## 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 $n = 20$.

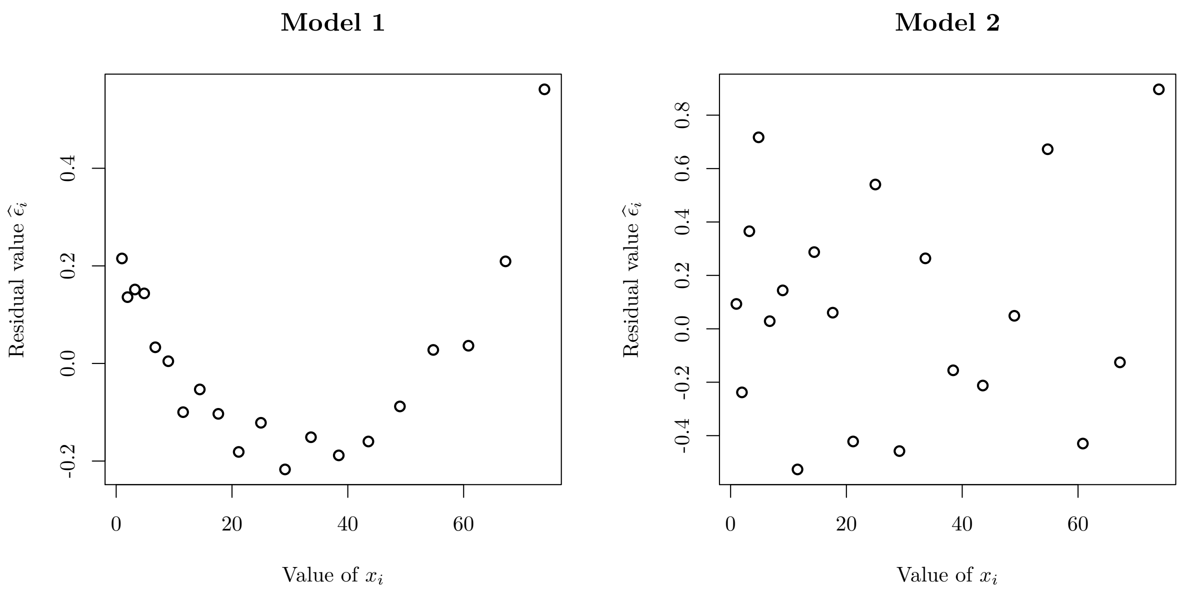
(a) What joint distribution are the errors $\epsilon_i$ assumed to follow?

(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.




**Solution to Question 1**

**Part (a):**

For $i \in \{1, 2, \ldots, n\}$, $\epsilon_i$ is assumed to follow a normal distribution with mean 0 and unknown variance $\sigma^2$, and, moreover, the $\epsilon_i$ are also assumed to be independent.

**Part (b):**

Suitable functions include $g(x) = x^k$, where $k$ is some rational (or real) number (this includes powers and square root), or $g(x) = \exp(x)$, or $g(x) = \log(x)$.

**Part (c):**

Model 1 does **not** appear to fit the data well; there is clearly trend in the residuals (or: they appear to follow a "U"-shape), which means they do not appear to be independently normally distributed.

Model 2 appears to fit the data well because the residuals appear to be normally distributed around 0, and they appear to be independent of each other.

## 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 $\sigma^2$ 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 $\sigma^2$ 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 $\sigma^2$ 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 $\beta$.

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

(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).

**Solution to Question 2**

**Part(a):**

Since $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$, and $\beta$ and each $x_i$ are assumed to be fixed, then for $i = 1, 2, \ldots, n$,

$$\begin{aligned} \beta x_i + \epsilon_i &\sim \mathrm{N}(\beta x_i, \sigma^2), \\ \Rightarrow Y_i &\sim \mathrm{N}(\beta x_i, \sigma^2). \end{aligned}$$

Note that since the $\epsilon_i$ are independent, the $Y_i$ are also independent, for $i = 1, 2, \ldots, n$.

