MATH40005 Coursework

Last, First CID: 12345678

Step 1: Reading in the Data (4 marks)

Use the setwd and getwd to find the file of the data,then use the read.table to read the data. Plot the graph y againts x to find the shape of the graph then we can guess the linear model

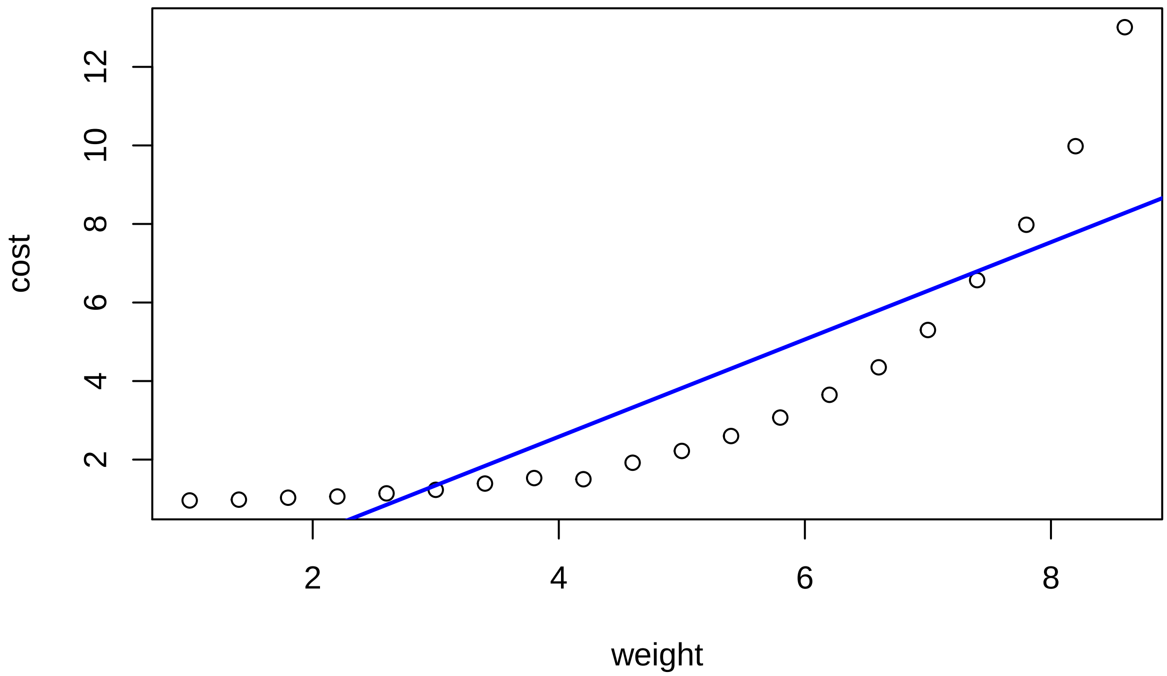
```
setwd("~/Desktop")
getwd()
```

```
## [1] "/Users/appler/Desktop"
```

```
df<-read.table("~/Desktop/data01712390.txt",sep=",",header=TRUE)
df
```

```
##    weight  cost
## 1     1.0  0.96
## 2     1.4  0.98
## 3     1.8  1.03
## 4     2.2  1.06
## 5     2.6  1.14
## 6     3.0  1.23
## 7     3.4  1.39
## 8     3.8  1.53
## 9     4.2  1.50
## 10    4.6  1.92
## 11    5.0  2.22
## 12    5.4  2.60
## 13    5.8  3.07
## 14    6.2  3.65
## 15    6.6  4.35
## 16    7.0  5.30
## 17    7.4  6.57
## 18    7.8  7.98
## 19    8.2  9.98
## 20    8.6 13.01
```

