

LAB 2 - Wildfire Spread Dynamics

Tanav Thanjavuru, Simran Gill, Ozkan Meric Ozcan

2024-07-23

Contents

1	Introduction	1
2	Description of the Data Source	1
3	Data Wrangling & Operationalization	1
4	Model Specification	2
5	Model Assuptions	3
6	Model Results and Interpretation	3
7	References	4

1 Introduction

Forest fires not only represent a significant threat towards human life and property, but also affect the ecosystem. The northeast region of Portugal in particular has a history of wildfires because more than a third is covered by forests. Adding in the factor of hot and dry summers can potentially increase the risk of wildfires¹. Even though wildfires are a normal part of the renewal process of forests, the forests are getting drier and people are living closer to them which causes an increase in safety issues². This is why it is important for us to understand the behavior of wildfires, in particular the speed at which fire spreads.

With this investigation, we can continue to discuss how wildfires affect the environment along with social-economics and explore opportunities to do our part to put in place appropriate measures of safety. Along with producing strategies to minimize the risk of wildfires. This research will be in the interest of Environmental Scientists and Ecologists along with Fire Management Agencies with the goal of protecting and decreasing any threats towards human life and the ecosystem. To create a starting point for this discussion, this research will investigate the following research question:

How does Fine Fuel Moisture Code (FFMC) help us understand how quick a fire spreads in the northeast region of Portugal?

To measure fire spread, we will use Initial Spread Index (ISI). Both FFMC and ISI are key components of the Canadian Forest Fire Weather Index system which are used to assess fire behavior. To provide context, FFMC a numeric rating which measures the moisture content of forest litter such as mosses and twigs. A higher FFMC value means these fine fuels are drier. ISI is a numeric rating that estimates how quickly a fire will spread after it starts. Exploring the relationship between these two variables can provide valuable insights into what conditions cause rapid wildfire to spread. To answer this research question, we will perform a bivariate analysis by developing a regression models by iteratively applying variable transformation. Then evaluating both the statistical and practical significance of results.

2 Description of the Data Source

The dataset we will be using for this research is from the UCI Machine Learning Repository which is called “Forest Fires” and can be found here: <https://archive.ics.uci.edu/dataset/162/forest+fires>. This data set has 517 instances with 13 feature that consists of climate and physical factors of the Montesinho natural park. The data was collected between January 2000 and December 2003. This can be used to understand forest fire behavior in northeast region of Portugal.

3 Data Wrangling & Operationalization

Given the dataset contains 517 rows, we will divide it into an exploration set and a confirmation set. The exploration set, will be 30% of the dataset (155 rows), which will be used to initially explore and analyze the data. The remaining 70% (362 rows) will form the confirmation set, which will be used to build the final model and interpret the results.

