Question 1

Suppose that the lifetime of a particular lightbulb is known to be distributed as $\text{Exp}(\theta)$, i.e. an exponential distribution with parameter θ. Suppose that a group of n friends bought a multipack containing n of these lightbulbs and each person keeps one lightbulb to use at home. In a few years' time, they get together and share how long their lightbulbs lasted, and these measurements (lifetimes) are written down as $x_1, x_2, \ldots, x_n$.

(a) Write down the probability density function of the $\text{Exp}(\theta)$ distribution.

(b) Given the sample of measurements $x_1, x_2, \ldots, x_n$, write down the likeihood function for θ based on these measurements.

(c) Write down the log-likelihood of θ given the measurements $x_1, x_2, \ldots, x_n$.

(d) Find the maximum likelihood estimate $\widehat{\theta}$.

(e) Are you sure that $\widehat{\theta}$ is a maximum, or could it be a minimum? If you have not already done so in (d), provide proof that $\widehat{\theta}$ is a maximum/minimum.

Solution to Question 1

Part (a):

Usually it is written with parameter $\lambda > 0$:

$$f(x) = \lambda e^{-\lambda x}$$

But we shall write it with parameter $\theta > 0$:

$$f_\theta(x) = \theta e^{-\theta x}$$

Part (b):

We can assume that each of the lightbulbs have an independent lifetime, and therefore the joint p.d.f. (and therefore joint likelihood) is a product of the individual p.d.f.'s (and likelihoods). Since the likelihood for θ given the sample point x_i ($i \in \{1, 2, \ldots, n\}$) is

$$L(\theta|x_i) = \theta e^{-\theta x_i} = \theta \exp(-\theta x_i),$$

(it will be more useful to use the notation using the function $\exp(\cdot)$), the joint likelihood is

$$L(\theta|\mathbf{x}) = L(\theta|x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \theta \exp(-\theta x_i) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) = \theta^n \exp(-\theta n\overline{x})$$

Part (c):

Taking logs of both sides of the equation above, we have

$$\begin{aligned}
\log L(\theta|\mathbf{x}) &= \log(\theta^n \exp(-\theta n\overline{x})) \\
&= \log(\theta^n) + \log(\exp(-\theta n\overline{x})) \\
&= n \log \theta - \theta n\overline{x}
\end{aligned}$$

Part (d):

Maximising the log-likelihood is equivalent to maximising the likelihood, and the log-likelihood is in a simpler form. Therefore, taking the derivative with respect to θ,

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log L(\theta|\mathbf{x}) = \frac{n}{\theta} - n\overline{x}$$

Setting the derivative to 0,

$$\begin{aligned}
&\frac{n}{\theta} - n\overline{x} = 0 \\
\Rightarrow &\frac{n}{\theta} = n\overline{x} \\
\Rightarrow &\frac{1}{\theta} = \overline{x} \\
\Rightarrow &\theta = (\overline{x})^{-1}
\end{aligned}$$

At this point, we do not know if this value for θ is a maximum or a minimum. We therefore need to take the second derivative:

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log L(\theta|\mathbf{x}) = -\frac{n}{\theta^2}$$

which is negative for all values of $\mathbf{x}$, because it does not even depend on $\mathbf{x}$. Therefore, $\widehat{\theta} = (\overline{x})^{-1}$ is the maximum likelihood estimate.

Part (e)

It is a maximum; in (d) it was shown that the second derivative of the log-likelihood is negative.

Question 2 (Knowledge of partial derivatives required)

Suppose the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are assumed to be independent and identically distributed as $\mathrm{N}\left(\mu, \sigma^2\right)$, where μ and σ^2 are unknown. Suppose further that $\mathbf{X}$ is observed as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

(a) Compute the likelihood $L(\mu, \sigma^2|\mathbf{x})$.

(b) Compute the log-likelihood $\log L(\mu, \sigma^2|\mathbf{x})$.

(c) Compute the partial derivative $\frac{\partial}{\partial \mu} \log L(\mu, \sigma^2|\mathbf{x})$.

(d) Set the partial derivative in (c) equal to 0 and solve for μ. Show that this value of μ maximises $L(\mu, \sigma^2|\mathbf{x})$ globally for fixed σ^2.

(e) Compute the partial derivative $\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2|\mathbf{x})$. **Hint**: it may help to set $z = \sigma^2$ and then compute the partial derivative with respect to z.

(f) Set the partial derivative in (e) equal to 0 and solve for σ^2. Show that this value of σ^2 is a (local) maximum for $L(\mu, \sigma^2|\mathbf{x})$, for fixed μ.

(g) Show that the value of σ^2 found in (f) maximises the likelihood $L(\mu, \sigma^2|\mathbf{x})$ globally for fixed μ.

(h) Write down the maximum likelihood estimators for μ and σ^2 given the random variables $\mathbf{X}$.

Solution to Question 2 (Knowledge of partial derivatives required)

Part (a):

The probability density function for each X_i is

$$f(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right).$$

Since the X_i are assumed to be independent, the joint distribution of $\mathbf{X}$ is

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i|\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) \end{aligned}$$

Therefore the likelihood of μ and σ^2 given $\mathbf{x}$ is

$$L(\mu, \sigma^2|\mathbf{x}) = f(\mathbf{x}|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

Part (b):

$$\begin{aligned}
\log L(\mu, \sigma^2|\mathbf{x}) &= \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\right) + \log\left(\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right) \\
&= \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\right) + \left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \\
&= \log\left(\left(2\pi\sigma^2\right)^{-n/2}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
\Rightarrow \log L(\mu, \sigma^2|\mathbf{x}) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}$$

Part (c):

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log L(\mu, \sigma^2|\mathbf{x}) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)(-1) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)
\end{aligned}$$

