**Essentials of Statistics and Math for Data Science - Project 3**

**Part A**

**Six different machines are being considered for use in manufacturing rubber seals. The machines are being compared with respect to the tensile strength of the product. A random sample of four seals from each machine is used to determine whether the mean tensile strength varies from machine to machine. In the Data.xlsx (Sheet Part A) file you find the tensile-strength measurements in kilograms per square centimeter**

**Perform the analysis of variance at the 0.05 level of significance and indicate whether or not the mean tensile strengths differ significantly for the six machines.**

To analyze the variance at the 0.05 level of significance (which corresponds to a 95% confidence level), we can conduct a one-way analysis of variance (ANOVA) test. Here we use the single-factor ANOVA test

Anova: Single Factor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| Machine 1 | 4 | 70 | 17.5 | 1.666667 |
| Machine 2 | 4 | 68 | 17 | 3.333333 |
| Machine 3 | 4 | 73 | 18.25 | 2.916667 |
| Machine 4 | 4 | 72 | 18 | 6.666667 |
| Machine 5 | 4 | 75 | 18.75 | 2.916667 |
| Machine 6 | 4 | 72 | 18 | 2.666667 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 7.333333 | 5 | 1.466667 | 0.436364 | 0.817329 | 2.772853153 |
| Within Groups | 60.5 | 18 | 3.361111 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 67.83333 | 23 |  |  |  |  |

The p-value associated with the F-statistic is 0.817, which is greater than the significance level of 0.05 (or 5%).

The critical F-value at a significance level of 0.05 is approximately 2.773

**Since the p-value is greater than 0.05, we fail to reject the null hypothesis. So we accept the null Hypothesis**

This means that there is not enough evidence to suggest that the average tensile strength for the six machines is significantly different.

In conclusion, based on the ANOVA test results, there is no significant difference in the average tensile strength among the six machines.

**Part B**

**Please refer to the file Data.xlsx (Sheet Part A) for this part.**

**A study measured the sorption (either absorption or adsorption) rates of three different types of organic chemical solvents. These solvents are used to clean industrial fabricated-metal parts and are potentially hazardous waste. Independent samples from each type of solvent were tested, and their sorption rates were recorded as a mole percentage.  Is there a significant difference in the mean sorption rates for the three solvents? Use a P-value for your conclusions. Which solvent would you use?**

ANOVA: Single Factor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| Solvent 1 | 9 | 8.48 | 0.942222222 | 0.028319444 |
| Solvent 2 | 8 | 8.05 | 1.00625 | 0.160769643 |
| Solvent 3 | 15 | 4.95 | 0.33 | 0.0431 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 3.305406944 | 2 | 1.652703472 | 24.5115048 | 5.8552E-07 | 3.327654499 |
| Within Groups | 1.955343056 | 29 | 0.067425623 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 5.26075 | 31 |  |  |  |  |

Since the p-value (5.8552E-07) is less than the significance level ( 0.05), we reject the null hypothesis.

we reject the null hypothesis, indicating that there is a significant difference in the mean sorption rates for the three solvents. Based on the above results

**To determine which solvent to use: -**

Solvent 1: Average sorption rate = 0.942

Solvent 2: Average sorption rate = 1.006

Solvent 3: Average sorption rate = 0.33

Conclusions-

Solvent 2 has the highest average sorption rate, so we would be the preferred choice is solvent 2