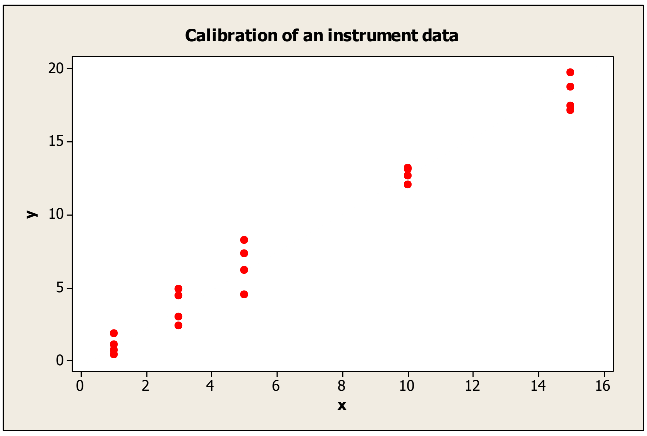
Queen Mary University of London
School of Mathematical Sciences

Minitab Project Report – Practical 1

Example 1



Regression Analysis: y versus x

```
The regression equation is
y = 0.061 + 1.23 x

Predictor      Coef   SE Coef      T      P
Constant     0.0612    0.3916   0.16  0.878
x           1.23144   0.04615  26.69  0.000

S = 1.04743   R-Sq = 97.5%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  781.28  781.28 712.12  0.000
Residual Error  18   19.75    1.10
Total           19  801.03
```

The relation between Y and X seems to follow an increasing pattern.

The linear model can be written as

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

independently

and β_0 and β_1 are unknown model parameters: the intercept and the slope of the line.

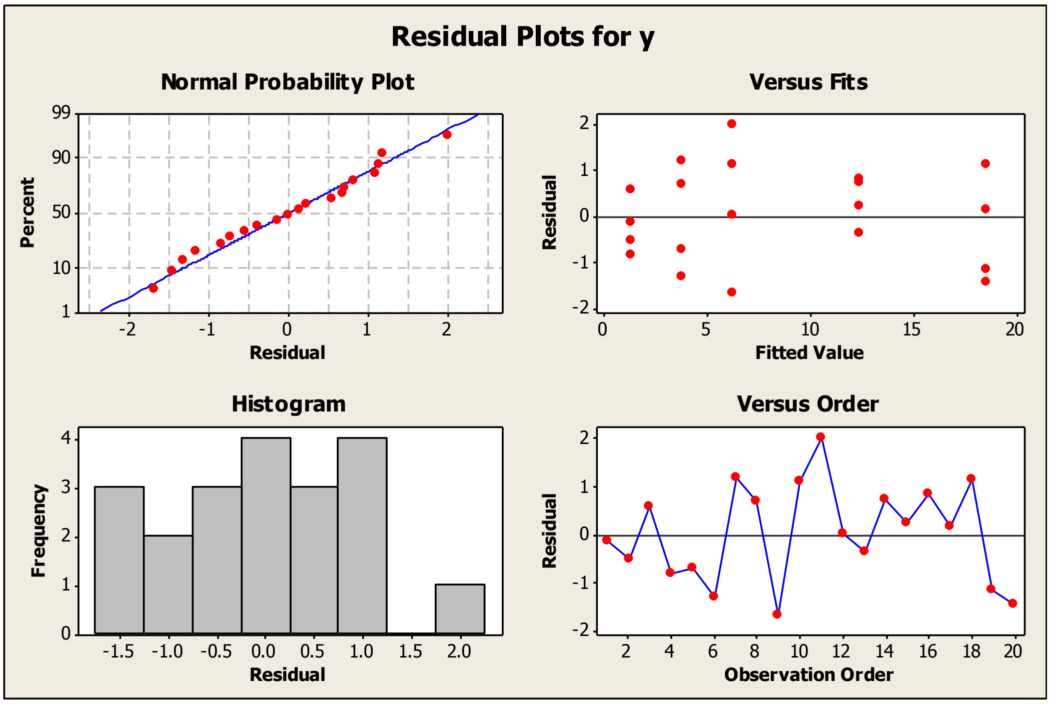
At x = 0 we would expect y = 0.
Hence, in fact we should fit a no-intercept model; this is supported by a large p-value for the intercept.

No-intercept model

```
The regression equation is
y = 1.24 x

Predictor       Coef   SE Coef       T      P
Noconstant
x            1.23722   0.02688   46.02   0.000


Analysis of Variance

Source          DF      SS      MS        F      P
Regression       1  2204.2  2204.2  2117.86  0.000
Residual Error  19    19.8     1.0
Total           20  2224.0
```

The interpretation of the model fit:
When lactic acid concentration changes by one unit, then, on average, the measurement of the concentration changes by 1.24 of the unit.

At $x = 7$ the measured concentration is estimated to be $1.24 \times 7 = 8.66$.

