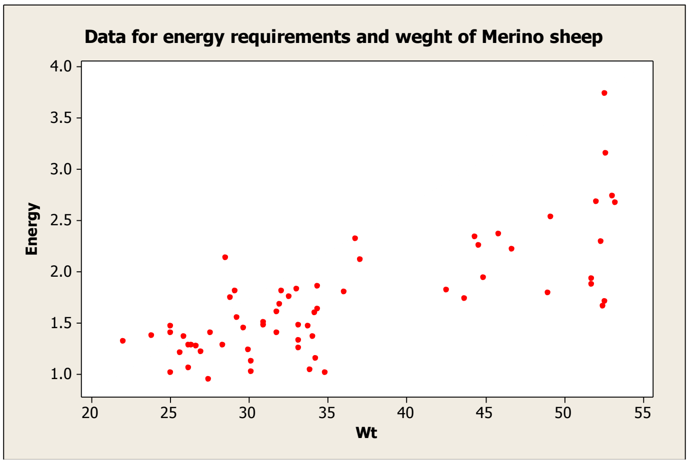
Queen Mary
University of London
School of Mathematical Sciences

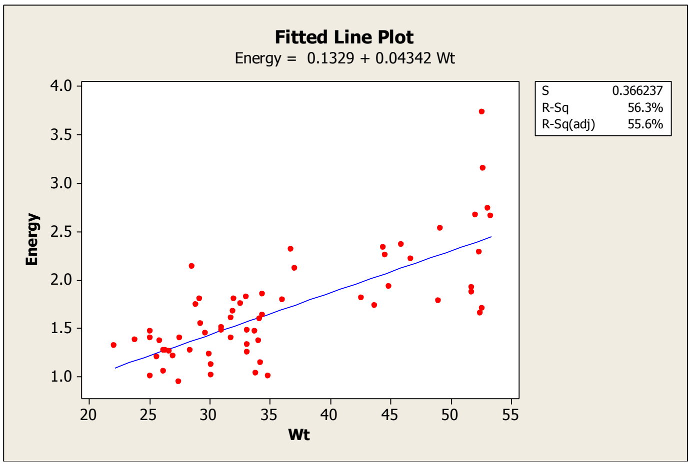
MTH5120 STATISTICAL MODELLING I

Practical 2 Solutions

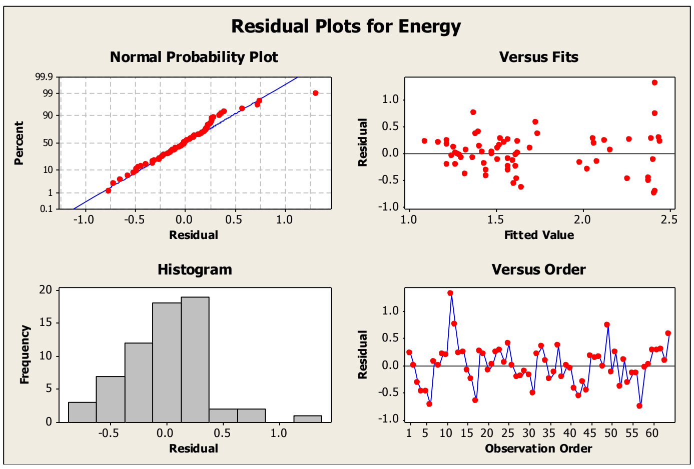


The energy requirements increase with weight of Merino sheep. There is a considerable variability in the data, but it seems to be stable. However, we see one observation of energy requirements which is clearly larger than other observations for similar weight.

The model we fit here is $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where $\varepsilon_i \sim N(0, \sigma^2)$ and are uncorrelated.



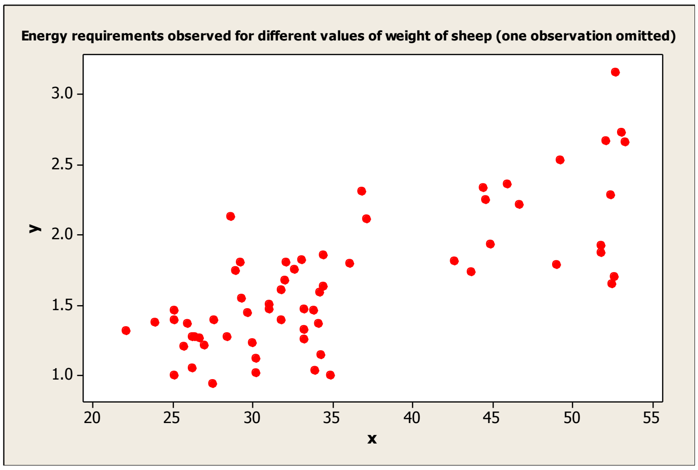
The model fit is $\hat{y}_i = 0.1329 + 0.04342\, x_i$.