Part (d):

$$\begin{aligned}
&\frac{\partial}{\partial\mu}\log L(\mu, \sigma^2|\mathbf{x}) = 0 \\
\Rightarrow &\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \\
\Rightarrow &\sum_{i=1}^{n}(x_i-\mu) = 0 \\
\Rightarrow &\mu = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{x}
\end{aligned}$$

To show this $\mu = \overline{x}$ maximises the likelihood globally, recall Exercise 1.2.10 which showed that for any $\mu \in \mathbb{R}$,

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

Therefore, for any $\sigma^2 > 0$,

$$
\begin{aligned}
&-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2 \geq -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&\Rightarrow \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right) \geq \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \\
&\Rightarrow \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right) \geq \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \\
&\Rightarrow L(\overline{x},\sigma^2|\mathbf{x}) \geq L(\mu,\sigma^2|\mathbf{x})
\end{aligned}
$$

Part (e):

One can compute the partial derivative with respect to σ^2 directly, but one needs to be careful not to compute the derivative with respect to σ (by accident). Solving for σ^2 for either $\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma^2|\mathbf{x}) = 0$ or $\frac{\partial}{\partial\sigma}\log L(\mu,\sigma^2|\mathbf{x}) = 0$ results in the same answer, but the second option is slightly more work, requiring the Chain Rule. Although we shall compute it directly below, if any of the steps are unclear, try setting $z = \sigma^2$ and then computing the partial derivative with respect to z.

The log-likelihood $\log L(\mu,\sigma^2|\mathbf{x})$ is

$$
\log L(\mu,\sigma^2|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2
$$

Part (f):

Now,

$$
\begin{aligned}
\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma^2|\mathbf{x}) &= 0 - \frac{n}{2}\cdot\frac{1}{\sigma^2} - \frac{1}{2}\cdot\frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}
$$

Setting this partial derivative equal to 0,

$$
\begin{aligned}
&\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma^2|\mathbf{x}) = 0 \\
&\Rightarrow -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \\
&\Rightarrow -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0 \\
&\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}
$$

Since the value $\mu = \overline{x}$ maximised the likelihood for any value σ^2, we set $\mu = \overline{x}$ here, so

$$
\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2.
$$

To show this is a local maximum, we need to compute the second derivative $\frac{\partial^2}{\partial(\sigma^2)^2}\log L(\mu,\sigma^2|\mathbf{x})$ and evaluate it at $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2$. In this case, it may be simpler to substitute values $z=\sigma^2$ first. Since

$$\log L(\mu,\sigma^2|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2$$
$$\Rightarrow \log L(\overline{x},z|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(z) - \frac{1}{2z}\sum_{i=1}^n (x_i-\overline{x})^2$$

And setting $A=\sum_{i=1}^n (x_i-\overline{x})^2$,

$$\log L(\overline{x},z|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(z) - \frac{A}{2z}$$

Now, we compute the second partial derivative with respect to z:

$$\frac{\partial^2}{\partial z^2}\log L(\overline{x},z|\mathbf{x}) = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial z}\left[-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(z) - \frac{A}{2z}\right]\right)$$

$$= \frac{\partial}{\partial z}\left(-\frac{n}{2z} + \frac{A}{2z^2}\right)$$

$$= \frac{n}{2z^2} + \frac{A}{2z^3}(-2)$$

$$\Rightarrow \frac{\partial^2}{\partial z^2}\log L(\overline{x},z|\mathbf{x}) = \frac{n}{2z^2} - \frac{A}{z^3}$$

Evaluating this second derivative at $z=\sigma^2=\frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2 = \frac{A}{n}$,

$$\left.\frac{\partial^2}{\partial z^2}\log L(\overline{x},z|\mathbf{x})\right|_{z=\frac{A}{n}} = \frac{n}{2(A/n)^2} - \frac{A}{(A/n)^3}$$

$$= \frac{n}{2}\cdot\frac{n^2}{A^2} - \frac{An^3}{A^3}$$

$$= \frac{n^2}{2A^2} - \frac{n^3}{A^2} = -\frac{n^2}{2A^2} < 0$$

since $A>0$, if we assume the x_i are not all equal. This shows that $\log L(\overline{x},\sigma^2|\mathbf{x})$ has a local maximum at $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2$.

Part (g):

We have already shown that $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2$ is a local maximum of $\log L(\overline{x},\sigma^2|\mathbf{x})$, and is therefore also a maximum of $L(\overline{x},\sigma^2|\mathbf{x})$. Furthermore, it is the only possible local maximum, since it is the only value of σ^2 which satisfies $\frac{\partial}{\partial\sigma^2}\log L(\overline{x},\sigma^2|\mathbf{x})=0$. However, we still need to check the boundary points for the range of σ^2, which is $(0,\infty)$. Setting $z=\sigma^2$ again and $A=\sum_{i=1}^n (x_i-\overline{x})^2$, this is equivalent to computing the limits

$$\lim_{z\to 0^+} z^{-n/2}\exp\left(-\frac{A}{2z}\right)$$

$$\lim_{z\to\infty} z^{-n/2}\exp\left(-\frac{A}{2z}\right)$$

(We ignore the coefficient of $(2\pi)^{-n/2}$ for now.) For the first limit, we can set $y = \frac{1}{z}$, and then consider the limit $y \to \infty$, and use L'Hôpital's rule:

$$\lim_{z\to 0^+} z^{-n/2} \exp\left(-\frac{A}{2z}\right) \lim_{y\to\infty} y^{n/2} \exp\left(-\frac{Ay}{2}\right) \lim_{y\to\infty} \frac{y^{n/2}}{\exp\left(\frac{Ay}{2}\right)}$$

