Forest4Life: Fighting fires with statistical inference

Lab section: C4
Group No# 3
Group Name: Forest4Life
Name of all the team members:

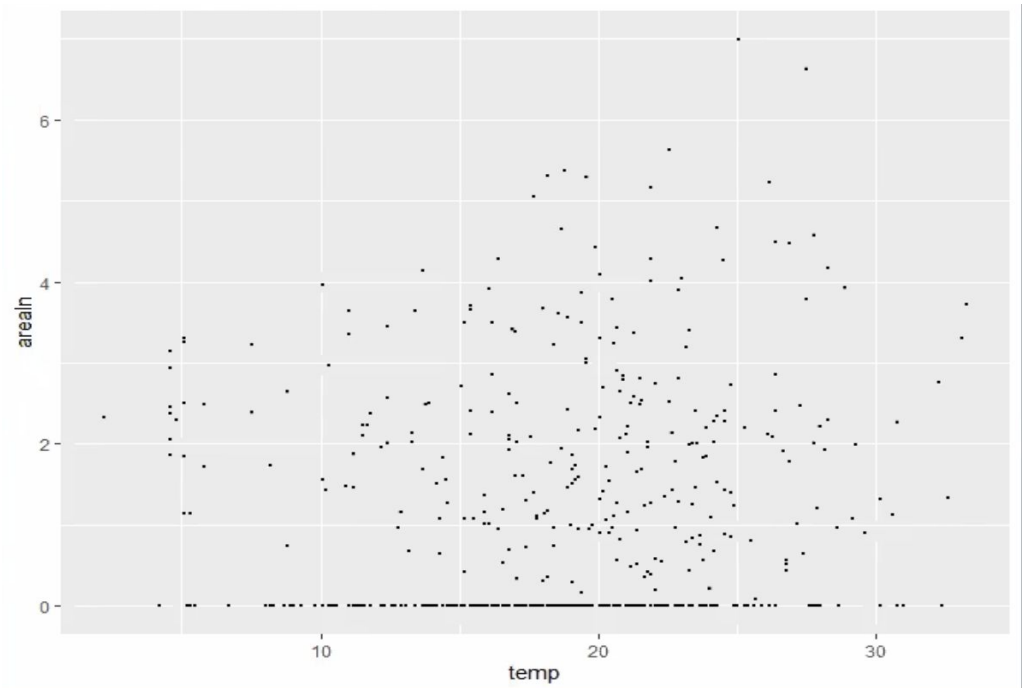
1. **Murtaza Shoeb Moiyadi**
2. **Daniel Cunha**
3. **Xuan Wu**
4. **Naznin Sultana**

1. Choose the response variable (Y) and one covariate (X). Please put some though for your response and covariate variable selection.

   Covariate (X): Temperature
   Response variable (Y): Area

When we think of fires, we think of summer weather: hot, arid, and a lack of rain. In narrowing our list of potential covariates to perform analysis on just one, we settled on temperature because we wanted to see the impact of the seasons on fires. Our hypothesis was that the middle of the summer produced the most and largest fires, which would coincide with high temperatures.

2. Plot Y vs X (i.e. a scatterplot) from the data.



3. Perform OLS using R on your response and covariate.

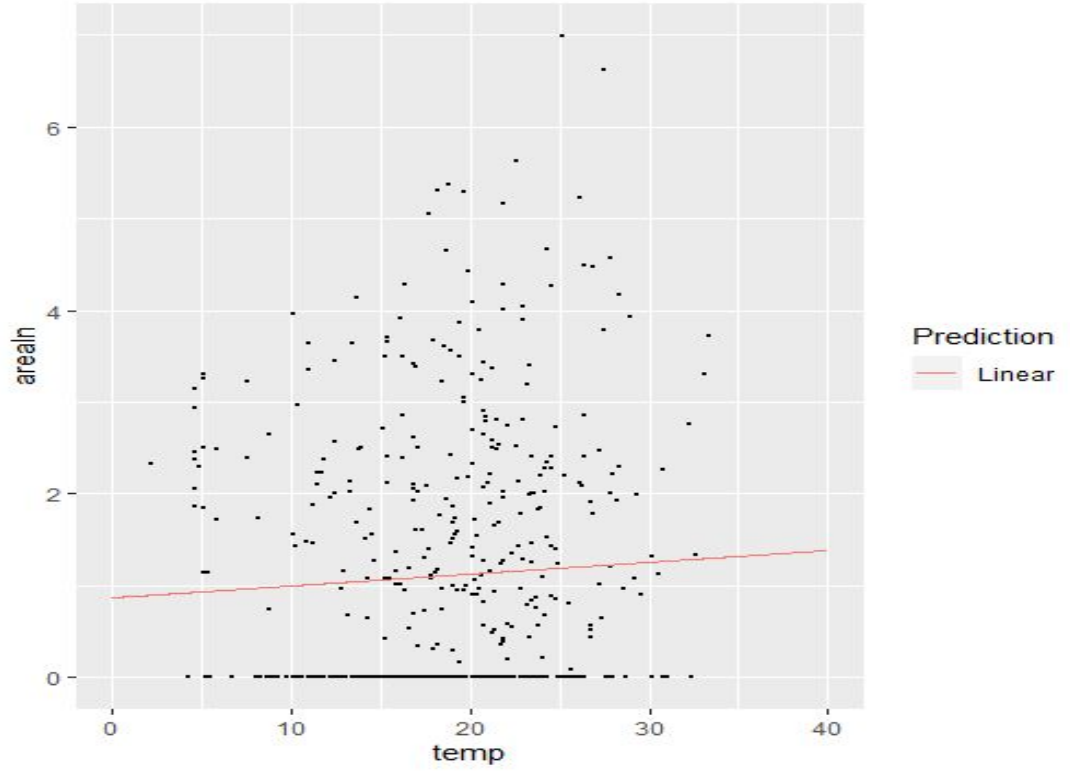
4. Please submit the output from R of OLS (coefficient estimate, t values, p values) and interpret the results.

```
Call:
lm(formula = arealn ~ temp)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2851 -1.1034 -0.7298  0.9278  5.8046

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.86771    0.20940   4.144 3.99e-05 ***
temp         0.01288    0.01060   1.216    0.225
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.398 on 515 degrees of freedom
Multiple R-squared:  0.002861,  Adjusted R-squared:  0.0009246
F-statistic: 1.478 on 1 and 515 DF,  p-value: 0.2247
```

These results suggest there is not enough evidence to reject the null hypothesis that

temperature does not have an effect on area. This is reflected in the p-value 0.225 and that we require a confidence coefficient of at least .95.

5. Plot also Y vs X for your choice of data set and overlay on your plot the linear regression fit obtained from R. Is this a good model ?



We see that there is no relationship between temperature and area in our current model, even after applying a logarithmic transformation on the area to account for the zero values. This is not a good model, which shows us we need to either choose another covariate or find a better way to account for the skewed distribution in the area.

This is the plot of transformed area (ln(area+1)) vs temperature along with the fitted linear regression line for least squares estimates of linear regression model. The predicted line indicates that this linear regression with our selected covariate temperature (temp) with transformed dependent variable “arealn” is not well fitted. So this is not a good model. So we may need to think about whether or not the relationship between the dependent (Y) and independent (X) is linear.