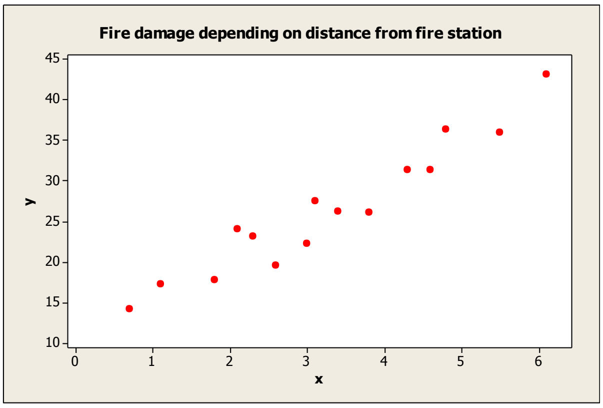


The normal plot gives no clear reason to doubt the assumption of normality. As for the residuals versus fits, the spread is a bit different for different values of x but there is no systematic pattern (it gets larger, then smaller, then larger again). I would not think this is sufficient to cast doubt on the assumption of a constant variance. None of the residuals are very large. The plot against order is probably irrelevant as we don't know if the data was collected in this order.

The ANOVA test
H_0: $\beta_1 = 0$ (the slope is zero)
H_1: $\beta_1 \neq 0$ (the slope is significantly different from zero)

The value of T- statistic is 46.02 (F-statistic 2117.86) and the p-value is < 0.0001. Hence we reject the null hypothesis. The measured values depend linearly on the actual value of the concentration.

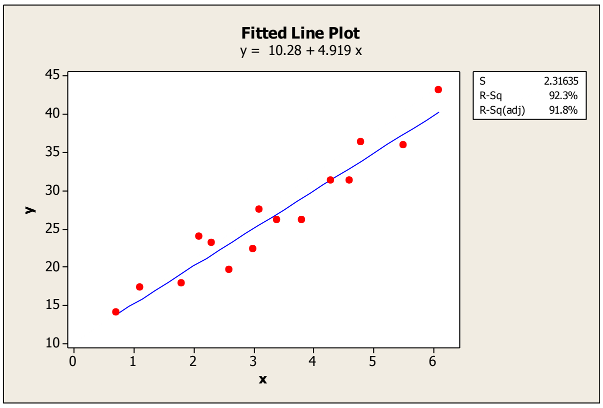
Example 2



The data suggest a simple linear regression model

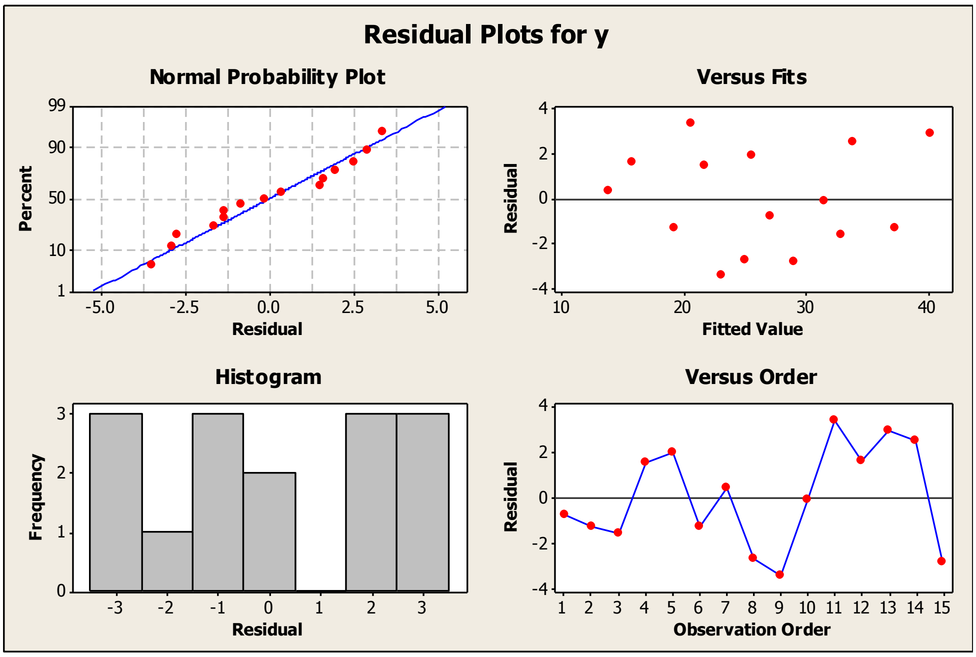
$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

independently



Regression Analysis: y versus x

```
The regression equation is
y = 10.3 + 4.92 x

Predictor     Coef  SE Coef      T      P
Constant    10.278    1.420   7.24  0.000
x           4.9193   0.3927  12.53  0.000

Analysis of Variance
Source          DF      SS      MS       F      P
Regression       1  841.77  841.77  156.89  0.000
Residual Error  13   69.75    5.37
Total           14  911.52
```



The residual plots give us no reason to doubt the model assumptions.

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

The value of the F-statistic is 156.89 and the p-value is < 0.0001. Hence we reject the null hypothesis.
The amount of damage caused by fire depends significantly on the distance of the property from the closest fire station.

```
y = 10.278 + 4.9193 * 3.5 = 27.496
```

The estimated mean cost of fire damage for a residence 3.5 km from the nearest station is £27,496.

The intercept (£10,280) gives the average cost for a fire which happens very close to a fire station. Even if the fire engines didn't have to go very far there would be a large amount of damage. The slope (£4,919) is the average extra cost for each further kilometer that the fire engine has to travel.