There are two cases to consider for n, which is a positive integer.. Either n is an even integer, or n is an odd integer. If n is even, $n = 2k$ for some integer k, and then

$$\begin{aligned}
\lim_{y\to\infty} \frac{y^{n/2}}{\exp\left(\frac{Ay}{2}\right)} = \lim_{y\to\infty} \frac{y^{k}}{\exp\left(\frac{Ay}{2}\right)} &= \lim_{y\to\infty} \frac{ky^{k-1}}{\frac{A}{2}\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule)} \\
&= \lim_{y\to\infty} \frac{k(k-1)y^{k-2}}{\left(\frac{A}{2}\right)^2\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule a second time)} \\
&\vdots \\
&= \lim_{y\to\infty} \frac{k!}{\left(\frac{A}{2}\right)^k\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule } k \text{ times)} \\
&= 0
\end{aligned}$$

If n is odd, the result is similar. Set $n = 2k-1$ for some positive integer k. Then

$$\begin{aligned}
\lim_{y\to\infty} \frac{y^{n/2}}{\exp\left(\frac{Ay}{2}\right)} = \lim_{y\to\infty} \frac{y^{k-1/2}}{\exp\left(\frac{Ay}{2}\right)} &= \lim_{y\to\infty} \frac{(k-1/2)y^{k-1-1/2}}{\frac{A}{2}\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule)} \\
&= \lim_{y\to\infty} \frac{(k-1/2)(k-3/2)y^{k-2-1/2}}{\left(\frac{A}{2}\right)^2\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule a second time)} \\
&\vdots \\
&= \lim_{y\to\infty} \frac{(k-1/2)(k-3/2)\cdots(1/2)y^{-1/2}}{\left(\frac{A}{2}\right)^k\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule } k \text{ times)} \\
&= \lim_{y\to\infty} \frac{(k-1/2)(k-3/2)\cdots(1/2)}{y^{1/2}\left(\frac{A}{2}\right)^k\exp\left(\frac{Ay}{2}\right)} \\
&= 0
\end{aligned}$$

Therefore, we have:

$$\lim_{z\to 0^+} z^{-n/2} \exp\left(-\frac{A}{2z}\right) = 0$$

For the second limit,

$$\lim_{z\to\infty} z^{-n/2} \exp\left(-\frac{A}{2z}\right) = \lim_{z\to\infty} \frac{\exp\left(-\frac{A}{2z}\right)}{z^{n/2}} = \frac{\lim\limits_{z\to\infty} \exp\left(-\frac{A}{2z}\right)}{\lim\limits_{z\to\infty} z^{n/2}} = \frac{\exp(0)}{\lim\limits_{z\to\infty} z^{n/2}} = \frac{1}{\lim\limits_{z\to\infty} z^{n/2}} = 0$$

Since for all $\sigma^2 > 0$,

$$L(\overline{x}, \sigma^2|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right) > 0,$$

this implies that $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ maximises $L(\overline{x}, \sigma^2|\mathbf{x})$ globally.

Overall, we can now state that the likelihood $L(\mu, \sigma^2|\mathbf{x})$ is maximised when we use the following maximum likelihood estimates for μ and σ^2:

$$\widehat{\mu}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\widehat{\sigma^2}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right).$$

Part (h):

Given the answers summarised in Part (g), the maximum likelihood estimators of μ and σ^2 given the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are

$$\widehat{\mu}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},$$

$$\widehat{\sigma^2}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = S_b^2.$$

Question 3

The data set `deliverycosts.txt` consists of 20 values in two columns named `weight` and `cost`. The `weight` column represents the weight (in kg) of a parcel and `cost` represents the cost (in British pounds) for delivering a package of that given weight from London to Edinburgh via an unnamed courier company.

Create an R Markdown document to answer the following questions:

(a) Read in the data contained in `deliverycosts.txt`, and plot `cost` vs `weight` using a scatterplot. Does there appear to be a linear relationship between the two variables?

(b) Use the `lm` function to fit a linear model on `weight` and `cost`, where `weight` is the predictor variable and `cost` is the regressor variable. Plot the line showing the estimated linear relationship. **Note:** you may wish to transform or rescale `weight` and/or `cost` before fitting the linear model.

(c) Using any output from the `lm` function as well as any additional plots, briefly discuss how well the data fits the proposed model.

(d) Given the results in Steps 2 and 3, if you think a better linear model can be found, perhaps by modifying the predictor/regressor variables and the data set, do Steps 2 and 3 again for your modified linear model. If you feel there are any unusual data points, provide justification.

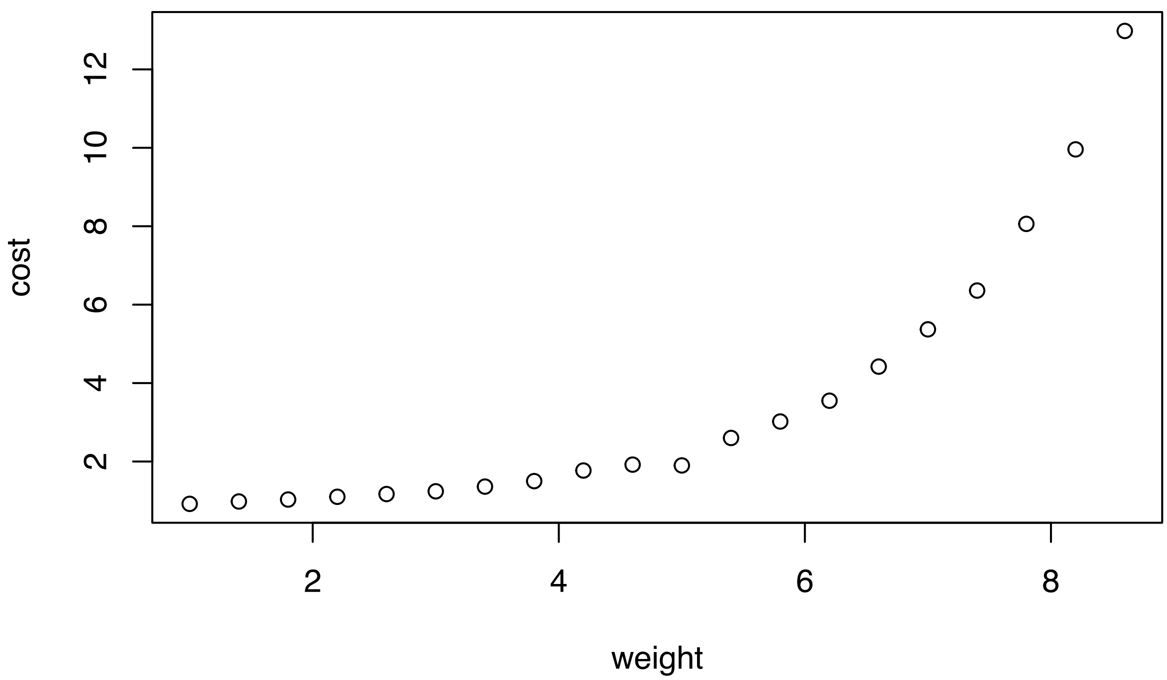
(e) Describe the results of your investigation into the relationship between `weight` and `cost`, and describe the relationship (if any) that exists between the two variables. Provide an interpretation of any estimated parameter values for the linear model.

