

Broken Stick Regression, Polynomial Regression, Splines

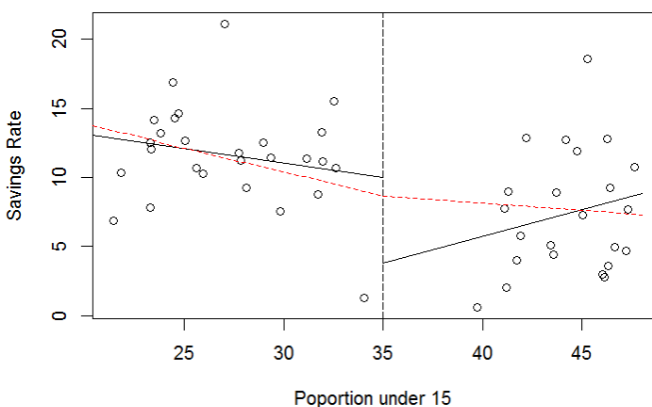
Transformations of the response and predictors can improve the fit and correct violations of model assumptions such as non-constant error variance. This chapter focuses on the transformation of predictors.

The idea behind **Broken Stick Regression/Segmented Regression** is that different linear regression models may apply in different regions of the data. We define two basis functions:

$$B_l(x) = \begin{cases} c - x & \text{if } x < c \\ 0 & \text{otherwise} \end{cases}$$

$$B_r(x) = \begin{cases} x - c & \text{if } x > c \\ 0 & \text{otherwise} \end{cases}$$

where c marks the division between groups. $B_l(x)$ and $B_r(x)$ form a first-order spline basis with knot point c . Sometimes $B_l(x)$ and $B_r(x)$ are called hockey-stick functions.



Polynomial Regression

Polynomial Regression models are frequently used curvilinear response models in practice. They make it easy to handle special cases of the linear regression model, however in reality this is not an exact representation.

$$y = \beta_0 + \beta_1 x + \dots + \beta_d x^d + \epsilon$$

We choose d by adding terms until the added term is not statistically significant. Start with a large d and eliminate non-statistically significant terms starting with the highest order term.

Regression Splines

- Polynomials regression has:
 - the advantage of smoothness
 - the disadvantage that each data point affects the fit globally
- Broken stick regression method
 - Localized the influence of each data point to its particular segment
 - Do not have the same smoothness as with polynomials
- Use B-spline basis functions
 - Have both smoothness and local influence
 - Can overfit the model

Revision #1

Created 29 September 2022 23:16:24 by Elkip

Updated 29 September 2022 23:37:52 by Elkip