare the means of the Road-Dist., Slope and Elevation variable for both subsamples of Lightning and Human-caused wildfire. The Mann–Whitney U-test will be conducted as the nonparametric alternative test.
6. Descriptive statistics for each individual WFS district will be generated for baseline data for future applied research in the agency.

The analysis was based on the assumption that there may be a need for both a parametric (ANOVA) test and non-parametric test (Kruskal-Wallis) comparisons of means by group or rank when testing the H_0 given the sample size and the skewed distribution of the dataset. The Exact Contingency Table analysis was chosen over the Chi-Square Test of Independence because it is expected that the some cells in the table analysis will have less than five expected outcomes and some zero outcomes may be expected. In addition, the Chi-Square Test of Independence is more accurate with sample sizes $n > 1000$ (McDonald, 2007) and (Lowry, 2008).

The computational analysis was performed by free statistical calculator websites developed and provided by the Department of Physics, College of St. Benedict and St. John's University in Minnesota (Kirkman, 1996); R. Lowry, Department of Psychology, Vassar College, New York; and P. Wessa of the Office for Research Development and Education in Belgium (Wessa, 2008). In addition, useful spreadsheets for data analysis were provided on a website maintained by J.H. McDonald, Department of Biology, University of Delaware.

Results

A multi-modal approach was used to test the H_0 as stated by the research questions. No assumptions of normality could be made for the sample distributions of the response variable of ha-burned or the explanatory variables in both datasets because both datasets were skewed right and log transformations and square root transformations did not bring the sample(s) back to normal. Appendix A contains the summary results from this test for normality. Therefore the One-way ANOVA and T-tests were not conducted. The Kruskal-Wallis test was used to observe any statistical differences in comparison of the populations of the datasets based on quartile treatments of the explanatory variables of elevation, slope, aspect, and road-distance to the response variable of ha-burned. The quartile treatments were selected to allow for balanced design of the Kruskal-Wallis test, with the first quartile (0-25%) containing the lowest values of the explanatory variable and the last quartile (75-100%) contains the highest values of the explanatory variable.

This analysis was used to calculate any difference in quartile sample *X*'s of the geospatial variables for the ranked data of ha-burned response variable. The K-W test statistic *H* was calculated against an $\alpha < .05$ and was found to be insignificant for all quartile treatments of ha-burned, with the exception of two explanatory variables. Significant *H* values were found for Lightning-Elevation, ($H = 9.00$, $df = 3$, $p = .03$); and HWC-Elevation, ($H = 8.6$, $df = 3$, $p = .04$). Given that the experiment design for the K-W test was balanced in sample size to take into account the effects of variation within treatments, it would seem reasonable to seriously look at rejecting the H_0 in favor of H_1 for those two geospatial variables if other test results do not produce conflicting information. However, it must be taken into account that the K-W test does not compare sample X's or medians between treatments, as does ANOVA. It compares the rank of the value of the variables grouped by treatment. If a significant difference is found in the mean ranks, then it could be determined that the sample did not come from the same population as the rest of the samples in the K-W test.

The third method of analysis was to investigate the frequency of occurrence of the ha-burned variable in both datasets by categorizing the occurrence into a nominal variable that could be tested against various treatments of the geospatial variables. The major limitation of the exact table analysis is that of table size, which necessitates that a much coarser scale must be used in creating the nominal variables used in the tables. Table 1 contains the nominal variables used in the analysis and grouping or classification technique that was used in creating the row (treatments) of the geospatial variables and the column (outcomes) of the frequency of the response variable (ha-burned).

Table 1

Row and Column (RxC) Variables and Values used for Contingency Table Analysis

Variable	Type	Values
Elevation	Row	1000 - 1200 m, 1200 - 1400 m, 1400 - 1600 m, > 1600 m
Slope	Row	< 10%, < 20%, > 20%
Aspect	Row	Flat, North, South
Road-Distance	Row	0 - 200 m, 200 – 400 m, 400 - 600 m, > 600 m
Road-Distance	Row	0 - 200 m, > 200 m
Frequency/ha	Column	0 – 1.00 ha, > 1.00 ha
Frequency/month*	Column	April-May-June, July-August-September
Frequency Season**	Column	Winter, Spring, Summer, Fall
HWC Cause**	Column	Debris, Equipment, Power lines
District	Column	Lead, Rapid City, Hot Springs, Custer St. Park.

Note. * is used only for Lightning dataset, ** is used only for the HWC dataset.

