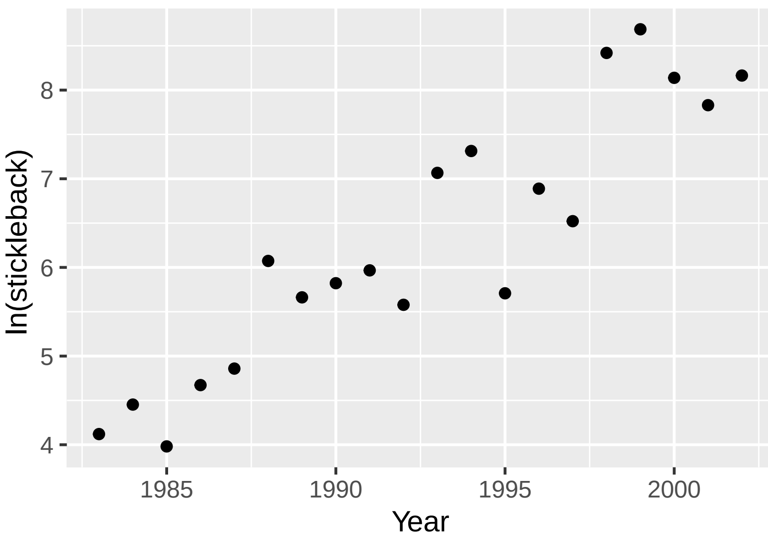
Level M Regression Models Examples

Modelling Stickeback catches



The number of stickleback (a species of fish) trapped in the filters at the pumping station at Ross Priory in Loch Lomond were recorded each year from 1983 to 200b) It was of interest to investigate the pattern over time for the number of stickleback. A natural log (ln) transformation has been applied to the stickleback data and it is plotted below for the years where data were available.

a) Use the plot to comment on the relationship between year and ln(stickleback).

b) Write down a possible statistical model to describe how ln(stickleback) depends on the year.

Solution

a) There appears to be a moderate positive relationship between ln stickleback catches and year and the variability appears fairly constant over time. There may be some evidence of cycles in the data.

b) A possible model is:

$$\ln(\text{stickleback})_i = \alpha + \beta \text{year}_i + \epsilon_i$$

Identifying linear models

Which of the following models are linear models? Justify your answer.

a) $Y_i = \alpha + \beta x_i + \gamma x_i^2 + \epsilon_i$

b) $Y_i = \alpha + \gamma x_i^{\beta} + \epsilon_i$

c) $E(Y_i) = e^{\beta x_i}$

Solution
Model (a): $E(Y_i) = \alpha + \beta x_i + \gamma x_i^2$ is a quadratic regression model
Model (b): $Y_i = \alpha + \gamma x_i^{\beta} + \epsilon_i$ is not a linear model as β appears in the exponent of x_i
Model (c): $E(Y_i) = e^{\beta x_i}$ is not a linear model as β appears β enters the model non-linearly.

Least Square Estimates

Consider the following model:
Data: $(y_i, x_i) \quad i = 1, \ldots, 24$
Model: $E(Y_i) = \beta x_i, Var(Y_i) = \sigma^2$

a) Using the sum-of-squares function $S(\beta) = \sum_{i=1}^{24}(y_i - \beta x_i)^2$ derive, from first principles, the least-squares estimator for β.

b) Use the summary statistics below to calculate the least-squares estimate for β and the residual sum-of-squares.

$$\sum_{i=1}^{24} x_i y_i = 12514, \sum_{i=1}^{24} x_i^2 = 29.518, \sum_{i=1}^{24} y_i^2 = 6511425$$

Solution

a)

$$S(\beta) = \sum_{i=1}^{n}(y_i - \beta x_i)^2$$

Calculus approach:

$$\begin{aligned}
S(\beta) &= \sum_{i=1}^{n}(y_i - \beta x_i)^2 \\
&= \sum_i y_i^2 - 2\beta \sum_i x_i y_i + \beta^2 \sum_i x_i^2 \\
\frac{\partial S}{\partial \beta} &= -2\sum_i x_i y_i + 2\beta \sum_i x_i^2 \\
\frac{\partial S}{\partial \beta} &= 0 \\
-2\sum_i x_i y_i + 2\beta \sum_i x_i^2 &= 0 \\
-\sum_i x_i y_i + \beta \sum_i x_i^2 &= 0 \\
\hat{\beta} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2}
\end{aligned}$$

b)

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{12514}{29.518} = 423.94$$

Residual Sum of Squares:

$$
\begin{aligned}
RSS = S(\hat{\beta}) &= \sum_{i=1}^{n}(y_i - \hat{\beta}x_i)^2 \\
&= \sum_{i=1}^{n} y_i^2 - 2\hat{\beta}\sum_{i=1}^{n} x_i y_i + \hat{\beta}^2 \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} x_i y_i)^2}{\sum_{i=1}^{n} x_i^2} \\
&= 6511425 - \frac{(12514)^2}{29.518} \\
&= 1206180.88
\end{aligned}
$$