Then, the probability density function of $y_i$ given $\beta$ (assuming $x_i$ is also given and fixed) is

$$\mathrm{L}(\beta|y_i) = f(y_i|\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right).$$

**Part(b):**

Writing $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, the likelihood of $\beta$ given $\mathbf{y}$ is

$$\begin{aligned} \mathrm{L}(\beta|\mathbf{y}) = f(\mathbf{y}|\beta) &= \prod_{i=1}^n f(y_i|\beta) \qquad \text{(since the } Y_i \text{ are independent)} \\ &= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right)\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2\right). \end{aligned}$$

**Part(c):**

In order to maximise the likelihood $\mathrm{L}(\beta|\mathbf{y})$, it is easier to maximise the log-likelihood,

$$\begin{aligned}\log \mathrm{L}(\beta|\mathbf{y}) &= \log\left[(2\pi\sigma^2)^{-n/2}\exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta x_i)^2\right)\right]\\ &= -\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta x_i)^2.\end{aligned}$$

where it assumed that $\sigma^2$ and each $x_1, x_2, \ldots, x_n$ are fixed and known.

Maximising $\log \mathrm{L}(\beta|\mathbf{y})$ is therefore the same as maximising

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

or, equivalently, minimising

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

One option is to take the derivative with respect to $\beta$ and set this equal to 0:

$$\begin{aligned}&\frac{d}{d\beta}\sum_{i=1}^{n}(y_i-\beta x_i)^2 = \sum_{i=1}^{n}\frac{d}{d\beta}(y_i-\beta x_i)^2 = \sum_{i=1}^{n}2(y_i-\beta x_i)(-x_i) = 0\\ &\Rightarrow \sum_{i=1}^{n}(x_i y_i - \beta x_i^2) = 0\\ &\Rightarrow \beta = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}\end{aligned}$$

Therefore, $\widehat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ is a candidate for the maximum likelihood estimate.

But one needs to compute the second derivative at this point to check that $\widehat{\beta}$ indeed minimises the expression $\sum_{i=1}^{n}(y_i-\beta x_i)^2$:

$$\begin{aligned}\frac{d^2}{d\beta^2}\left[\sum_{i=1}^{n}(y_i-\beta x_i)^2\right] &= \frac{d}{d\beta}\left[\sum_{i=1}^{n}2(y_i-\beta x_i)(-x_i)\right] = \left[\sum_{i=1}^{n}2(-x_i)(-x_i)\right]\\ &= 2\sum_{i=1}^{n}x_i^2 > 0,\end{aligned}$$

as required (unless all $x_i = 0$, in which case the value of $\beta$ is meaningless).

However, this only shows that $\widehat{\beta}$ is a local minimum; to show it is a global minimum, we need to check the boundary cases for $\beta$. Since $\beta \in \mathbb{R}$,

$$\lim_{\beta\to\infty}\sum_{i=1}^{n}(y_i-\beta x_i)^2 = \infty, \qquad \lim_{\beta\to-\infty}\sum_{i=1}^{n}(y_i-\beta x_i)^2 = \infty,$$

which shows that $\widehat{\beta}$ is indeed a global minimum for $\sum_{i=1}^{n}(y_i-\beta x_i)^2$.

Therefore $\widehat{\beta}$ is a global maximum for $-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\beta x_i)^2$, which shows that $\widehat{\beta}$ maximises $\log \mathrm{L}(\beta|\mathbf{y})$ and hence $\widehat{\beta}$ maximises $\mathrm{L}(\beta|\mathbf{y})$.

Alternatively, instead of finding the value of $\beta$ that minimises $\sum_{i=1}^n(y_i-\beta x_i)^2$ using calculus (which involves additional checks), one can simply complete the square. Defining

$$Q_y=\sum_{i=1}^n y_i^2, \qquad Q_{xy}=\sum_{i=1}^n x_i y_i, \qquad Q_x=\sum_{i=1}^n x_i^2,$$