Appendix B contains the probabilities associated with these exact tests. There are some significant relationships between the geospatial variables and response variables within the HWC dataset. In particular, the H_0 that the frequency of fire starts and the geospatial variables are independent of each other must be rejected in the case of the HWC-Elevation tests and some of the HWC Slope and Road-Dist. frequency tests. It does appear that the geospatial variable of Aspect is independent of any measure of frequency of wildfire tested with an exact test.

The last test conducted on the data was that of the comparing the sample *X's* of like geospatial variables between the Lightning and HWC datasets (e.g., *X* of HWC Road-Dist. vs. *X* Light Road-Dist). The H_0 is that both *X's* are from the same population for each set of explanatory variables. The test that was conducted was the nonparametric Mann–Whitney U-test. This test showed the probability of both *X*'s between HWC Road Dist. and Lightning Road-Dist; and HWC Slope and Lightning Slope; are not from the same population and H_0 to be rejected in the case of both comparisons.

Descriptive statistics were generated from the data analysis for WFS field office use and can be found in Appendix C.

Discussion

The major purpose of this study was to describe the effect of geospatial variables of elevation, slope, aspect and road-distance on wildfire occurrence within state and private land jurisdiction of WFS in the Black Hills of South Dakota. The research questions designed to explore this purpose were twofold: primarily, that the null hypothesis(s) are tested to explore the factors that influence wildfire occurrence in the study area; and to ensure the proper statistical analysis was used in the procedures so that the proper inferences would be drawn from the analysis.

The results of this applied research project were compared to other research findings that were published in the last decade in Missouri, Kentucky, Florida, California, Minnesota, Wisconsin, and Michigan. The research by Maingi and Henry (2007) in the Appalachian mountains of eastern Kentucky reported a different relationship for slope and elevation as compared to the applied research project. The Kentucky study on the Daniel Boone National Forest (DBNF) showed a strong statistical relationship between wildfire frequency and elevation with higher elevations exhibiting an increase in wildfire occurrence. Additionally, a strong association was found for the geospatial variable of slope and its influence on wildfire occurrence in that steeper slopes on the DBNF exhibited more fire occurrence than slopes that were less steep. Maingi and Henry also reported that the variable of aspect had negligible influence on wildfire occurrence in the DBNF.

The Kentucky study by Maingi and Henry (2007) analyzed a database of wildfires on the DBNF from 1985 to 2002 with $n = 2245$, with > 10% of fire occurrence started by lightning and the remainder were human-caused. Of the < 90% human-caused fires, approximately 75% were categorized into arson starts, leading to the premise that most of the wildfire occurrence on the DBNF is started by arson activity on steeper slopes in the higher elevations of the forest.

Contrasting this study area to the DBNF, approximately 47% of the fire occurrence was originated by lightning and less than 3% had arson as a casual factor on state and private lands in the Black Hills from 2005 to 2007. The analysis of the contingency tables from the HWC data showed that a significant frequency of fires (62%) started in elevation zones < 1400 m in a range from 1000 m to 2100 m. Furthermore the contingency table analysis showed that the geospatial variable of slope influence on wildfire occurrence in the Black Hills study area showed a significant probability that on average, 76% percent of all human-caused fires will start on slopes that are $\leq$ 10% given a random sample of the HWC data from 2005-2007. The slope results make sense from an intuitive fire investigation standpoint, in that humans tend to domicile, work, and recreate in areas that are more flat than steep.

Lightning-caused fires in the applied research project study area showed no significant probabilities of causing wildfire occurrence with regard to an increase or decrease in elevation, slope, aspect or road-distance measurements. The frequency appeared to be stochastic in nature across the range for these explanatory variables. However, the Kruskal-Wallis tests did show significant results with regard to differences in elevation in response to ha-burned for both the Lightning and the HWC datasets. Given that the elevation treatment zones were kept as balanced as possible in the K-W test to ensure homogeneity of variance, then it could be reasonably assumed that one of the treatment quartiles contains a range of the response variable of ha-burned that is derived from a different population than the other quartiles. The post-hoc Mann-Whitney tests showed that this difference was found between the lower quartile (0 – 25%) vs. the highest quartile treatment (75 – 100%), meaning that significant differences in mean, median and variation of ha-burned can be expected from these two different treatment (elevation) zones, given that they come from two different populations. It appears that for both lightning and human-caused starts more hectares burned in lower elevations of the Black Hills than the higher elevations over the time period of the datasets.

This finding shows an opposite relationship than what Maingi and Henry found in Kentucky (2007) where cooler and wetter weather conditions are found in the lower elevation valleys of the DBNF and it will get hotter and drier as elevation rises in that study area during fire season. These contrasts in the Black Hills and the Appalachian mountains can be attributed to the difference in fire regimes in the western part of North America as opposed to the eastern part of the United States. In the Black Hills, under the drier conditions of lower humidity and unstable atmosphere, the cooler air temperatures and resulting increase in humidity as elevation increases will cause changes in fuels and weather variables that will decrease the spread potential of the flaming front of a wildfire under normal conditions.

