GADS12-NYC: Regression Project

Wine Data

Response Variable

- "quality" of wine as determined by score from 3 to 8
  - count 1599.000000
  - mean 5.636023
  - std 0.807569
  - min 3.000000
  - 25% 5.000000
  - 50% 6.000000
  - 75% 6.000000
  - max 8.000000
  - dtype: float64

Explanatory Variable

- 8 Different Explanatory Variable that can potentially explain variables in how wines score in the "quality" metrics
- i.e.) Fixed Acidity, pH, level of chlorides, level of sulfur dioxide etc…
- Are any of the values of the explanatory variables missing?
  - Nope
- Calculate descriptive statistics (min, max, mean, etc.) for the variables
  - Next Slide

Explanatory Variable - Descriptions

	fixed acidity	volatile acidity	citric acid	residual sugar
count	1599.000000	1599.000000	1599.000000	1599.000000
mean	8.319637	0.527821	0.270976	2.538806
std	1.741096	0.179060	0.194801	1.409928
min	4.600000	0.120000	0.000000	0.900000
25%	7.100000	0.390000	0.090000	1.900000
50%	7.900000	0.520000	0.260000	2.200000
75%	9.200000	0.640000	0.420000	2.600000
max	15.900000	1.580000	1.000000	15.500000

	chlorides	free sulfur dioxide	total sulfur dioxide	density
count	1599.000000	1599.000000	1599.000000	1599.000000
mean	0.087467	15.874922	46.467792	0.996747
std	0.047065	10.460157	32.895324	0.001887
min	0.012000	1.000000	6.000000	0.990070
25%	0.070000	7.000000	22.000000	0.995600
50%	0.079000	14.000000	38.000000	0.996750
75%	0.090000	21.000000	62.000000	0.997835
max	0.611000	72.000000	289.000000	1.003690

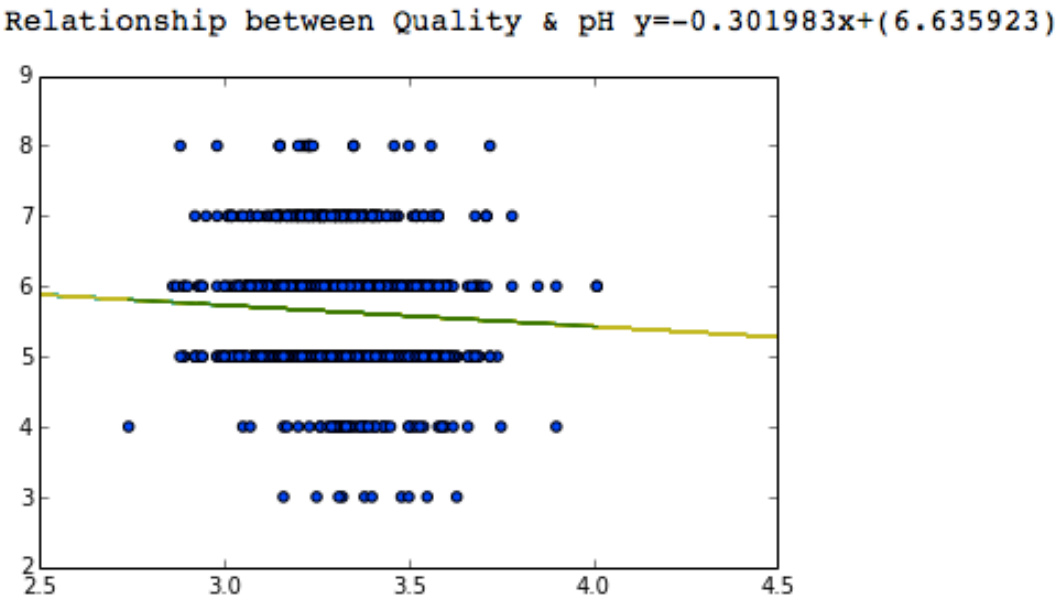
	pH	sulphates	alcohol
count	1599.000000	1599.000000	1599.000000
mean	3.311113	0.658149	10.422983
std	0.154386	0.169507	1.065668
min	2.740000	0.330000	8.400000
25%	3.210000	0.550000	9.500000
50%	3.310000	0.620000	10.200000
75%	3.400000	0.730000	11.100000
max	4.010000	2.000000	14.900000

Explanatory Variable – Covariance between pH and Quality

- Calculate the covariance of the response variable and some of the explanatory variables. Plot the response variable against some of the explanatory variables. What relationships do you observe?
  - Covariance calculated for pH and Quality
  - Covariance between pH and Quality is -0.00719782232481
  - Variance in pH is 0.0238351805454
  - beta value for pH and Quality is -0.301983125787
  - alpha value for pH and Quality is 6.63592282676

Relationship between pH & Quality

- pH and Quality are negatively correlated where the wine quality score tends to decrease as the wine becomes less acidic



Other Relationship between Expl and Resp shown in ipython notebook