The normality assumption may be violated, the distribution seems to be skewed.

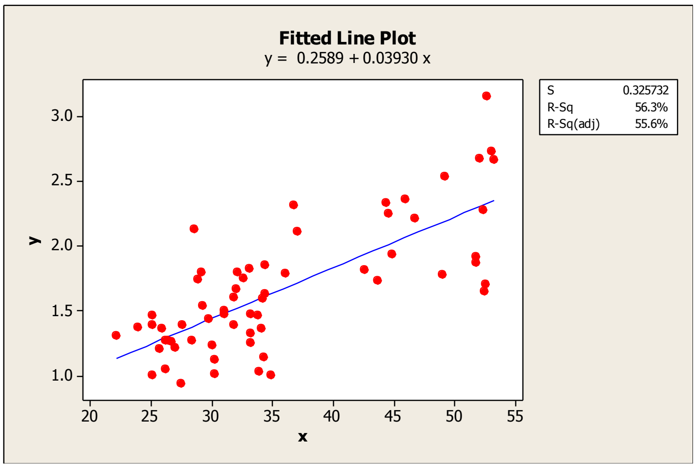
As to the assumption of constant variance, for large fitted values it seems that the variation is larger than for the other values. It may seem so because of the one residual whose value for the unusual observation.

The large residual is for observation 11.



This is a scatter plot of the data after removing observation 11, that is (52.60, 3.73).

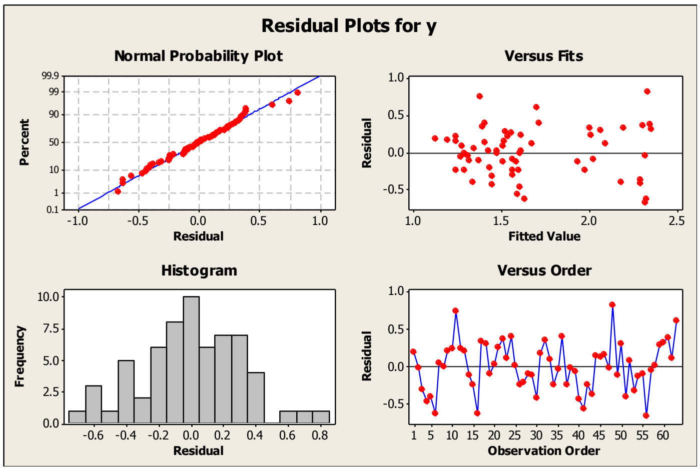
As before the energy levels increase with weight, but there is no clearly outstanding observation.



The model fit now is

$$\hat{y}_i = 0.2589 + 0.0393\, x_i$$

The estimates of the intercept and the slope have changed; the slope is smaller while the intercept is larger.



The residuals have improved.

Now, there is no clear evidence against the assumption of normality, neither against the assumption of constant variance of the errors.

Regression Analysis: y versus x

```
The regression equation is
y = 0.2589 + 0.03930 x

S = 0.325732   R-Sq = 56.3%   R-Sq(adj) = 55.6%

Analysis of Variance
Source      DF       SS        MS      F      P
Regression   1   8.3475   8.34747  78.67  0.000
Error       61   6.4722   0.10610
Total       62  14.8197
```

The Variance Ratio F = 78.67 is large and the respective p-value is very small, p < 0.001. Hence, we reject the hypothesis that the slope is zero, that is, we reject

