WQD.3 - Application of Polynomial Regression

In order to investigate whether a polynomial relationship fits the model better, an alternative model with squared terms of the significant variables is tried, which improves R^2 value to 31%.

Regression Coefficients	Estimate	Std. Error	t	Pr(>|t|)
(Intercept)	5.78	0.11	51.91	0.00
poly(alcohol, 2)1	18.03	0.92	19.52	0.00
poly(alcohol, 2)2	2.01	0.72	2.78	0.00
poly(volatile.acidity, 2)1	-6.52	0.70	-9.31	0.00
poly(volatile.acidity, 2)2	2.70	0.68	3.96	0.00
residual.sugar	0.02	0.00	4.57	0.00
poly(free.sulfur.dioxide, 2)1	2.57	0.73	3.52	0.00
poly(free.sulfur.dioxide, 2)2	-0.24	0.68	-7.70	0.00
chlorides	-0.64	1.83	-3.63	0.00
sulphates	0.65	0.16	4.21	0.00
poly(pH, 2)1	1.75	0.70	2.52	0.00
poly(pH, 2)2	2.39	0.68	3.50	0.00

Sample R output:

```
Residual standard error: 0.6727 on 1999 degrees of freedom
Multiple R-squared:  0.3146,   Adjusted R-squared:  0.3108
F-statistic: 83.42 on 11 and 1999 DF,  p-value: < 2.2e-16
```

Application of this model on test data gives sum of square of differences between the actual response and predicted response to be 1139.41 whereas sum of square of deviations of actual response is 1554.754. Ratio of these two may be taken as the ratio of Error sum of squares and total sum of squares. Hence a measure similar to that of R^2 may be computed as 1 – 1139.41/1554.754 = 0.2671.

Source URL: https://onlinecourses.science.psu.edu/stat857/node/226