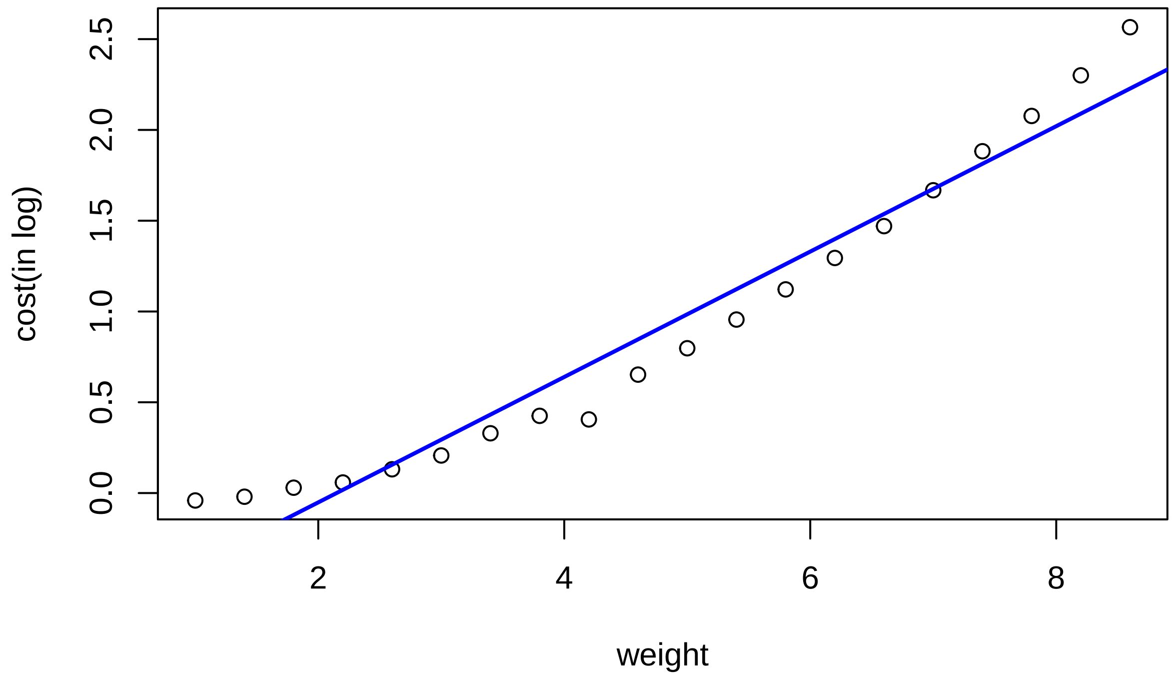
```
xlab="weight"
ylab="cost"
x<-df$weight
y<-df$cost
plot(x,y,xlab=xlab,ylab=ylab)
model<-lm(y~x)
abline(a=model$coefficients[1],b=model$coefficients[2],lwd=2,col="blue")
```



Step 2: Fitting a linear model (4 marks)

Then plot the graph log(y) against x,define the model of y with x the use the abline to find the straight line of linear relationship in this graph.

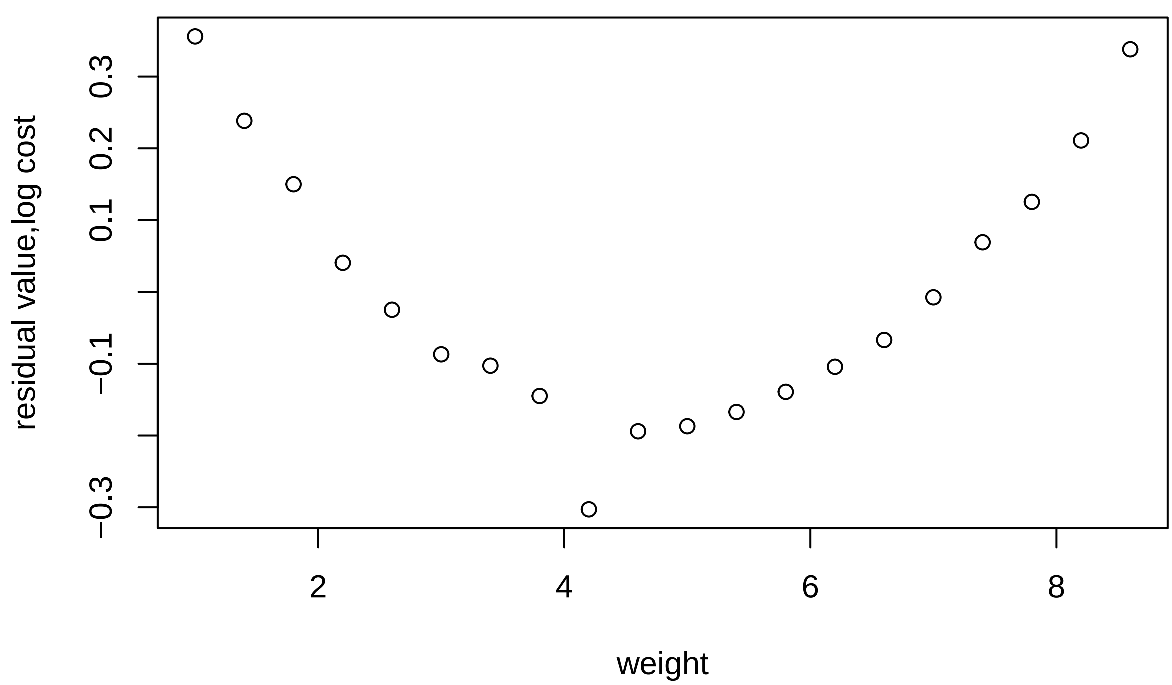
```
xlab="weight"
ylab="cost(in log)"
x<-df$weight
y<-log(df$cost)
plot(x,y,xlab=xlab,ylab=ylab)
model<-lm(y~x)
abline(a=model$coefficients[1],b=model$coefficients[2],lwd=2,col="blue")
```