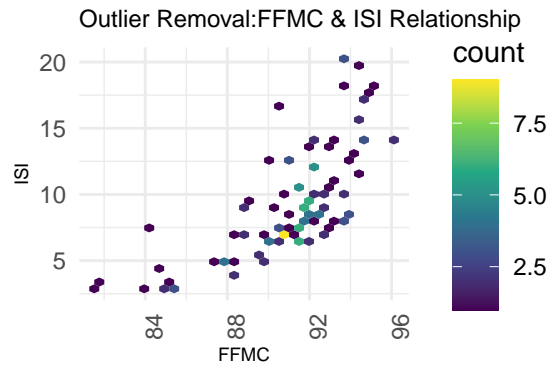
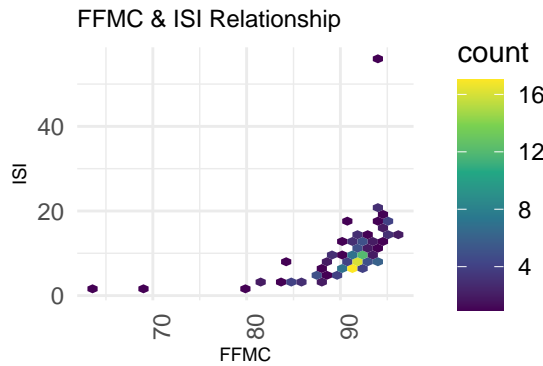
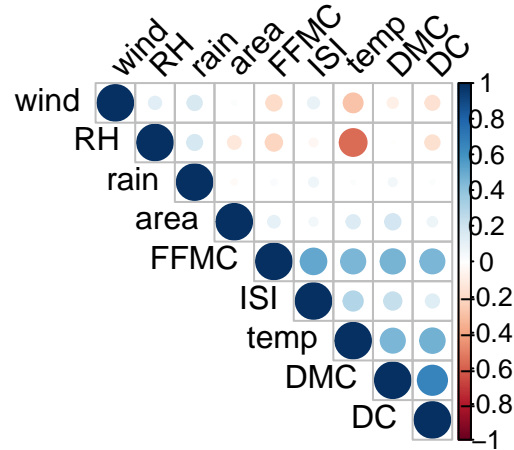
Our research question has two parts, the Fine Fuel Moisture Code (FFMC) and how quickly a fire spreads. To start exploring the data set, we created a correlation heat map to see if FFMC has any correlation with other variables, even though correlation does not mean causation. From the plot, we saw there are 4 variables that have correlation, which are Initial Spread Index (ISI), Temperature, Duff Moisture Code (DMC), and the Drought Code (DC). Because we are interested in the speed of a fire spreading, we created a scatter plot for FFMC and ISI to see if there is a linear relationship. From the scatter plot called “FFMC & ISI Relationship”, we found that as FFMC increases, so does ISI. Since there is a relationship here, and our research question involves fire spread-ability, we decided to select ISI as our dependent variable and FFMC as our independent variable.

From the scatter plot, we noticed there are some outliers, and decided it would be a good idea to remove those using an interquartile range. After removing the outliers, the exploration set has 151 rows and the confirmation set still has 362 rows.

¹The Sunday Times. "Portugal wildfires: what to expect if you're travelling in summer 2024" (2024).

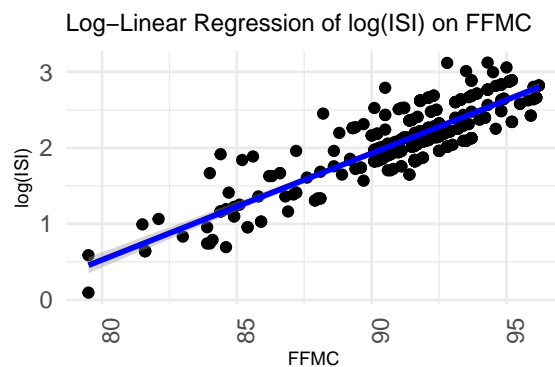
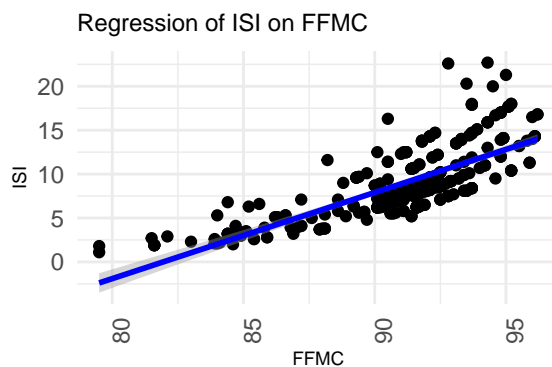
²National Geographic. "How to live with mega-fires? Portugal's feral forests may hold the secret" (2019).

Forest Fire Variables



4 Model Specification

We started with a simple linear regression model to establish a baseline relationship between FFMC and ISI. After plotting the model, we saw a strong linear relationship between the variables. After this we applied log transformation to address skewness and improve model fit, and again saw a good line of best fit. The log transformation helped in normalizing the distribution of the dependent variable, which can lead to a better fit and more reliable statistical inference. We used log-transform for the ISI variable was based on an initial observation of skewness in the data.