As it was pointed out on a study in the northern Rocky Mountains about the relationship of climate change and wildfire occurrence, that a potential barometer to gauge climate change is an increase in ha-burned at higher elevation zones of western forests in the U.S. (Westerling, et al., 2006). Tree-ring studies conducted in Colorado showed that from 1650 to 1920, lower elevation zones had higher frequency of wildfire and more ha-burned than higher elevation zones (Veblen, et al, 2000). Both frequency and ha-burned in the Black Hills as observed in this study area is congruent with the long-term trends of elevation and wildfire occurrence observed in Colorado, Idaho, and Montana.

Comparing this study to the research done on the Mark Twain National Forest in Missouri showed that elevation played a more important role in determining ha-burned than aspect or slope in wildfire occurrence, but all three geospatial variables had low importance overall as compared to ecological and anthropogenic factors (Brosofske, et. al, 2007) . The statistical analysis showing only significant results for elevation and ha-burned as compared to slope or aspect in this study showed the same pattern as the Mark Twain National Forest. Studies dealing with slope, aspect and elevation were not conducted in the research done in Michigan, Wisconsin, and Minnesota, because of the lack of topographical variation in those study areas (Cardille, et al., 2001). The research completed on wildfire occurrence in California (Syphard, et al., 2007) and Florida (Prestemon, et al., 2002) focused primarily on anthropogenic factors.

The Road-Dist. variable was observed in this study to be not independent of wildfire frequency. An analysis of the contingency tables showed significant probabilities associated with frequency of wildfire within 200 m of a road feature on the map generated from the Graphical Locator. In the study years of 2005-2007, for 72% of the time, the x-y coordinates for fire location from the HWC dataset could be expected to be mapped out within 200 m of a road. As discussed earlier, lightning fires displayed a more random pattern when compared to road distance and the H_0 of no relationship in the case of lightning-caused fires and road-distance was not rejected by statistical analysis.

The observation in this study that human-caused wildfires occur more frequently closer to road features is similar to many research findings in the last 20 years. The studies discussed earlier in Kentucky, Missouri, Missouri and the upper Midwest all show correlations between wildfire occurrence and WUI features, such as housing density, population density, and roads even when researched at coarser scales than this project. In addition, the national study by Stephens and Ruth in 2005 titled *Federal Forest Fire Policy in the United States* observed on a national scale the significant frequency of human-caused wildfire next to transportation corridors. Prestemon, et al. in their 2002 Florida study of human factors and wildfire risks did not document any measurable influence of WUI interface to wildfire risk, but these researchers did call for a finer scale analysis of WUI features such as roads and settlement patterns before settling the question. The California study by Syphard, et al. (2007) did not report any measurable significance with the variable of average distance to road in explaining fire frequency, but that may have been hard to observe in their large scale analysis due to the fact that the study area involved 54 of the 58 counties in California.

The results of this study showed that when comparing the average distance from the road for a fire location does vary in this study area depending on what started the wildfire. Significant differences were found between the means of road distance variables between the HWC (≈ 200 m) and Lightning (≈ 500 m) datasets. In addition, comparing mean slope measurements, a significant difference was found between Lightning-Slope (≈ 11%) and HWC-Slope (≈ 7.4%), although comparison of the means is not possible through the Mann-Whitney test, the test does show that these mean values for both slope and road-distance were produced from different populations during the 2005-2007 calendar years.

During the study period, HWC fires started more frequently and were closer to roads than lightning-caused fires in this study area, and it appeared that road features allowed for more efficient access for fire suppression resources to quickly suppress fires and keep the average fire HWC wildfire size small ($X \approx 7$ ha) as compared to lightning wildfires ($X \approx 40$ ha). Factoring in the national initial attack efficiency ratio (97%-99%) observed by Stephens and Ruth (2005) and removing 3 % of the largest fires from both the HWC and Lightning datasets, then recalculating the means for both Lightning ($X \approx 0.7$) and HWC ($X \approx 0.7$), it can be seen that road-distance may play an important part in escaped-fire analysis as it applies to lightning-caused fires. That same attribute for suppression efficiency can be referenced in the California study *Human Influence on California Fire Regimes* (Syphard, et al., 2007). In that study, it was noted that a positive relationship existed between road-distance and ha-burned when grasslands and shrub land components were incorporated in the regression models. However, the K-W tests of ha-burned with HWC fires in the Black Hills study did not produce any results of any significance ($H = 3.241$, $df = 3$, $p = 0.36$) that would reject the H_0 of no relationship between ha-burned and the HWC Road-Dist. variable.

