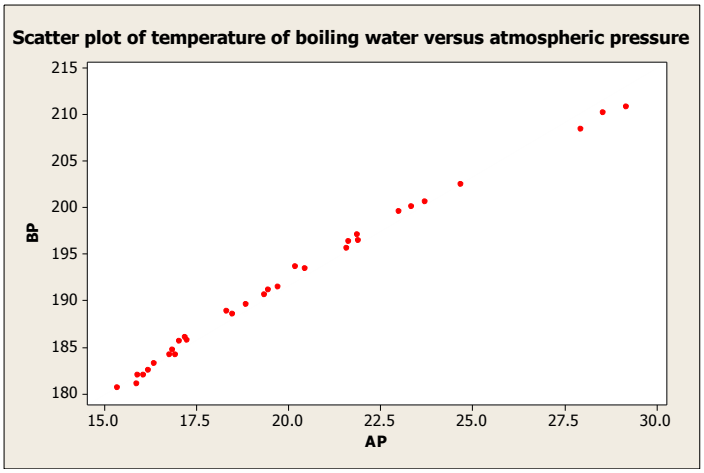
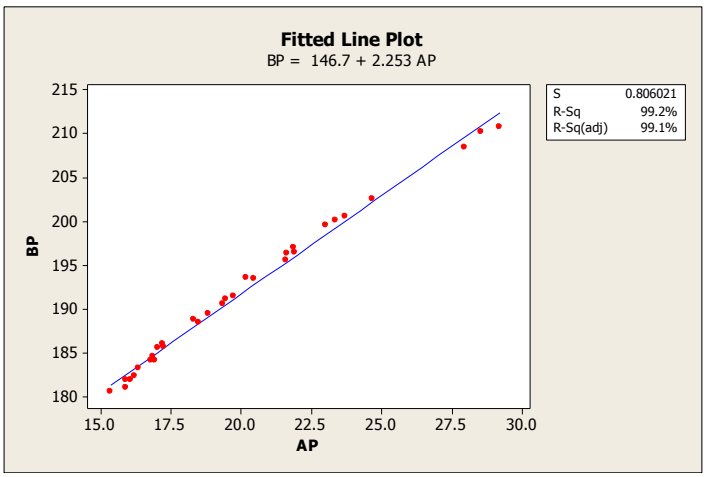
Queen Mary
University of London
School of Mathematical Sciences

MTH5120 STATISTICAL MODELLING I — Practical 3 Solutions



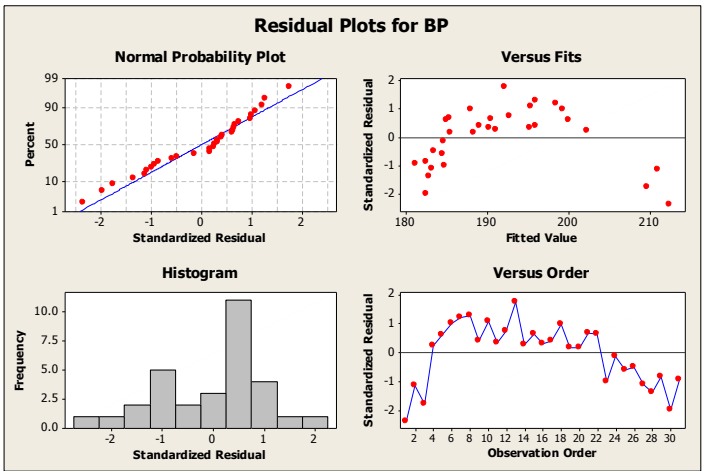
The scatter plot clearly shows that the boiling point of water increases with the atmospheric pressure.



The fitted model is

$$\hat{y}_i = 146.7 + 2.253\, x_i$$

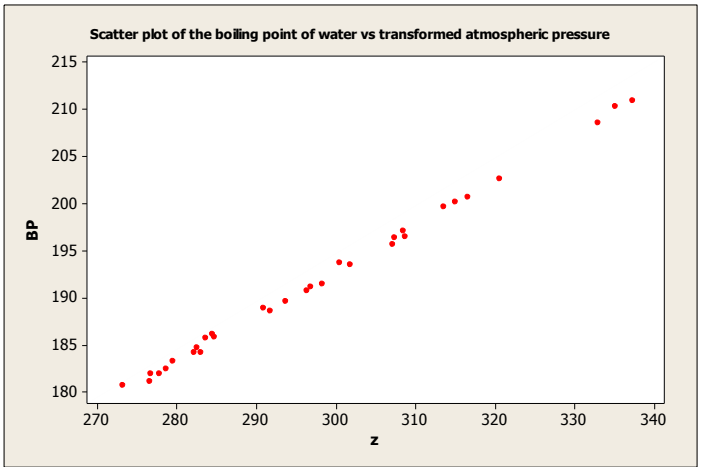
where $\hat{y}_i$ denotes the estimate of the mean boiling point at x_i - the i-th value of the atmospheric pressure.



