**Question 1a (file: Question 1a output.SPSS)**

For Normality test, both Kolmogorov-Smirnov and Shapiro-Wilk show that all formulations are non-parametric data since p-value < 0.05.

Next, Since all formulations are non-parametric there are meaningless to compare in mean. By comparing the median, we may raise the concern of whether significance difference between two or more than two formulations which have a close median. For example, formulation A (1.5177) versus D (1.36) versus C (3.48), B (13.3) versus G (8.600), and F (0.5550) versus H (0.0000) versus I (0.0000). Other than that, E (72.79) have an obviously significance difference from others therefore we can choose to ignore the significance test for this.

Last, the correlation test performed to help the company to identify if there have any formulations have a positive or negative relationship to each other, if p-value < 0.05 then meaning both is correlated. For example, there is an outstanding data performed in between formulation A and formulation G where both are strong positively correlated of 0.704 to each other highly significance. This might because of some external elements made both formulations burn patterns data direction moving together up and down.

**Question 1b (file: Question 1b output.SPSS)**

Upon observing the distribution through histograms and boxplots for each formulation, we noticed burn patterns’ distribution. Formulations F, I, and H exhibit right-skewed distributions with similar curve shapes, indicating that the majority of the burn pattern data cluster at lower values. While, formulations G and D also display right-skewed distributions, but their majority values cluster at higher levels compared to formulations F, I, and H. On the other hand, formulations A, B, C, and E are left-skewed, with the majority of the burn pattern data distributed at higher values. Notably, formulation C its data also cluster significantly at lower values.

**Question 1c (file: Question 1c output.SPSS)**

Based on the output from question 1a and 1b, we are going to clustering as table below together with test conclusions.

|  |  |
| --- | --- |
| **Formulations clustering test** | **Conclusion** |
| A, C and D | There is a significant difference between formulations A, C, and D. |
| B and G | There is a significant difference between formulations B and G. |
| F, H and I | There is an *insignificant* difference between formulations H and I. |
| E | Obviously significance difference with other formulations |

Other than formulations H and I, all the formulation are distinctive number in the dataset.

**Question 2**

By utilize the R programming, we observed SoilMoisture and Precipitation are the variables effecting FFB Yield. After excluding insignificance variables, we have the final regression model as below:

FFB Yield = 1.9318334 – 0.0011517 SoilM + 0.0014715 Precipit + residual

When Soil Moisture increases by 1000 millimetres in the top 0 – 1.6 meters of soil, the FFB Yield decreases by 1.15 Tonnes/Hectare, holding other variables constant. However, the relationship between Soil Moisture and FFB Yield is not very strong, as indicated by the insignificant Pearson correlation result.

Conversely, when accumulative precipitation in a month increases by 1000 millimetres, the FFB Yield increases by 1.4715 Tonnes/Hectare. There might be other potential external factors to explore, such as fertilization.

**Question 3**

Kindly refer to ‘Question 3’ file format PyCharm/Python