Statistical Analysis

The research question of the proper statistical analysis to use in this study was important given the hypotheses tested and the type of datasets that were used. The underlying basis of the study is that of applied research that can be implemented by agency firefighters in the course of their work. Therefore, Haines, et al., (1974) in the Pennsylvania study of wildfire occurrence, set forth a principle of applied research for wildfire data analysis in that just enough analysis should be provided for straightforward interpretation in the field, even though many different types of complex analysis could be conducted. By applying the Causal-Comparative research method outlined by Gay and Airasion (2002) and by supplying enough descriptive and inferential statistical analysis, the basis of applied research can be carried forward for application in the field.

In addition, Anderson and Burnham, in their book *Model Selection and Multi-Modal Inference: A practical information-theoretic approach* cautioned researchers not to engage in data dredging but use critical thinking skills and to understand "If a particular model (parameterization) does not make biological sense, this is reason to exclude it from the set of candidate models, particularly in the case where causation is of interest." (Anderson & Burnham, 2002, p. 17). Therefore two-pronged procedure (parametric vs. nonparametric) in this study were designed to avoid data-mining or egregious post-hoc analysis by testing the data after the hypothesis were stated.

In this study, both the Lightning and the HWC datasets could not be transformed into a normal distribution, so the nonparametric Kruskal-Wallis and Mann-Whitney tests were chosen for analysis. This right-skewed datasets used in this study presented the same problems for the assumption of normality that have faced other researchers of wildfire occurrence data, even when these researchers had the ability to work with large datasets > 10,000 records (Brosofske, et al., 2007), (Stephens, Forest fire causes and extent on United States Forest Service lands, 2005) and (Syphard, et al., 2007). The *Fire Data Analysis Handbook, 2nd Edition* (FEMA, 2004) does mention skewed data and the Central Limit Theorem for making the assumption of normality if $n > 20$ for working with inferential statistics and fire data, but it appears that academic researchers working with wildfire occurrence data make no such assumptions of normality.

Internet-based statistical analysis worked well with the Graphical Locater program for creating and analyzing the dataset of geospatial variables. This approach would serve firefighters in small fire departments that do not have the budget to purchase and maintain expensive and sophisticated software packages for GIS and statistical work that are used in the academic setting. The internet-based statistical and GIS applications that were used for this project are available for non-commercial academic use and more than sufficed for the basic statistical analysis used for this study. (Environmental Statistics Group, Montana State University, 2005), (Kirkman, 1996), (Lowry, 2008), (McDonald, 2007), and (Wessa, 2008).

Organizational Impact

Implicit in this study is the impact that the analysis of geospatial variables can bring to the day to day operations of State of South Dakota, Division of Wildland Fire Suppression. Users of this research will be able to: (a) target prevention efforts; (b) provide focus for direction in the investigation of human-caused wildfire; (c) provide more information for fire suppression strategy on days with multiple fires; (d) provide descriptive statistics for agency budget presentations, inquiries from the Governor's Office or Legislative staff, and requests for news media releases; (e) provide more direction in redirecting and allocating expensive and scarce wildfire suppression resources such as aircraft and handcrews; and (f) providing initial attack incident commanders on wildfires with more information to size up and deploy resources on wildfires.

Recommendations

This study described the relationships between geospatial variables and wildfire occurrence – specifically area-burned (ha-burned) and frequency. Also researched was the correct statistical analysis needed to correctly infer the results from the procedures performed.

In the case of where statistical relationships were observed between a geospatial variable and a response variable, the following recommendations are made based on the acceptance of the $H_{1:}$ that there is a relationship between the geospatial variable and ha-burned or wildfire frequency:

1. Those periods where multiple lightning-caused fires are ignited, or wind-event days were multiple debris-burns escape control, priority suppression efforts should be targeted at lower-elevation fires in the Black Hills, after such factors as life safety and critical values at risk are taken into account, in order to reduce area burned.
2. In order to reduce the seasonal occurrence of spring and summer human-caused wildfires, prevention and patrol efforts should be targeted in those areas where slope is < 10% and elevation is < 1400 m and this effort should probably focus in the Rapid City Field Office area, given the high frequency of human-caused fires in that district.
3. From a fire investigation standpoint, most human caused fires (72%) start within 200 m of a road and the data further showed that all suspected arson fires started within

4. Arson task force investigators assigned to road patrols should continue to maintain surveillance on road systems.
5. For 2005-2008, results indicated that all fires started by equipment use started on slopes less than 10% and within 135 m of a road. Careful consideration should be given to those fires in the future where cause cannot obviously be determined that if they start near roads and in areas where slope is < 10 %, that equipment use be suspected as highly probable cause (e.g. spark arrestor failure, etc).
6. Given the difference between the average distances from road for lightning-caused fires increases from North to South in the Black Hills; then maintaining both the heavy-lift contract helicopter and the Single Engine Air-Tanker (SEAT) Base in the southern Black Hills for initial attack in the Hot Springs and Custer State Park Districts is sound strategy.
7. That WFS continue to make investments into its aviation program of contract SEAT's and continue the cost-share contributions to the interagency helicopter program given the average distances from roads that lightning fires can start from in the study area.