Problem Sheet 14, Question 3

Legendre, Adrien-Marie, CID: 10101010

Question 3(a)

```
# reading in the data into a data frame
df <- read.table("deliverycosts.txt", sep=",", header=T)

# extracting the columns of the data frame to vectors
weight <- df$weight
cost <- df$cost

# plotting the data
plot(weight, cost)
```



The model seems nonlinear, as the slope seems to change.

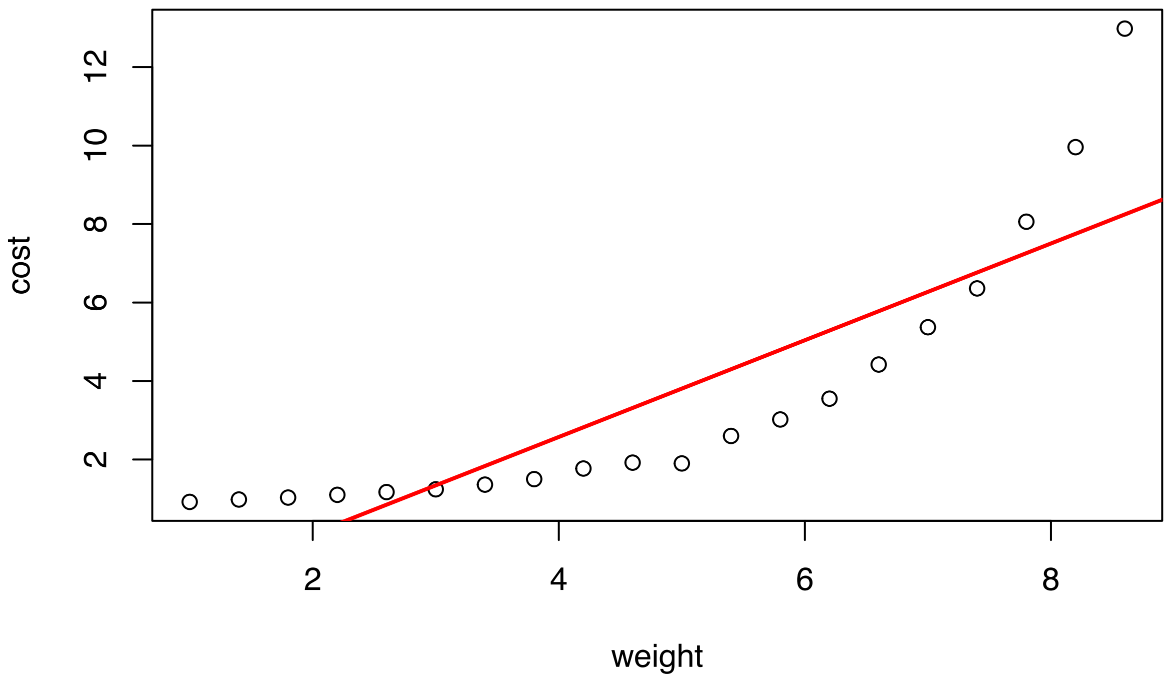
Question 3(b)

Option 1: no transformation

Model 1: Fitting `weight` and `cost` without any transformation, i.e.