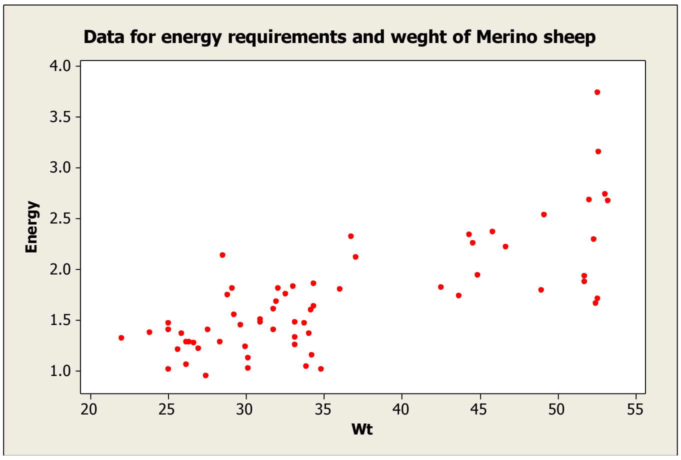
$$H_0: \beta_1 = 0$$

versus

$$H_0: \beta_1 \neq 0$$

at a significance level $\alpha < 0.001$.

The slope is strongly significant.

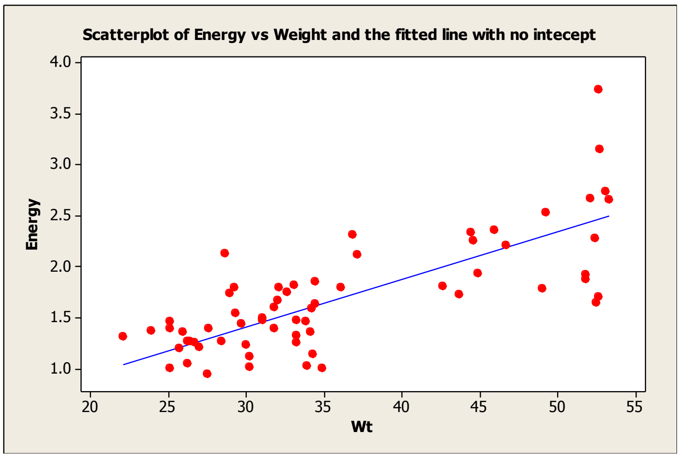


The energy requirements increase with weight of Merino sheep. There is a considerable variability in the data, but it seems to be stable. However, we see one observation of energy requirements which is clearly larger than other observations for similar weight.

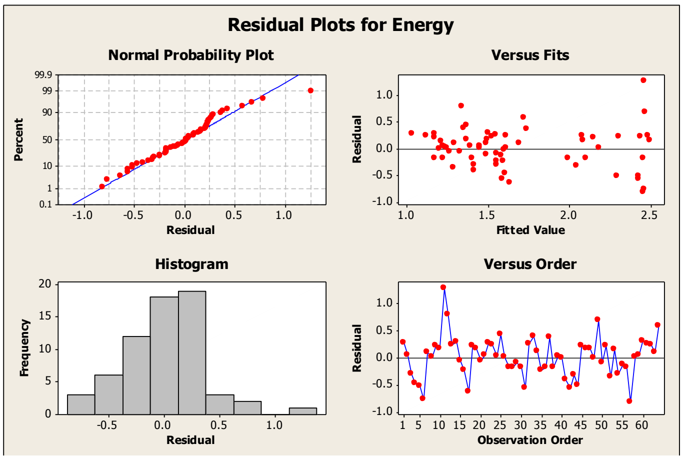
We will fit a no-intercept model:

$$Y_i = \beta_1 x_i + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and are uncorrelated.



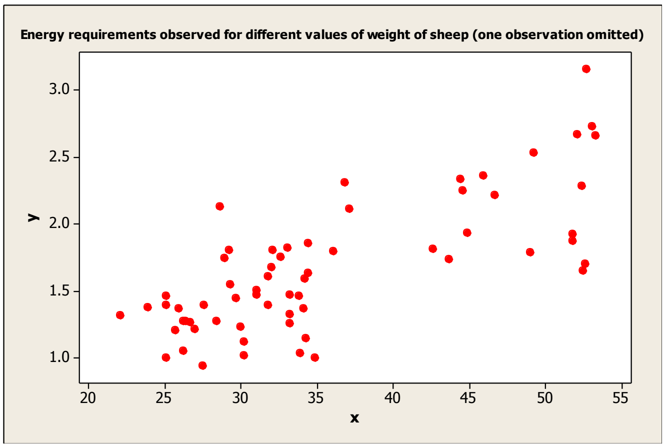
The model fit is $\hat{y}_i = 0.0469\, x_i$.