$$\begin{aligned}\sum_{i=1}^n(y_i-\beta x_i)^2 &= \sum_{i=1}^n(y_i^2-2\beta x_i y_i+\beta^2 x_i^2)=Q_y-2\beta Q_{xy}+\beta^2 Q_x\\ &= Q_x\left(\beta^2-2\beta\frac{Q_{xy}}{Q_x}\right)+Q_y=Q_x\left(\beta^2-2\beta\frac{Q_{xy}}{Q_x}+\frac{Q_{xy}^2}{Q_x^2}\right)+Q_y-\frac{Q_{xy}^2}{Q_x}\\ &= Q_x\left(\beta-\frac{Q_{xy}}{Q_x}\right)^2+Q_y-\frac{Q_{xy}^2}{Q_x}\end{aligned}$$

which shows that $\beta=\dfrac{Q_{xy}}{Q_x}=\dfrac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$ is a global minimum.

**Part(d):**

Since $\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$ is the maximum likelihood estimate for $\beta$, the maximum likelihood estimator is then

$$\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$$

**Part(e):**

See the solution for Part (c).

## 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 $\beta_0$ and $\beta_1$ 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 $\alpha_0$ and $\alpha_1$ 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.

**Solution to Question 3**

**Part(a):**

From Section 9.2 in the notes,

$$\begin{aligned}
\widehat{\beta}_0 &= \overline{y} - \left(\frac{S_{xy}}{S_{xx}}\right)\overline{x} \\
\widehat{\beta}_1 &= \frac{S_{xy}}{S_{xx}},
\end{aligned}$$

where

We also note that

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} = n\overline{x} - n\overline{x} = 0,$$

and so

$$\begin{aligned}
S_{xy} &= \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \\
&= \sum_{i=1}^{n}(x_i - \overline{x})y_i - \overline{y}\sum_{i=1}^{n}(x_i - \overline{x}) \\
&= \sum_{i=1}^{n}(x_i - \overline{x})y_i - \overline{y} \cdot 0 \\
&= \sum_{i=1}^{n}(x_i - \overline{x})y_i.
\end{aligned}$$

**Part(b):**

Let us define the new constants $z_i = x_i - \overline{x}$, for $i = 1, 2, \ldots, n$. Note that

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} (x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} = n\overline{x} - n\overline{x} = 0,$$

which implies that

$$\overline{z} = \frac{1}{n} \sum_{i=1}^{n} z_i = \frac{1}{n} \cdot 0 = 0.$$

Then the reparametrised model becomes

$$Y_i = \alpha_0 + \alpha_1 z_i + \epsilon_i, \qquad i \in \{1, 2, \ldots, n\},$$

and using the same calculation as for $\widehat{\beta}_0$ and $\widehat{\beta}_1$, we can straight away write down that

$$\begin{aligned} \widehat{\alpha}_0 &= \overline{y} - \left( \frac{S_{zy}}{S_{zz}} \right) \overline{z} \\ \widehat{\alpha}_1 &= \frac{S_{zy}}{S_{zz}}, \end{aligned}$$

where

$$\begin{aligned} S_{zz} &= \sum_{i=1}^{n} (z_i - \overline{z})^2 \\ &= \sum_{i=1}^{n} z_i^2 \qquad \text{(using } \overline{z} = 0\text{, shown above)}, \\ &= \sum_{i=1}^{n} (x_i - \overline{x})^2 \\ \Rightarrow S_{zz} &= S_{xx}, \end{aligned}$$

$$\begin{aligned} S_{zy} &= \sum_{i=1}^{n} (z_i - \overline{z})(y_i - \overline{y}) \\ &= \sum_{i=1}^{n} (z_i - \overline{z}) y_i \qquad \text{(using Part (a))} \\ &= \sum_{i=1}^{n} z_i y_i \qquad \text{(using } \overline{z} = 0\text{, shown above)} \\ &= \sum_{i=1}^{n} (x_i - \overline{x}) y_i \\ \Rightarrow S_{zy} &= S_{xy}. \end{aligned}$$

Now, we can show

$$\widehat{\alpha}_1 = \frac{S_{zy}}{S_{zz}} = \frac{S_{xy}}{S_{xx}} = \widehat{\beta}_1.$$

**Part(c):**

In Part (b) it was mentioned that using the same calculation for $\widehat{\beta}_0$, one can show that

$$\widehat{\alpha}_0 = \overline{y} - \left(\frac{S_{zy}}{S_{zz}}\right)\overline{z}.$$

However, also in Part (b) it was shown that $\overline{z} = 0$, so

$$\widehat{\alpha}_0 = \overline{y} - \left(\frac{S_{zy}}{S_{zz}}\right) \cdot 0 = \overline{y},$$

as required.