$$cost = \beta_0 + \beta_1 weight + error$$

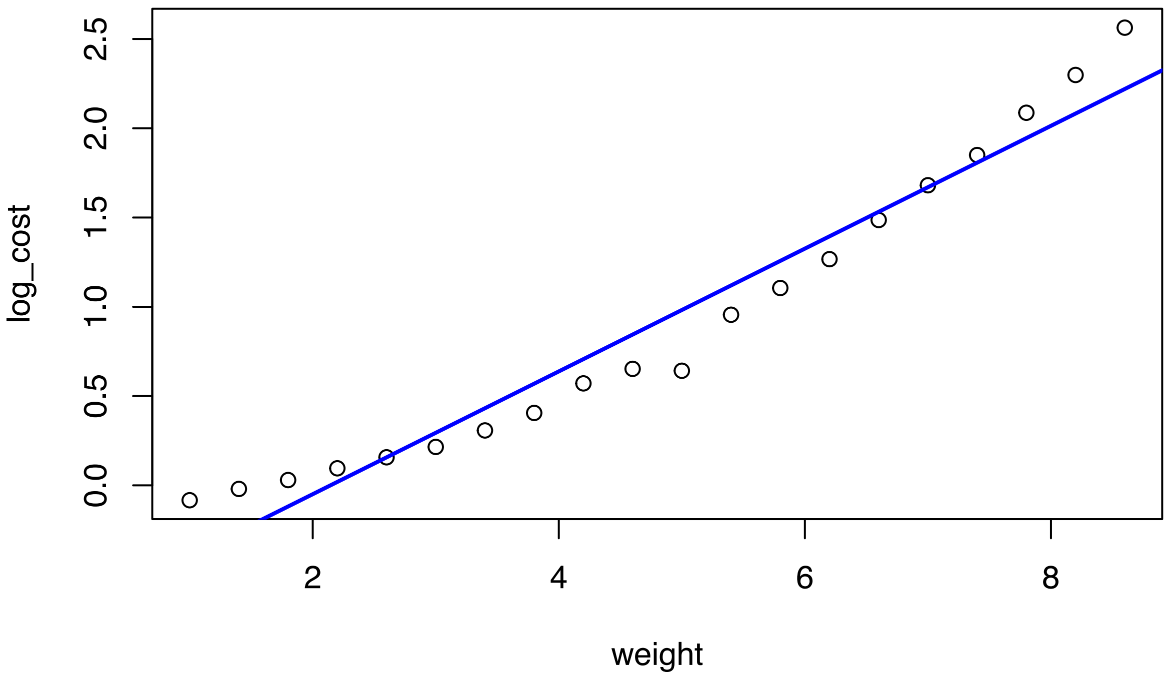
```
# using the lm function to compute the coefficients of the linar model
# between cost and weight
model_1 <- lm(cost~weight)

# obtaining the coefficients from the model object
beta0hat <- model_1$coefficients[1]
beta1hat <- model_1$coefficients[2]

# plotting the data
plot(weight, cost)

# adding the line showing the best linear fit
abline(a=beta0hat, b=beta1hat, col="red", lwd=2)
```



Option 2: log(cost) vs weight

Or fitting we could try Model 2: fitting `weight` and log(`cost`):

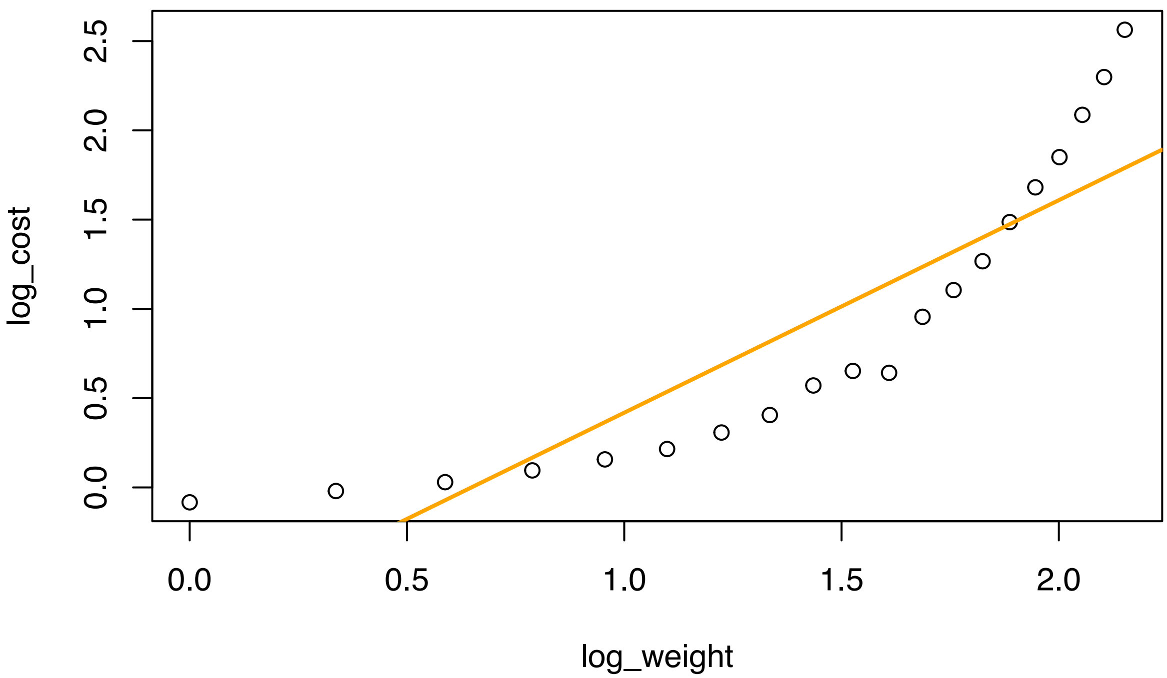
```
# taking the log of the cost
log_cost <- log(cost)

# computing the linear model coefficients using lm
model_2 <- lm(log_cost~weight)

# obtaining the coefficients from the model object
beta0hat <- model_2$coefficients[1]
beta1hat <- model_2$coefficients[2]

# plotting the data
plot(weight, log_cost)

# adding the line showing the best linear fit
abline(a=beta0hat, b=beta1hat, col="blue", lwd=2)
```



Option 3: log(cost) vs log(weight)

Or fitting we could try Model 3: fitting log(`weight`) and log(`cost`):

```
# taking the log of the cost
log_cost <- log(cost)
log_weight <- log(weight)

# computing the linear model coefficients using lm
model_2 <- lm(log_cost~log_weight)

# obtaining the coefficients from the model object
beta0hat <- model_2$coefficients[1]
beta1hat <- model_2$coefficients[2]

# plotting the data
plot(log_weight, log_cost)

# adding the line showing the best linear fit
abline(a=beta0hat, b=beta1hat, col="orange", lwd=2)
```