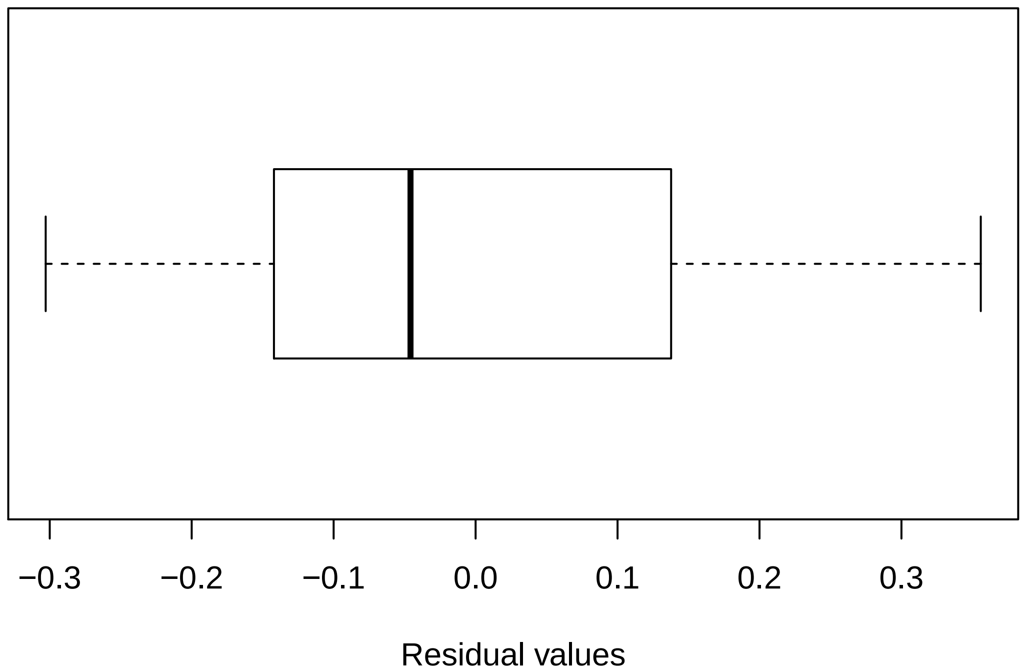
Step 3: Analysis of results (4 marks)

Use the residuals graph of linear model to find whether the points are randomly distributed around 0. Use the boxplot to see if some residuals are outliers.

```
plot(x,model$residuals,xlab="weight",ylab="residual value,log cost")
```




```
boxplot(model$residuals,horizontal=TRUE,xlab="Residual values")
```

