Question 1

Suppose one fits a simple linear regression model to the data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ as

$$Y_i = \beta_0 + \beta_1 g(x_i) + \epsilon_i, \qquad i \in \{1, 2, \ldots, n\},$$

where $g : \mathbb{R} \to \mathbb{R}$ is some univariate transformation and $n = 20$.

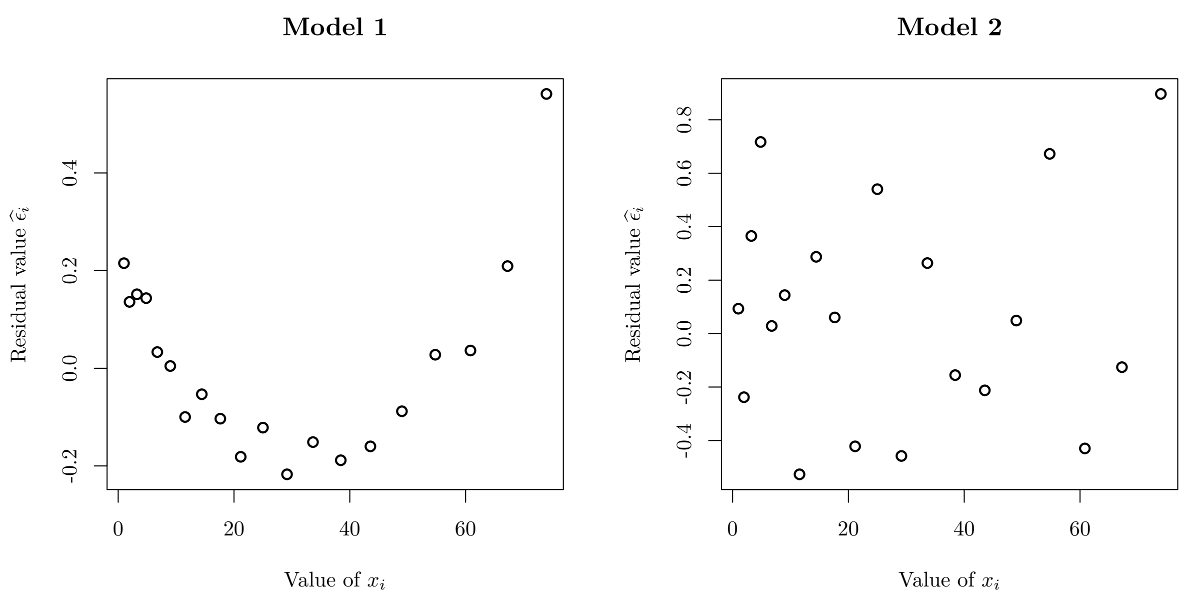
(a) What joint distribution are the errors ϵ_i assumed to follow?

(b) Give examples of suitable transformation functions that could be used for g.

(c) For two different choices of transformation g, one has two models with the fitted residuals

$$\widehat{\epsilon}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 g(x_i),$$

shown in the figures below. For each model, state whether the model fits the data well or not and justify your answer.



Question 2

Suppose one is provided with the data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and one is then asked to fit the regression model

$$Y_i = \beta x_i + \epsilon_i, \qquad i \in \{1, 2, \ldots, n\},$$

to this data, where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are assumed to be independent and identically distributed according to a $\mathrm{N}(0, \sigma^2)$ distribution, where σ^2 is known.

(a) For the above model compute the likelihood $\mathrm{L}(\beta|y_i)$, for $i = 1, 2, \ldots, n$, assuming that the values of σ^2 and $x_1, x_2, \ldots, x_n$ are known.

(b) Using Part (a), compute the likelihood $\mathrm{L}(\beta|\mathbf{y})$ where $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and the values of σ^2 and $x_1, x_2, \ldots, x_n$ are known.

(c) Prove that $\widehat{\beta} = \dfrac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$ is the maximum likelihood estimate of β.

(d) Write down $\widehat{\boldsymbol{\beta}}$, the maximum likelihood estimator of β.

(e) If you used calculus to find the MLE in Part (c), try to find the MLE without using calculus (and if you did not use calculus, then find the MLE using calculus).

Question 3

Suppose one observes the data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ which one then assumes follows the conditional normal model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i \in \{1, 2, \ldots, n\},$$

where the $x_1, x_2, \ldots, x_n$ are constants and $\epsilon_1, \epsilon_1, \ldots, \epsilon_n \sim \mathrm{N}(0, \sigma^2)$ are independent. The maximum likelihood estimates of β_0 and β_1 are denoted $\widehat{\beta}_0$ and $\widehat{\beta}_1$, respectively, and it is well known that they can be defined in terms of $\overline{x}$, $\overline{y}$, S_{xx} and S_{xy}, where

$$S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2, \qquad S_{xy} = \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}).$$

Now suppose the model is reparametrised to

$$Y_i = \alpha_0 + \alpha_1 (x_i - \overline{x}) + \epsilon_i, \qquad i \in \{1, 2, \ldots, n\},$$

and denote the maximum likelihood estimates of α_0 and α_1 as $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$, respectively.

(a) Write down $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (refer to your notes if necessary), and show that $S_{xy} = \sum_{i=1}^n (x_i - \overline{x}) y_i$.

(b) Show that $\widehat{\alpha}_1 = \widehat{\beta}_1$.

(c) Show that $\widehat{\alpha}_0 = \overline{y}$.

(d) Under what conditions does $\widehat{\alpha}_0 = \widehat{\beta}_0$?

(e) Write down $\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}$ and $\widehat{\boldsymbol{\alpha}}_{\mathbf{1}}$, the maximum likelihood **estimators** of $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$.

(f) Show that $\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}$ and $\widehat{\boldsymbol{\alpha}}_{\mathbf{1}}$ are uncorrelated.