Forest Fires A/B Test Analysis

Bonus Assignment - DS: Statistics and Probability

College: Cornerstone International Community College of Canada - CICCC

Student: Amir Lima Oliveira

Date: May 20th, 2025

1. Objective

To conduct a complete A/B (two-sample) hypothesis test using a real-world environmental dataset. We aim to determine whether the presence of rain significantly affects the burned forest area.

2. Dataset Information

Dataset title: Forest Fires
Source: UCI Machine Learning Repository
Link: https://archive.ics.uci.edu/dataset/162/forest+fires
Description: The dataset contains meteorological data and fire area measurements from Montesinho Natural Park in Portugal.

3. Exploratory Data Analysis (EDA)

- Descriptive statistics: Mean, median, standard deviation computed.
- Distribution: Right-skewed burned area data.
- No missing values or duplicates were found after cleaning.
- Variables such as FFMC, ISI, and rain were visually inspected.

4. Hypothesis Formulation

H0: There is no significant difference in the burned area between days with high and low FFMC.
H1: There is a significant difference in the burned area between days with high and low FFMC.
Significance Level (alpha): 0.05

5. Test Execution

- Group A: Low FFMC days
- Group B: High FFMC days
- Statistical test: Welch's t-test (assumes unequal variances)
- Assumptions: Continuous data, independent groups, unknown population std dev
- Levene's test was used to check for equal variances.

6. Results & Conclusion

Test statistic = 1.134
P-value = 0.2575
95% Confidence Interval: (-4.3507, 16.2043)
Conclusion: Fail to reject the null hypothesis.
Even though FFMC is a relevant fire index, it was not statistically significant in this sample to explain the burned area variation.

Forest Fires A/B Test Analysis

7. Brief Discussion

- A larger dataset from all of Portugal could increase the statistical power.
- Burned area data is skewed, which can violate parametric test assumptions.
- Future work could include log-transforming area or using non-parametric tests like Mann-Whitney.