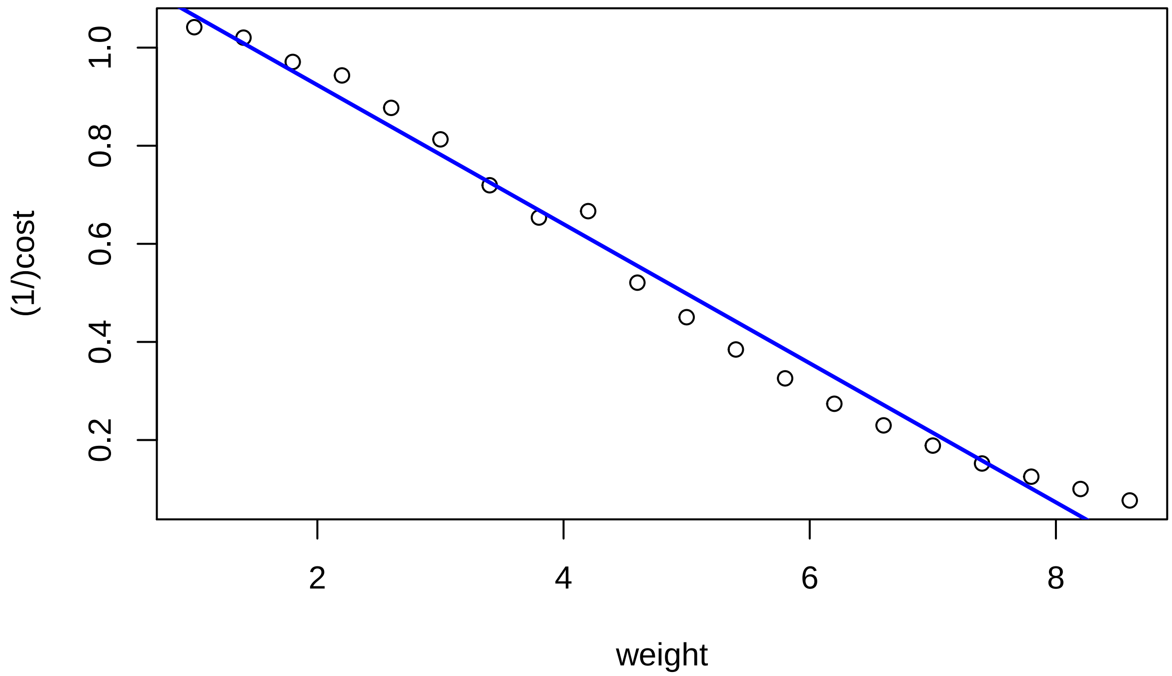


Most of the points are near the 0 point ,this is the evidence to the normal distribution.There is no outlier,but the most of the points are within the 0.3 and -0.3 so the fit is not good enough.

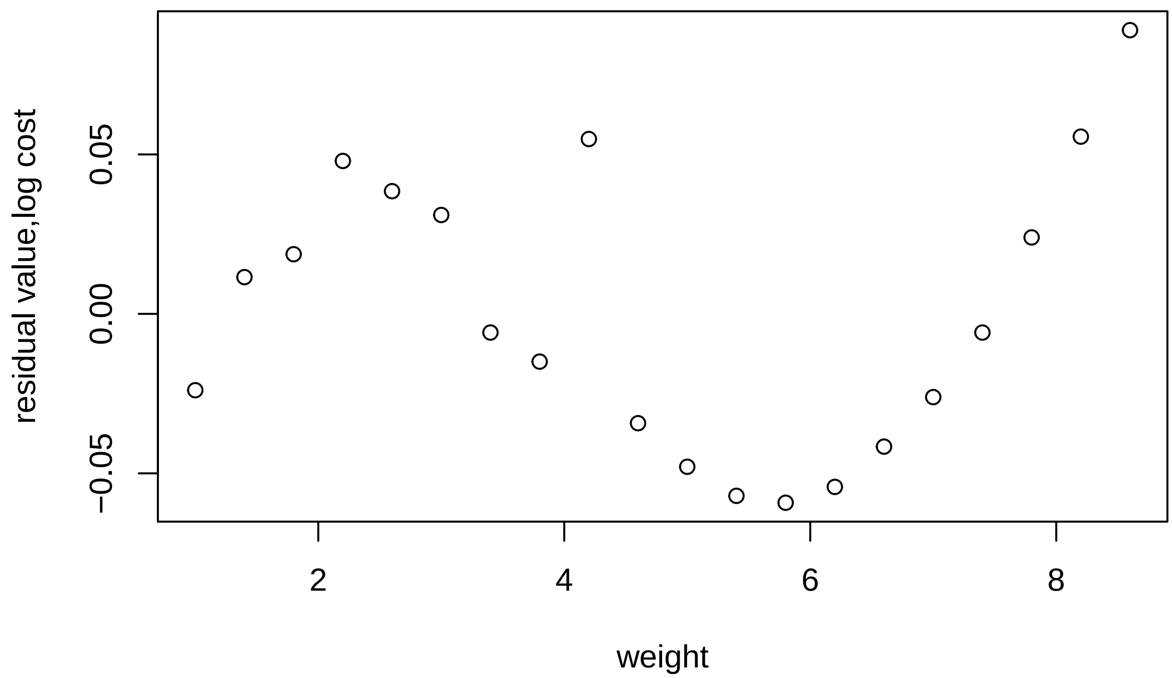
Step 4: Fitting a second linear model (if desired) (4 marks)

Use the guess of 1/y and x then plot the graph of the z against x when z is 1/y,and use the residuals graph of linear model to find whether the points are randomly distributed around 0. Use the boxplot to see if some residuals are outliers