The residual plots contradict the assumptions of normality and also the linearity of the model.

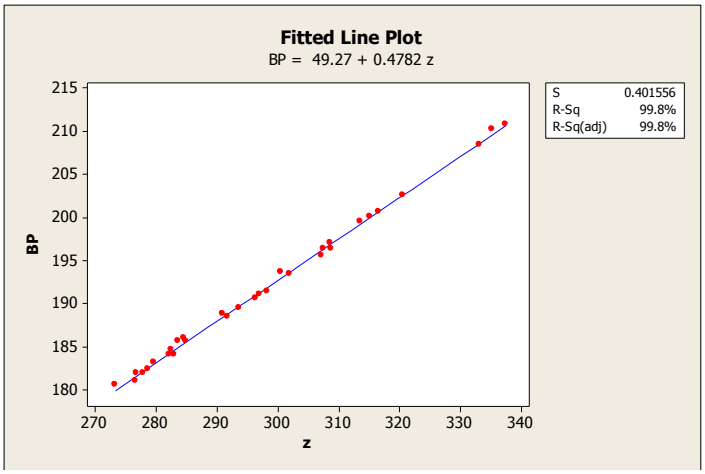
The sample distribution seems to be skewed.

The plot of residuals versus fitted values shows that the residuals are grouped into negative, then positive and then again negative clusters. This suggests that the expected response may not be a linear function of the explanatory variable. We would get the same picture by plotting residuals versus the explanatory variable.



Here also the scatter plot clearly shows that the boiling point of water increases with the atmospheric pressure. This plot suggests that the boiling point depends linearly on the transformed pressure

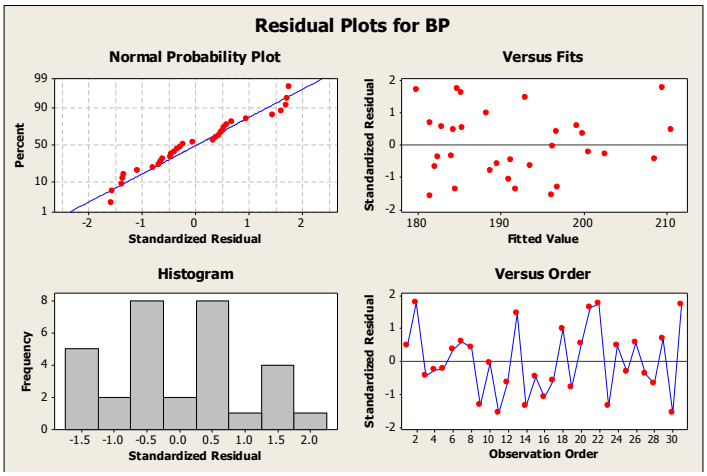
Z = 100 ln(AP).



The fitted model is

$$\hat{y}_i = 49.27 + 0.4782\, z_i$$

where $\hat{y}_i$ denotes the estimate of the mean boiling point at z_i which is the i-th value of the transformed atmospheric pressure, that is $z_i = 100\, ln(BP)$.



The residuals did improve.

The linearity of the model is not contradicted.

The normality assumption is not clearly violated. The data set is not large and there is no obvious departure from the assumption.

There is no apparent problem regarding the assumption of constant variance.

Regression Analysis: BP versus z

```
The regression equation is
BP = 49.27 + 0.4782 z

S = 0.401556   R-Sq = 99.8%   R-Sq(adj) = 99.8%

Analysis of Variance
Source       DF       SS       MS         F       P
Regression    1  2286.64  2286.64  14180.95   0.000
Error        29     4.68     0.16
Total        30  2291.31
```

The F-test indicates that the regression is highly significant, that is we can reject the null hypothesis

$H_0: \beta_1 = 0$ versus the alternative $H_1: \beta_1 \neq 0$

at the significance level $\alpha < 0.001$.

The data suggest that the boiling point of water increases linearly with the transformed atmospheric pressure.