**Part (d)**

Setting $\widehat{\alpha}_0 = \widehat{\beta}_0$, this implies

$$\begin{aligned} \overline{y} &= \overline{y} - \left(\frac{S_{xy}}{S_{xx}}\right)\overline{x} \\ \Rightarrow &\left(\frac{S_{xy}}{S_{xx}}\right)\overline{x} = 0 \\ \Rightarrow &S_{xy} = 0 \text{ or } \overline{x} = 0. \end{aligned}$$

Note that $S_{xy} = 0 \Leftrightarrow r^2 = 0$, where $r^2$ is the sample correlation.

So, $\widehat{\alpha}_0 = \widehat{\beta}_0$ implies that either the sample correlation of the $x_i$ and $y_i$ values is equal to 0, or $\overline{x} = 0$.

Equivalently, if $\overline{x} \neq 0$ and $S_{xy} \neq 0$, then $\widehat{\alpha}_0 \neq \widehat{\beta}_0$.

**Part (e)**

Replacing the sample $y_i$ values with the random variables $Y_i$, one has

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{\mathbf{0}} &= \overline{Y} \\ \widehat{\boldsymbol{\alpha}}_{\mathbf{1}} &= \frac{S_{zY}}{S_{zz}} = \frac{\sum_{i=1}^{n} z_i Y_i}{\sum_{i=1}^{n} z_i^2} \end{aligned}$$

**Part (f)**

Showing the correlation of two random variables is 0 is equivalent to showing the covariance is 0:

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}, \widehat{\boldsymbol{\alpha}}_{\mathbf{1}}\right) = \mathrm{Cov}\left(\overline{Y}, \frac{\sum_{i=1}^{n} z_i Y_i}{\sum_{i=1}^{n} z_i^2}\right) = \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \mathrm{Cov}\left(\overline{Y}, \sum_{i=1}^{n} z_i Y_i\right) = \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \sum_{i=1}^{n} z_i \mathrm{Cov}\left(\overline{Y}, Y_i\right)$$

Now recalling that $\overline{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$, and again using the bilinearity of the covariance function,

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}, \widehat{\boldsymbol{\alpha}}_{\mathbf{1}}\right) = \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \sum_{i=1}^{n} z_i \mathrm{Cov}\left(\frac{1}{n}\sum_{j=1}^{n} Y_j, Y_i\right) = \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \frac{1}{n} \sum_{i=1}^{n} z_i \sum_{j=1}^{n} \mathrm{Cov}\left(Y_j, Y_i\right)$$

And now noting that $\mathrm{Cov}\left(Y_j, Y_i\right) = 0$ when $i \neq j$, since the random variables $Y_i$ are independent, since the errors $\epsilon_i$ were assumed to be independent,

$$\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}, \widehat{\boldsymbol{\alpha}}_{\mathbf{1}}\right) &= \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \frac{1}{n} \sum_{i=1}^{n} z_i \mathrm{Cov}\left(Y_i, Y_i\right) \\
&= \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \frac{1}{n} \sum_{i=1}^{n} z_i \mathrm{Var}\left(Y_i\right) \\
&= \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \frac{1}{n} \sum_{i=1}^{n} \sigma^2 z_i \\
&= \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \frac{\sigma^2}{n} \sum_{i=1}^{n} z_i \\
&= \left(\frac{1}{\sum_{i=1}^{n} z_i^2}\right) \frac{\sigma^2}{n} \cdot 0 \\
\Rightarrow \mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}, \widehat{\boldsymbol{\alpha}}_{\mathbf{1}}\right) &= 0,
\end{aligned}$$

using $\sum_{i=1}^{n} z_i = 0$, which was shown in Part (a).