Another set of recommendations can be made by reviewing the fact that the H_0: independence (no relationship) between the geospatial variable and ha-burned or wildfire frequency was not rejected in the other statistical tests. These two recommendations are as follows:

1. That during times of expected lightning storms, move up and cover procedures take place by ensuring that both crew and engine resources have good and fast access to roads to be able to move quickly to new fire starts in any direction of their location, given the stochastic pattern of lighting-caused fires across the landscape.
2. That aviation personnel working on lightning detection flights for WFS be briefed on the need to observe on both sides of ridge-top features when conducting detection flights in the Black Hills given that it appears that lightning started fires are discovered just as frequently on north -facing aspects as south -facing aspects and flat surfaces.

These recommendations will be discussed internal to the agency at staff meetings and shared in interagency manner at seasonal fire meetings for discussion and future implementation. As more data is gathered in future fire seasons via handheld GPS units with accurate x-y coordinates specific to the point of origin, a larger sample will be available for advanced research using sophisticated statistical analysis under the direction of more skilled researchers. Hopefully new interactions will be discovered between fuel, weather and topography that will improve firefighter and public safety and increase the knowledge of fire behavior.

References

Anderson, D., & Burnham, K. (2002). *Model selection and multi-modal inference: A practical information-theoretic approach.* New York: Springer-Verlag.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* Belmont, CA: Wadsworth International Group.

Brosofske, K. D., Cleland, D. T., & Saunders, S. (2007). Factors influencing modern wildfire occurence in the Mark Twain National Forest, Missouri. *Southern Journal of Applied Forestry* , *31* (2), 73-84.

Brown, P. M., & Sieg, C. H. (1996). Fire history in interior ponderosa pine communities of the Black Hills, South Dakota, USA. *International Journal of Wildland Fire* , *6(3)*, 97-105.

Cardille, J., Ventura, S., & Turner, M. (2001). Environmental and social factors influencing wildfires in the upper midwest, USA. *Ecological Applications* , *11*(1), 111-127.

De'ath, G., & Fabricius, K. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* , *81*(11), 3178-3192.

Donoghue, L., & Main, W. (1985). Some factors influencing wildfire occurrence and measurement of fire prevention effectiveness. *Journal of Environmental Management* , *20* (1)1, 87-96.

Environmental Statistics Group, Montana State University. (2005, June 23). *Graphical Locater Home Page.* Retrieved October 16, 2007, from Montana State University, Environmental Statistics Group: http://www.esg.montana.edu/gl/

ESRI. (2004, March 31). *ArcUser Online, Jan-Mar 2004 Newsletter.* Retrieved January 23, 2007, from ESRI, GIS and Mapping Software: http://www.esri.com/news/arcuser/0104/rec-gps.html

FEMA. (2004). *Fire Data Analysis Handbook, Second Edition.* U.S. Dept. of Homeland Security, FEMA.

Furman, R. (1978). Wildfire zones on a mountain ridge. *Annals of the Association of American Geographers* , *68*(1), 89-94.

Gay, L. R., & Airasian, P. (2003). *Educational research -- Competencies for analysis and applications, 7th ed.* Upper Saddle River, NJ: Merrill Prentice Hall.

Haines, D. A., Main, W. A., & McNamara, E. F. (1978). *Forest fires in Pennsylvania. Research Paper NC-158.* St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station.

Kirkman, T. (1996). *Statistics to use -- Index*. Retrieved March 6, 2008, from Statistics to Use: www.physics.csbsju.edu/stats/

Lewis, R. J. (2000). *An introduction to classification and regression tree (CART) analysis.* Retrieved January 17, 2008, from Society for Academic Emergency Medicine: http://www.saem.org/download/lewis1.pdf

Lowry, R. (2008). *The Fischer exact probability test.* Retrieved January 18, 2008, from Concepts and Applications of Inferential Statistics: http://faculty.vassar.edu/lowry/webtext.html

Maingi, J., & Henry, M. (2007). Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. *International Journal of Wildland Fire* , 16, 23-33.