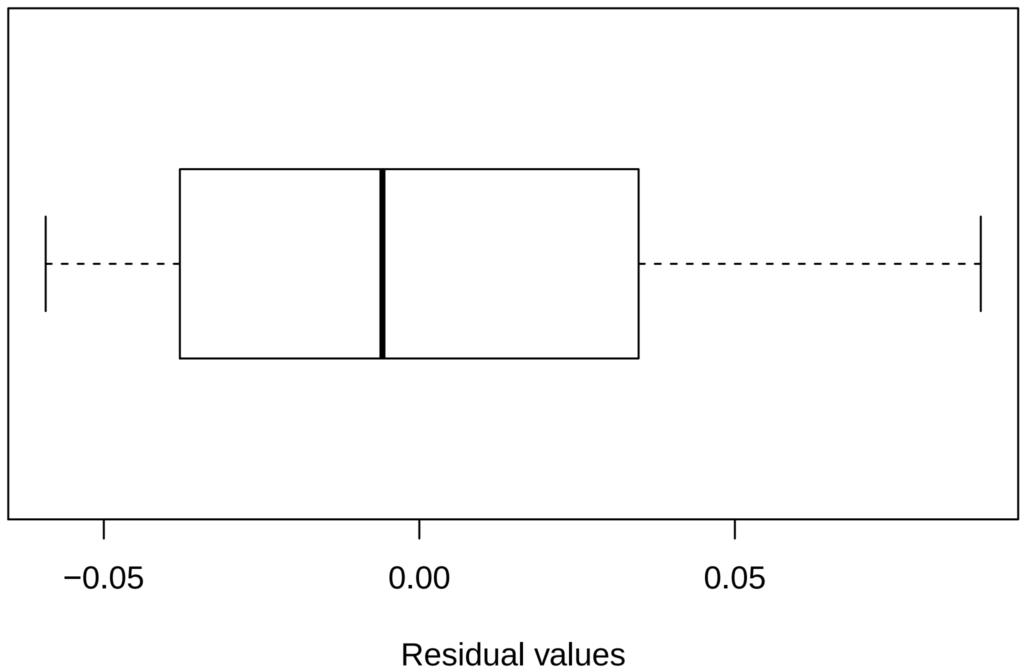
```
x<-df$weight
z<-1/(df$cost)
model2<-lm(z~x)
xlab="weight"
zlab="(1/)cost"
plot(x,z,xlab=xlab,ylab=zlab)
abline(a=model2$coefficients[1],b=model2$coefficients[2],lwd=2,col="blue")
```




```
plot(x,model2$residuals,xlab="weight",ylab="residual value,log cost")
```




```
boxplot(model2$residuals,horizontal=TRUE,xlab="Residual values")
```



There is no outlier,and the most of the points are within the 0.05 and -0.05 But in step 2 that is within 0.3 .And most of the points are near the 0 point ,this is the evidence to the normal distribution. In this graph of boxplot shows that values centred around 0,so in this step is a better fit to the data

Step 5: Interpreting results (4 marks)

When we get this data we first use the setwd and getwd to find the file of the data,then use the read.table to read the data. Plot the graph y againts x to find the shape of the graph then we can guess the linear model.

Then we can try many types of models and use the residuals graph of linear model to find whether the points are randomly distributed around 0. Use the boxplot to see if some residuals are outliers.Then we can define if that linear module is well, then find the best one.

In step4 the residuals appear to be centred around 0 which suggests that this model is a better fit to the data

and $1/Yi = \beta_0+\beta_1 Xi+\epsilon i$ may be a better model than simply $Yi = \beta_0+\beta_1 Xi+\epsilon i$ and $log(Yi) = \beta_0+\beta_1 Xi+\epsilon i$.

But this may not be the best fit of the data,and we can not sure that if this data is exactly in linear model.We can only find that in the step4 that is a good fit of the data.

Calling model<-lm(y~x) and model1<-(z~x),we can obtain the parameters $\widehat{\beta_0}$ and $\widehat{\beta_1}$ from modelcoefficients, and the residuals ϵ from modelresiduals.

An inline equation: $Y = \beta_0 + \beta_1 X + \epsilon$.

Another equation:

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

Some estimated parameters: $\widehat{\beta_0}$ and $\widehat{\beta_1}$.

An expression with some common mathematical functions

$$\cos(A) + B^3 + \exp(C) - \log(D) - 2E$$

It is definitely **not necessary** to use these (LaTeX) mathematical expressions, but they are provided here in case one wishes to use them.