Concepts in non-linear regression

Non-linear models

Recall that a linear model takes on the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

This is a linear function.

A model which is not linear is called a **non-linear model**. There are many different approaches to non-linear modelling.

We will look at two approaches here:

- linearisation
- non-linear least squares

Linearisation

One way to create a non-linear model is to:

- transform the data in some way that makes it linear
- apply linear regression to the transformed data
- reverse the transformation on the model

For example, if a power law model was proposed:

$$y = a \cdot x^b,$$

then transforming it with logarithms yields:

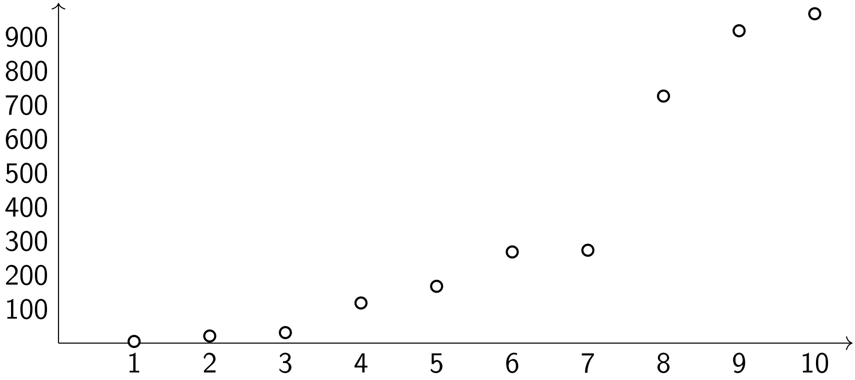
$$\log(y) = \log(a \cdot x^b) = \log(a) + b \log(x)$$

You then treat "$\log(y)$" as the response variable and "$\log(x)$" as the explanatory variable to estimate $\log(a)$ and b.

Example

Consider the following set of data points:

$$(1,5),(2,21),(3,31),(4,118),(5,167),(6,268),(7,273),(8,726),(9,918),(10,968)$$



We propose a model of the form $y = a \cdot x^b$. We need to estimate the coefficient a and the power b.

Example

We propose a model of the form $y = a \cdot x^b$. We need to estimate the coefficient a and the power b.

Linearising the model:

Since the model is "log(y)" in terms of "log(x)", the data needs to be transformed accordingly as well:

```
x <- 1:10
y <- c(5,21,31,118,167,268,273,726,918,968)
log.x <- log(x)
log.y <- log(y)
linearised.model <- lm(log.y ~ log.x)
```

Let's view this in R...

Example

The estimated linearised model is

$$\widehat{\log(y)} = 1.3617 + 2.3732 \log(x)$$

The transformation was $y \to \log(y)$, so to undo that, we need to substitute both sides into the exponential function:

$$\widehat{y} = e^{\widehat{\log(y)}} =$$

A problem

Accounting for the error term, suppose that the proposed model is

$$y = a \cdot x^b + \varepsilon$$

If we try to linearise:

$$\log(a \cdot x^b + \varepsilon) = ???$$

To satisfy the linear model assumptions, we need a multiplicative error term instead:

$$y = a \cdot x^b \cdot \varepsilon$$

$$\implies \log(y) =$$

We require homoskedasticity of $\log(\varepsilon)$, and normality for hypothesis testing/confidence intervals.

We can check for model violations using the techniques from Week 10.

Other transformations

We previously looked at a model of the form $y = a \cdot x^b$.
You could apply similar transformations to:

- $y = a \cdot b^x \implies$
- $y = a \cdot e^{bx} \implies$

It need not necessarily be logarithmic/exponential:

$$y = \frac{a}{b + x} \implies$$

Non-linear least squares

If we prefer, for example, a model of the form

$$y = a \cdot x^b + \varepsilon,$$

then we can use **non-linear least squares regression** instead.

Homoscedasticity and normality of the error terms is a requirement for inferencing.

This is easily performed using the `nls` function in R.