Marchand, K. J.; (2008) *Black Hills National Forest – State of South Dakota 2008 (display map)* [graphics file].Custer SD: Black Hills National Forest Supervisors Office, GIS Group

McDonald, J. H. (2007, January 1). *Handbook of biological statistics -- Small numbers in Chi-Square and G-tests.* Retrieved February 19th, 2008, from Handbook of Biological Statistics: http://udel.edu/~mcdonald/statsmall.html

McDonald, J. H. (2008, January 8th). *Handbook of biological statistics: Normality*. Retrieved February 24, 2008, from Handbook of Biological Statistics: http://udel.edu/~mcdonald/statnormal.html

Mercer, D., & Prestemon, J. P. (2005). Comparing production function models for wildfire risk analysis in the wildland-urban interface. *Forest Policy and Economics ,7*, 782-795.

National Interagency Fire Center, B. I. (2008, January 21). *Incident Management Situation Report, July 12, 2007*. Retrieved January 21, 2008, from National Interagency Fire Center, Boise Idaho: http://www.nifc.gov/nicc/IMSR/2007/20070712IMSR.pdf

Prestemon, J., Pye, J., Butry, D., Holmes, T., & Mercer, D. (2002). Understanding broadscale wildfire risks in a human-dominated landscape. *Forest Science* , 48(4), 685-693.

Pyne, S., Andrews, P., & and Laven, R. (1996). *Introduction to wildland fire.* New York: John Wiley and Sons.

Stephens, S. L. (2005). Forest fire causes and extent on United States Forest Service lands. *International Journal of Wildland Fire*, *14*, 213-222.

Stephens, S., & Ruth, L. (2005). Federal forest fire policy in the United States. *Ecological Applications* , *15*, 532-542.

Syphard, A., Radeloff, V., J.E., K., Hawbaker, T., Clayton, M., Stewart, S., et al. (2007). Human influence on California fire regimes. *Ecological Applications* , *17*(5) ,1388-1402.

United States Forest Service, Rocky Mountain Region *Alabaugh Canyon Fire entrapment and shelter deployment: Final accident investigation report, July 8th 2007*, Retrieved April 1, 2008, from Wildfire Lessons Learned: http://iirdb.wildfirelessons.net/main/DisplayReport.aspx?ID=914

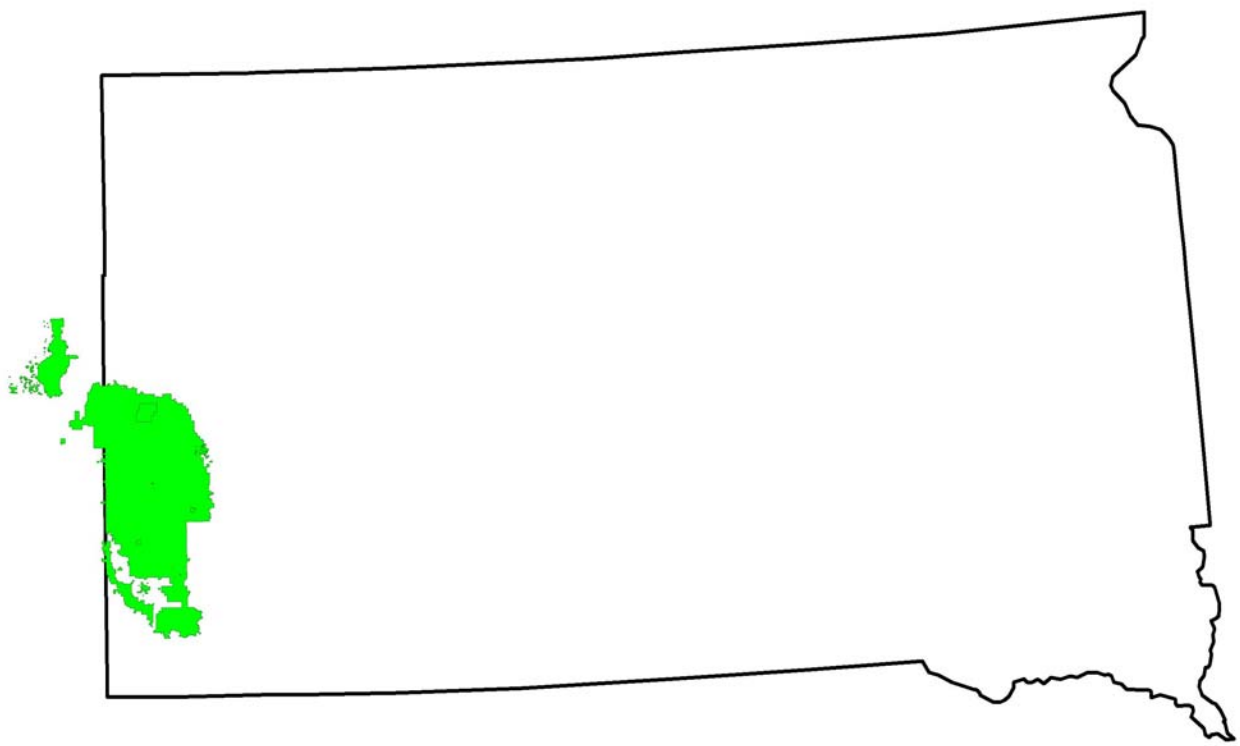
Veblen, T. T., Kitzberger, T., & Donnegan, J. (2000). Climatic and human influences on fire regimes in ponderosa pine forests in the Colorado front range. *Ecological Applications* , *10*, (4), 1178-1195.

