Analysis of residuals

Scatter diagram usually helps us understand whether a simple linear regression model is appropriate. When scatter diagram is not enough, we first fit the simple linear regression model and then analyze the residuals $Y_i - (\hat{\alpha} + \hat{\beta} x_i)$. We calculate the "*standardized residuals*" as follows:

$$\frac{Y_i - (\hat{\alpha} + \hat{\beta} x_i)}{\sqrt{SS_e/(n-2)}}$$

When the simple linear regression model is correct, the standardized residuals are approximately independent standard normal random variables, and should be randomly distributed above and below 0 with about 95 percent of their values being between −2 and +2. In addition, a plot of the standardized residuals should not indicate any distinct pattern. Indeed, any indication of a distinct pattern should make us suspicious about the validity of the assumed simple linear regression model. (*See the plots on pages 381 and 382 of textbook*.)

Transforming to linearity

In many cases, a linear regression is not appropriate. In such cases, if the form of the relationship can be determined, it is sometimes possible to transform it into a linear form. For instance, in certain applications it is known that $W(t)$, the amplitude of a signal a time *t* after its origination, is approximately related to *t* by the functional form

$$W_t \approx c\exp(-dt)$$

so that

$$\log W_t \approx \log c - dt$$

or

$$\log W_t = \log c - dt + \epsilon$$

or

$$Y_t = \alpha + \beta t + \epsilon$$

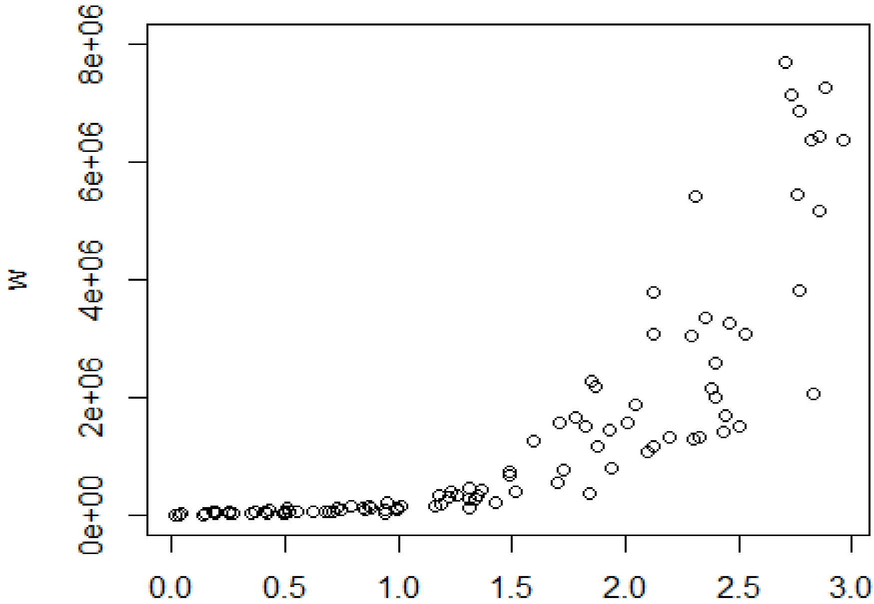
The regression parameters α and β would then be estimated by the usual least squares approach and the original functional relationships can be predicted from