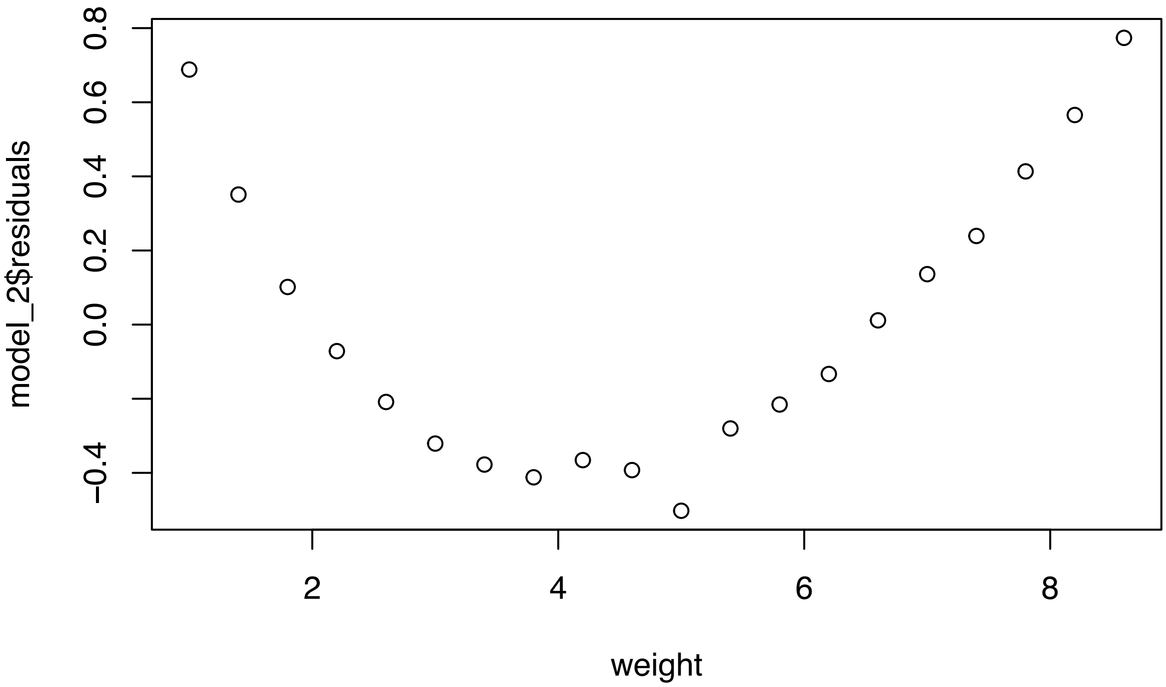


Question 3(c)

For Model 2 (plot above), although the points are close to the regression line, the fit is not perfect because for very small and very large weights the `log(cost)` is above the regression line, but for moderate values of weight the `log(cost)` is below the regression line.

If we look at the residual plot of Model 2:

```
plot(weight, model_2$residuals)
```

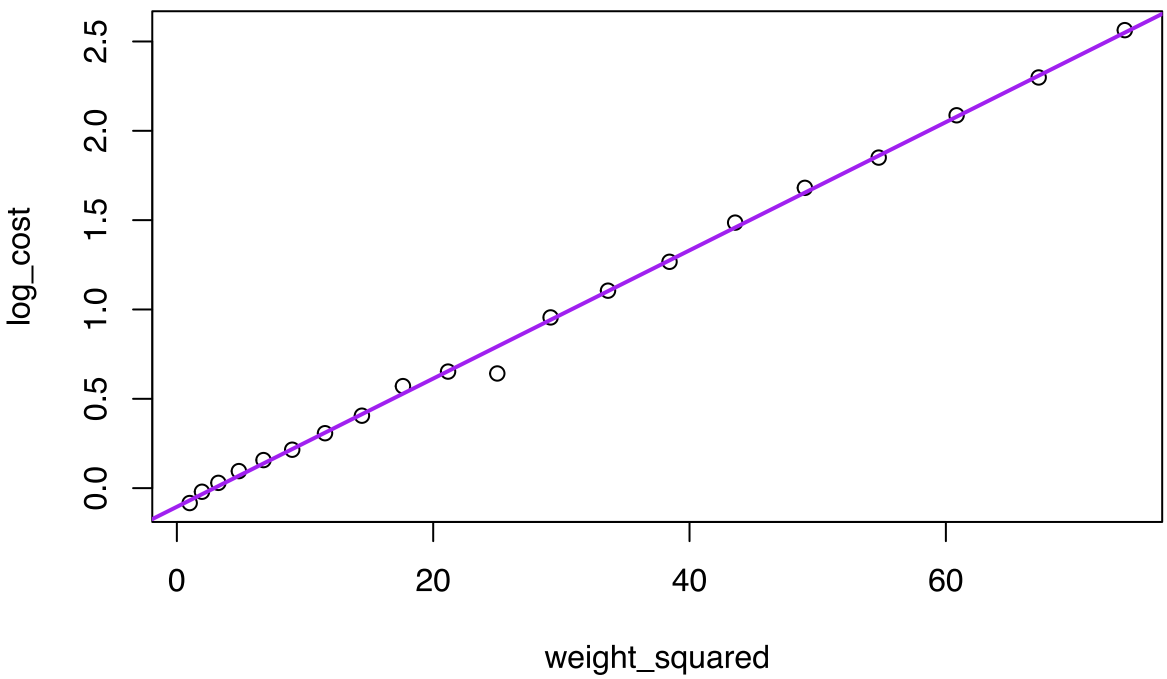


This confirms our observation above, since the residuals are in the form of a "U"-shape, which means that our model may be incorrect.