Wessa, P. (2008, January). *Free statistics and forecasting software*. Retrieved February 23, 2008, from Office for Research Development and Education: www.wessa.net

Westerling, A., Hidalgo, H., Cayan, D., & Swetnam, T. (2006). Warming and earlier spring increase western U.S. forest wildfire activity. *Science* , *313*, 5789, 940-943.

Figure Caption

Figure 1. Outlined border of the State of South Dakota, with the general location of the study area displayed in the left-hand side of the figure. The shaded area is the Black Hills National Forest (BHNF) that is located in both SD and WY. The study area comprises those private and state-owned lands adjacent to and within the BHNF in SD. (Marchand, 2008)



Appendix A

Table A1 displays the results of tests of the sample distribution of the response variable ha-burned and when it was transformed by square root and a log + 1 transformation. The spreadsheet provided by McDonald calculated the transformation values (McDonald, 2008). The Kolmogorov-Smirnov calculator was provided by the Physics Department at the College of St. Benedict and St. John's University in Minnesota (Kirkman, 1996) and it was used to produce the $X_{\text{normal dist}}$ and the probability of a normal sample distribution . The skewness rating of the variable was provided by the spreadsheet calculator, any range outside of -.8 to .8, indicates a skewed distribution.

Table A1

Analysis of Sample Distribution

Variable	X	$X_{\text{normal dist.}}$	p of norm. dist.	skew
HWC ha-burned	7.04	215	0.00	12.9
Light ha-burned	39.80	910	0.00	11.0
HWC $\sqrt{}$ha-burned	0.77	6.5	0.00	10.1
Light $\sqrt{}$ha-burned	1.35	14.2	0.00	8.3
HWC $_{\log+1}$ ha-burned	0.16	.069	0.00	4.2
Light $_{\log+1}$ ha-burned	0.18	0.92	0.00	4.6

Appendix B

Exact Table Analysis

The following tables results are given with the probability assigned from the Exact Table analysis by an internet-based calculator provided by the Physics Department at the College of St. Benedict and St. John's University in Minnesota (Kirkman, 1996). The term "*total probability*" in this analysis means that the table result you are looking at (the given table) and the all the other tables that were calculated, if any of those tables have probabilities less than or equal to the given table, and the sum of those probabilities is small, meaning that we do not expect to see those types of table very often, then the H_0 may be rejected and the H_1 alternative be accepted that the variables are not independent of each other in the design of the table. Conversely, a high probability number > 0. 05%, indicates that the given table that was created from the ha-burned frequency count, is not unusual at all, when compared to other tables that could have been calculated from same design of rows and columns and the same data, and there are many tables similar to it. It is important to keep the design of the tables the same throughout the test, in order to test the H_0 in a scientific manner. Continually resorting the treatments in order to find a relationship in order to reject the H_0 may result in finding relationships that do not make sense or are spurious.

Table B1

Design of Tables and Probabilities Associated with R x C Table Contingency Analysis

Dataset	Design	*R x C*	Values
Lightning	Elevation x Frequency/ha	4 x 2	$p = 0.36$

Appendix B

Table B1, cont.