$$W_t \approx \exp(\hat{\alpha} + \hat{\beta} t)$$

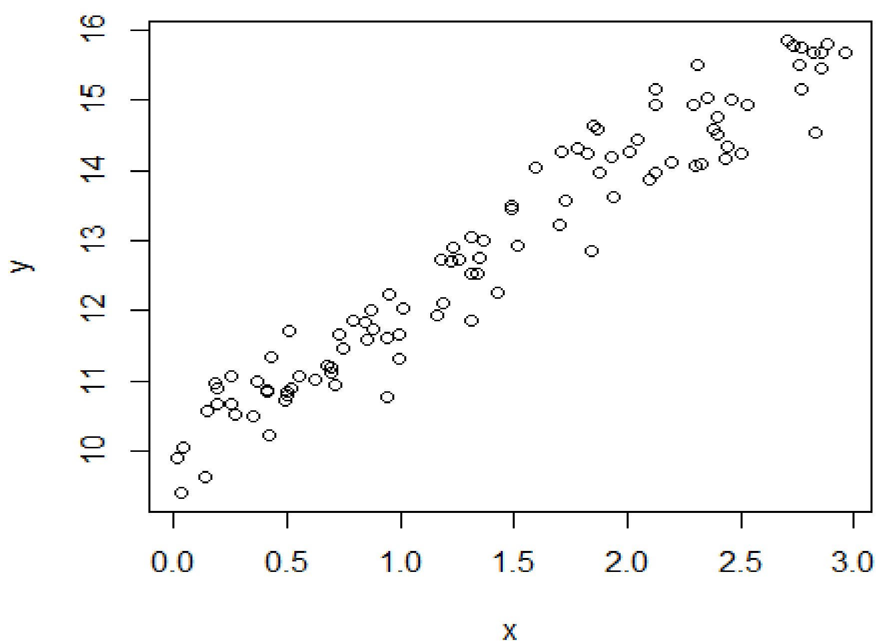
Example

Consider the following scatter diagram. The relationship appears to be

$$W_x \approx c\,\exp(dx)$$

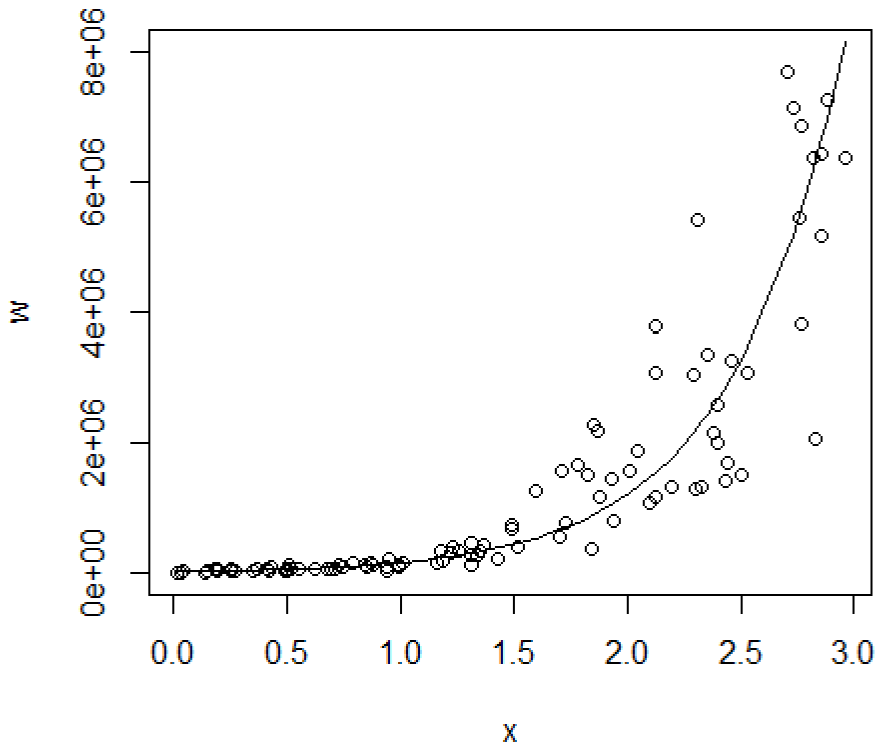


Using $Y_x = \log W_x$ we have the following plot



A simple linear regression model is now fitted: $\hat{\alpha} = 10.03$ and $\hat{\beta} = 1.99$. Thus,

$$W_x \approx \exp(10.03 + 1.99x)$$



Weighted least squares

Sometimes the variance of a response is not constant but depends on the value of x. then the regression parameters α and β should be estimated by minimizing a weighted sum of squares. That is, if

$$V(Y_i) = \frac{\sigma^2}{w_i}$$

then to estimate the regression parameters, we minimize

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} w_i \, (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

That is, we minimize

$$\sum_{i=1}^{n} w_i \, (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

We try to choose an appropriate set of weights and then use them in the above expression.