Question 3(d)

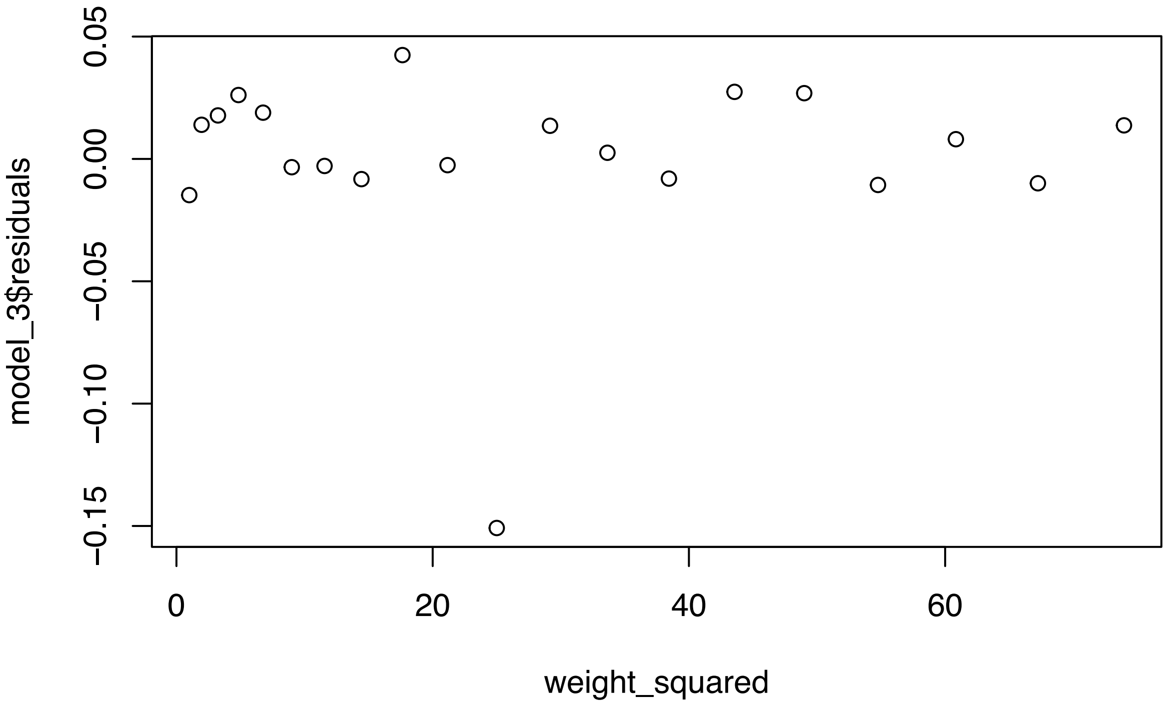
Let us try another model. Model 3: `weight^2` and log(`cost`), i.e. the logarithm of the cost and the square of the weight.

```
# taking the log of the cost
log_cost <- log(cost)

# taking the square of the weight
weight_squared <-weight^2

# using the lm function and obtaining coefficients, and plotting
model_3 <- lm(log_cost~weight_squared)
beta0hat <- model_3$coefficients[1]
beta1hat <- model_3$coefficients[2]
plot(weight_squared, log_cost)
abline(a=beta0hat, b=beta1hat, col="purple", lwd=2)
```



We also plot the residuals:

```
# We also plot the residuals
plot(weight_squared, model_3$residuals)
```

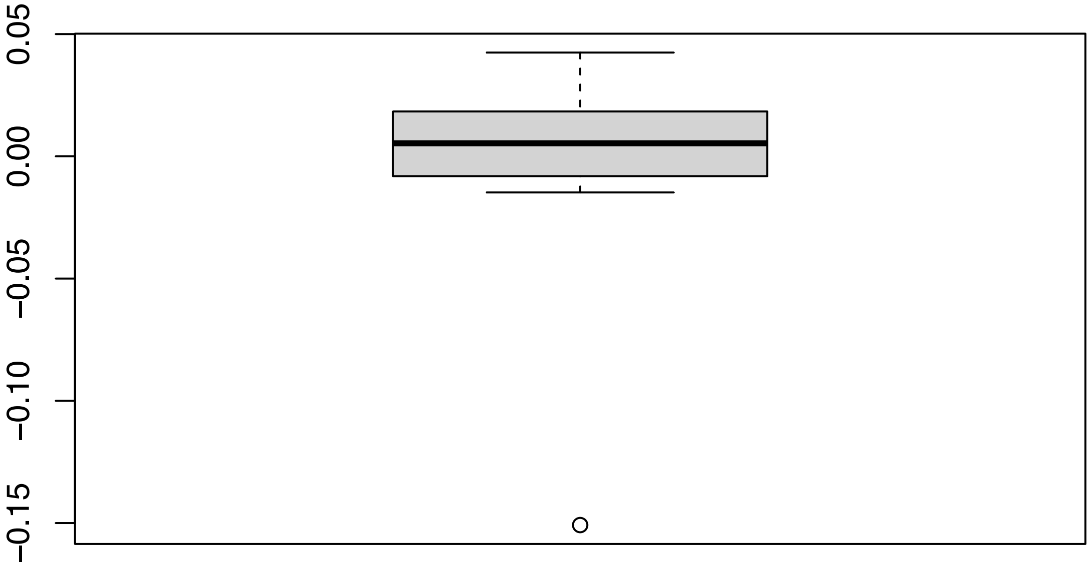


The residual plot looks much better, as the residuals appear to be centred around 0.

However, we notice that there is one data point that does not seem to fit the line and gives a larger/smaller residual than the others.

