MULTICOLLINEARITY

The below given dataset contains 40 instances of the wine and also its attributes such as Density of wine (in g/c.c), pH level, Alcohol (in percentage) and finally the quality of wine (a score between 0 and 10). We will try to see if there is any relation between the attributes by checking for the Multicollinearity condition.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No. | Quality | Density | pH | Alcohol |
| 1 | 5 | 0.9978 | 3.51 | 9.4 |
| 2 | 5 | 0.9968 | 3.2 | 9.8 |
| 3 | 5 | 0.997 | 3.26 | 9.8 |
| 4 | 6 | 0.998 | 3.16 | 9.8 |
| 5 | 5 | 0.9978 | 3.51 | 9.4 |
| 6 | 5 | 0.9978 | 3.51 | 9.4 |
| 7 | 5 | 0.9964 | 3.3 | 9.4 |
| 8 | 7 | 0.9946 | 3.39 | 10 |
| 9 | 7 | 0.9968 | 3.36 | 9.5 |
| 10 | 5 | 0.9978 | 3.35 | 10.5 |
| 11 | 5 | 0.9959 | 3.28 | 9.2 |
| 12 | 5 | 0.9978 | 3.35 | 10.5 |
| 13 | 5 | 0.9943 | 3.58 | 9.9 |
| 14 | 5 | 0.9974 | 3.26 | 9.1 |
| 15 | 5 | 0.9986 | 3.16 | 9.2 |
| 16 | 5 | 0.9986 | 3.17 | 9.2 |
| 17 | 7 | 0.9969 | 3.3 | 10.5 |
| 18 | 5 | 0.9968 | 3.11 | 9.3 |
| 19 | 4 | 0.9974 | 3.38 | 9 |
| 20 | 6 | 0.9969 | 3.04 | 9.2 |
| 21 | 6 | 0.9968 | 3.39 | 9.4 |
| 22 | 5 | 0.9982 | 3.52 | 9.7 |
| 23 | 5 | 0.9966 | 3.17 | 9.5 |
| 24 | 5 | 0.9968 | 3.17 | 9.4 |
| 25 | 6 | 0.9968 | 3.43 | 9.7 |
| 26 | 5 | 0.9955 | 3.34 | 9.3 |
| 27 | 5 | 0.9962 | 3.28 | 9.5 |
| 28 | 5 | 0.9966 | 3.17 | 9.5 |
| 29 | 5 | 0.9972 | 3.47 | 9.4 |
| 30 | 6 | 0.9964 | 3.38 | 9.8 |
| 31 | 5 | 0.9958 | 3.35 | 10.1 |
| 32 | 6 | 0.9966 | 3.46 | 10.6 |
| 33 | 5 | 0.9966 | 3.17 | 9.8 |
| 34 | 6 | 0.9993 | 3.45 | 9.4 |
| 35 | 5 | 0.9957 | 3.38 | 9.2 |
| 36 | 6 | 0.9986 | 3.4 | 9.6 |
| 37 | 6 | 0.9975 | 3.42 | 10.8 |
| 38 | 7 | 0.9968 | 3.23 | 9.7 |
| 39 | 4 | 0.994 | 3.5 | 9.8 |
| 40 | 5 | 0.9963 | 3.31 | 9.1 |

**MODEL BUILDING: -**

Here we take the quality of wine to be the dependent variable dependent on the independent variables Density, Alcohol, pH since it is an intrinsic truth that the standards of the wine may depend on these variables.

**DESCRIPTIVE STATISTICS: -**



**CORRELATION: -**



We see that there is either Low degree of correlation or Negligible correlation between the variables which means there is no relation among the variables and by looking at the values we can infer that there is hardly any chance of Multicollinearity existing over here.

**SCATTER PLOT: -**



**LINE & SYMBOL PLOT: -**



**REGRESSION OUTPUT: -**



Estimated model: quality = (-57.23) +(58.92) \*(density) +(-0.62) \*(ph) +(0.61) \*(alcohol)

**CONCLUSION: -**

The values of R-squared and Adjusted R-squared are 0.136 and 0.064 respectively which are very low and thus indicates the low explanatory power of our proposed model by calculated by the OLS method. On the other hand, the probability values of the coefficient, density, pH and the F-static are also acting as added negatives (since these values are above 0.05) thus making the model statistically insignificant. Since we concluded that Multicollinearity is non-existent between the independent variables of the equation, hence we postulate that the estimated coefficients in the equation possess the characteristics of BLUE.

NAME: S.KRISHNAN

REG.NO: 200968068

SEC: DSE-A

ROLL.NO: 19