Design of Tables and Probabilities Associated with R x C Table Contingency Analysis

Dataset	Design	*R x C*	Values
Lightning	Elevation x Frequency/month	4 x 2	$p = 0.75$
Lightning	Slope x Frequency/ha	3 x 2	$p = 0.83$
Lightning	Slope x Frequency/month	3 x 2	$p = 0.59$
Lightning	Aspect x Frequency/ha	3 x 2	$p = 0.53$
Lightning	Aspect x Fire Frequency/month	3 x 2	$p = 0.93$
Lightning	Road-Dist. x Fire Frequency/ha	4 x 2	$p = 0.07$
Lightning	Road Dist. x District	2 x 2	$p = 0.07$
HWC	Elevation x Fire Frequency/ha	4 x 2	$p = 0.03$*
HWC	Elevation x Fire Frequency/season	4 x 4	$p = 0.00$**†
HWC	Elevation x HWC cause	4 x 3	$p = 0.00$**
HWC	Slope x Fire Frequency/ha	3 x 2	$p = 1.00$
HWC	Slope x Fire Frequency/Season	3 x 4	$p = 0.00$**
HWC	Slope x HWC cause	3 x 3	$p = 0.00$**

Appendix B

Table B1, cont.

Design of Tables and Probabilities Associated with R x C Table Contingency Analysis

Dataset	Design	*R x C*	Values
HWC	Slope x HWC cause	3 x 6	p = 0.00**††
HWC	Aspect x Fire Frequency/ha	3 x 2	p = 1.00
HWC	Aspect x Fire Frequency/season	3 x 4	p = 0.67
HWC	Aspect x HWC cause	3 x 3	p = 0.52
HWC	Road-Dist. x Fire Frequency/ha	4 x 2	p = 0.00**
HWC	Road-Dist. x Fire Frequency/ha	2 x 2	p = 0.00**
HWC	Road Dist. x District	2 x 4	p = 0.21
HWC	Road Dist. x Fire Frequency/season	2 x 4	p = 0.48
HWC	Road Distance x HWC cause	2 x 6	p = 0.34††

Note. * p < .05, ** p > .01,

† indicates chi-square test was conducted instead of exact test.

†† indicates that 6 fire causes frequencies were tested: undetermined, debris, power lines, equipment, campfires and miscellaneous.

Appendix C

The following tables contain descriptive statistics from the results section that are classified by WFS District for CY 2005-2007.

Table C1

Sample Means (X) of Explanatory Variables Classified by District

Variable	Unit	*X*	Hot Springs	Custer State Park	Rapid City	Lead
HWC--Elevation	m	1346	1400	1452	1266	1432
Light - Elevation	m	1321	1238	1460	1220	1406
HWC – Slope	%	7.0	4.0	16.0	8.0	12.0
Light – Slope	%	11.0	8.0	15.0	10.0	13.0
HWC – Road	m	201	227	346	194	163
Light—Road	m	497	572	554	406	305

Table C2

Sample Means (X) of the Response Variable Ha-burned Classified by District

Variable	Unit	*X*	Hot Springs	Custer State Park	Rapid City	Lead
HWC-ha-burned	ha	7.08	1.48	137.7	1.03	1.15
Light-ha-burned	ha	39.8	87.1	1.13	37.5	0.46

Appendix C

Sample Means (X) of the Response Variable Ha-burned Classified by District at 97% Efficiency

Variable	Unit	*X*	Hot Springs	Custer State Park	Rapid City	Lead
HWC-ha-burned	ha	0.72	0.50	3.80	0.48	1.15
Light-ha-burned	ha	0.70	1.45	0.19	0.46	0.46

Table C3

Frequency Occurrence by District, 2005-2007

Dataset	*n*	Hot Springs	Custer State Park	Rapid City	Lead
HWC	171	59	7	78	27
Lightning	157	56	47	35	19

Table C4

Percentage of Occurrence by District, 2005-2007

Dataset	*n*	Hot Springs	Custer State Park	Rapid City	Lead
HWC	171	35%	4%	46%	16%
Lightning	157	36%	30%	23%	13%

Note: due to rounding error, total percentage may not equal 100%

Author Note

I would like to thank my family and fellow co-workers at the South Dakota Division of Wildland Fire Suppression for their patience with me and picking up additional duties while I was working on this project. It is much appreciated.

I also want to thank individuals such as Dr. McDonald of the Univ. of Delaware, Dr. Lowry of Vassar College, NY, Dr. Wessa at Office for Research Development and Education, and the Physics Department at College of St. Benedict and St. John's University in MN, for taking the time to put together such informative websites on statistical analysis, and provide free calculators and software for use for non-commercial purposes. Many fire departments cannot afford expensive and sophisticated statistical software packages for occasional and intermittent use, and the ability to use these at no charge is much appreciated

The Environmental Statistics Group at Montana State University provides the easy to use Graphical Locator program that has served the wildland fire community in South Dakota so well for so many years, and their work in maintaining the database is always appreciated.

Any errors or omissions in this article are mine alone.

Correspondence concerning this article should be addressed to:

James E. Strain Asst Chief – Operations
SD Div of Wildland Fire Suppression
4250 Fire Station Road
Rapid City, SD 57703
e-mail: jim.strain@state.sd.us