The normality assumption may be violated, the distribution seems to be skewed.

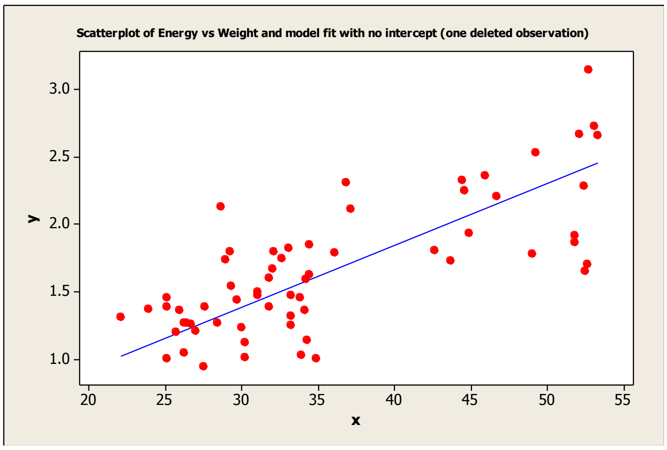
As to the assumption of constant variance, for large fitted values it looks that the variation is also large compared to other values. It may seem so because of the one residual whose value is much larger than other residual values.

The large residual is for observation 11.



This is a scatter plot of the data after removing observation 11, that is (52.60, 3.73).

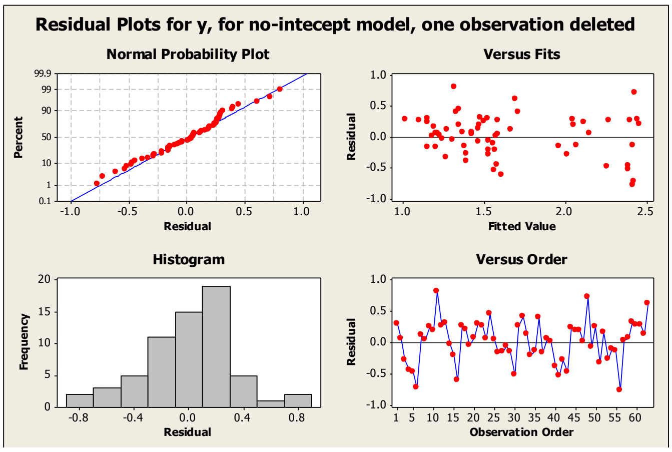
As before the energy levels increase with weight, but there are no clearly outstanding observations.



The model fit now is

$$\hat{y}_i = 0.0461\, x_i$$

The estimate of the slope has changed; now it is smaller compared to the fit with the full data set.



The residuals have improved.

Now, there is no clear evidence against the assumption of normality, nor against the assumption of constant variance of the errors.

```
The regression equation is
y = 0.0461 x

Predictor        Coef    SE Coef       T      P
Noconstant
x            0.046100   0.001127   40.90  0.000

S = 0.329682

Analysis of Variance
Source          DF      SS      MS        F      P
Regression       1  181.85  181.85  1673.10  0.000
Residual Error  62    6.74    0.11
Total           63  188.59
```

The Variance Ratio F = 1673.10 is very large and the respective p-value is very small, p < 0.001. Hence, we reject the hypothesis that the slope is zero, that is, we reject

$$H_0\colon \beta_1 = 0$$

versus

$$H_0\colon \beta_1 \neq 0$$

at a significance level $\alpha < 0.001$.

The slope is strongly significant.

2.2

Assuming that for zero weight we should expect zero energy requirements, the model with no intercept would be more appropriate than the model with intercept. Out of 64 observations there was one rather unusual observation which caused the residuals not to meet the model assumptions. Hence, the model with no-intercept and with the unusual observation deleted fits best for this data set.

Also, the F value is very large for this model. This suggests that the variability in the data accounted for by the regression is much higher than the variability accounted for by the residuals. The slope is highly significant.

2.3

The estimate of an average increase of energy requirements for an increase of weight of Merino sheep by 1 kg is 0.0461 Mcal/sheep/day.