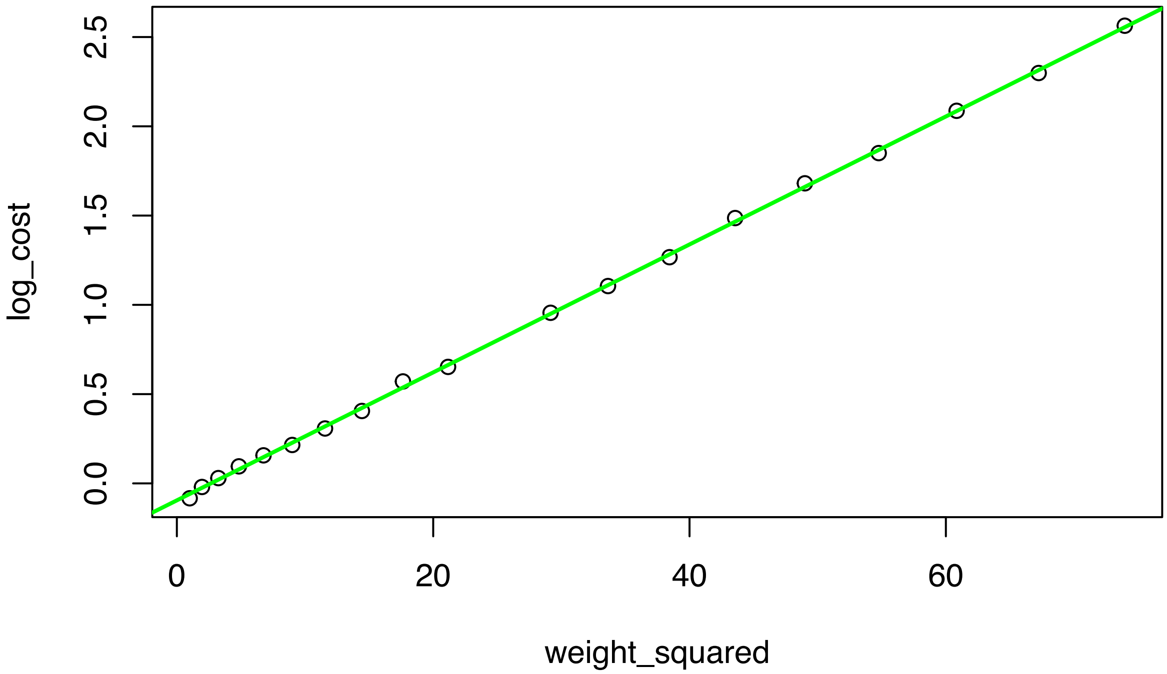
If we plot the residuals using a boxplot, this point appears to be an outlier, following Tukey's recommendation for outliers.

```
boxplot(model_3$residuals)
```



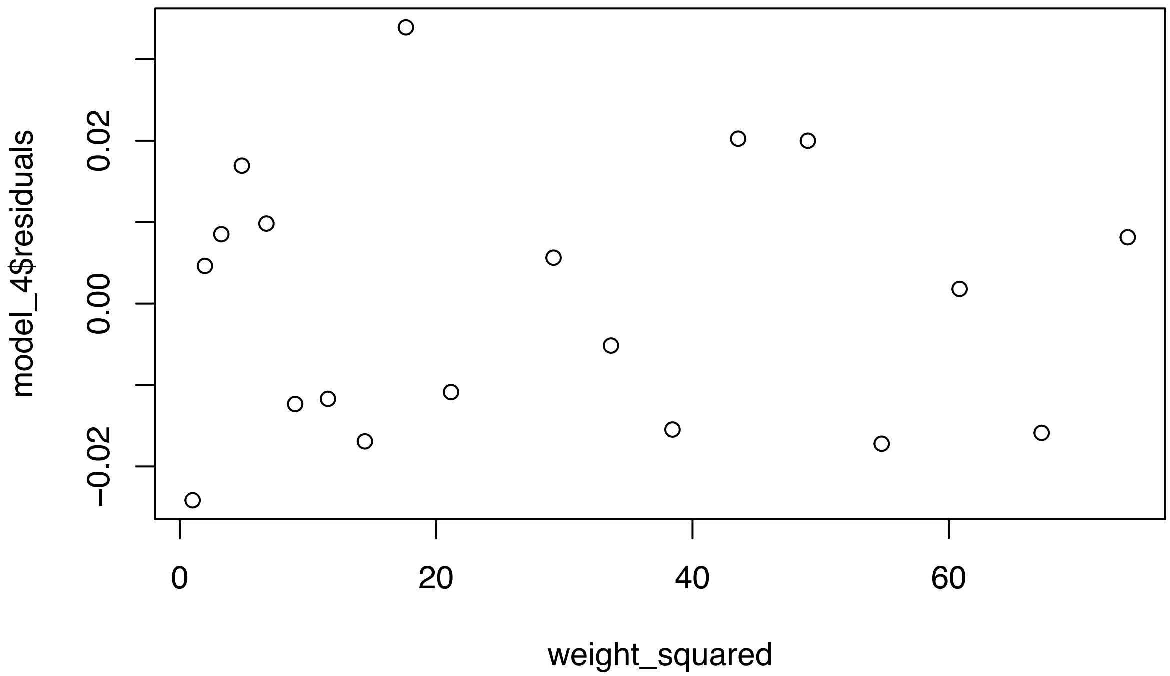
We can try removing this value (in this case it appears to be the 11th value) and rerun the analysis. This will be Model 4.

```
log_cost <- log_cost[-11]
weight_squared <-weight_squared[-11]
model_4 <- lm(log_cost~weight_squared)
beta0hat <- model_4$coefficients[1]
beta1hat <- model_4$coefficients[2]
plot(weight_squared, log_cost)
abline(a=beta0hat, b=beta1hat, col="green", lwd=2)
```



And we see these residuals, with the outlier removed, appear to be independently normally distributed.

```
plot(weight_squared, model_4$residuals)
```



Question 3(e)

From the residual plot for Model 3, it appears that the model of

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where Y is the logarithm of the cost and X is the weight squared, seems to be a good fit. This is because the residuals are small, and they seem to be normally distributed around 0 without any obvious pattern.

Clearly, as the weight increases, so does the cost (we did not need regression to see this!). However, it appears that there is a linear relationship between the logarithm of the cost and the square of the weight.

It seems that as the weight of the package increases, so does the cost, and the logarithm of the cost increases as a linear function of the square of the weight.

If we let the cost be C and the weight be W, then the model becomes $\log(C) = \beta_0 + \beta_1 W^2 + \epsilon$.

Rearranging this, $C = e^{\beta_0} \cdot e^{\beta_1 W^2} \cdot e^{\epsilon}$.

There is some multiplicative effect between the intercept β_0 (a fixed cost?) and a function of the weight squared.