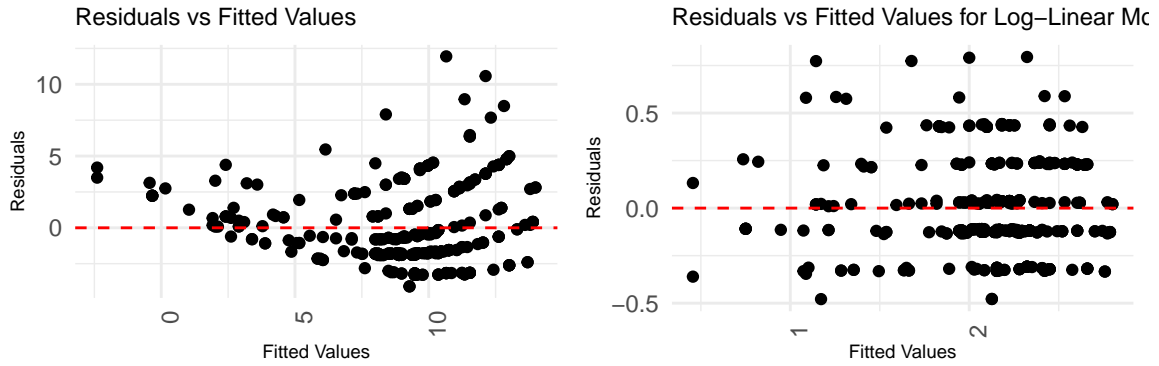


5 Model Assumptions

The first assumption (IID) states that the data points are independent of each other and follow the same probability distribution. For our wildfire dataset, based on the correlation plot, each observation of fire spread in the dataset appears to be independent of the others. The conditions that affect fire spread (temperature, moisture, and fuel type) are assumed to be consistent across observations as well. The second assumption (BLP) states that there is a linear relationship between the independent variable(s) and the dependent variable, and that the linear model is the best predictor of the dependent variable. Based on the FPMC (Fine Fuel Moisture Code) and ISI (Initial Spread Index) plot, The relationship appears to be approximately linear.

6 Model Results and Interpretation

The residuals vs fitted plot will help us judge two key model assumptions, homoscedasticity and non-linearity. We can see in the residuals plot for the simple linear model that the residuals increase with the fitted values, indicating heteroscedasticity and non-linearity. In contrast, the log-linear model's residuals appear more randomly distributed. The absence of a clear pattern suggests that the log transformation has improved the model fit by stabilizing the variance and improving the linearity.



Our chart below shows that for our log-linear regression our coefficient for FPMC is statistically significant. Overall our model has an adjusted R^2 of 0.74. This high adjusted R^2 value shows that FPMC has strong explanatory power in our model and is a good predictor of ISI.

Table 1: Log-Linear Regression Model Results

	Dependent variable:
	$\log(\text{ISI})$
FFMC	0.14*** (0.004)
Constant	-10.71*** (0.40)
Observations	355
R^2	0.74
Adjusted R^2	0.74
Residual Std. Error	0.25 (df = 353)
F Statistic	1,022.88*** (df = 1; 353)
Note: *p<0.1; **p<0.05; ***p<0.01	

7 References

Howard, B. C. (2023, October 4). How to live with mega-fires: Portugal's forests may hold the secret. National Geographic. <https://www.nationalgeographic.com/science/article/how-to-live-with-mega-fires-portugal-forests-may-hold-secret>

Sampson, L. (2024, July 18). Where are the Portugal wildfires and is it safe to travel? The Times. <https://www.thetimes.com/travel/advice/where-are-the-portugal-wildfires-is-it-safe-to-travel-ttflwnmnr>