

Final Assignment_Group11

Name: Jiamian Liu VU student number: 2632301

Name: Xiaoyu Yang VU student number: 2640948

Name: Fangzheng Lyu VU student number: 2644757

We first read the “gala.txt” data via the following code and the output is shown in the same figure. We read the gala.txt to ‘galadata’ instead of ‘gala’ to avoid name overlapping.

```
> galadata=read.table("gala.txt",header=TRUE)
> galadata[1:5,]
      Species Area Elevation Nearest Scrutz Adjacent
Baltra      58 25.09      346      0.6      0.6      1.84
Bartolome   31  1.24      109      0.6     26.3     572.33
Caldwell     3  0.21      114      2.8     58.7      0.78
Champion    25  0.10       46      1.9     47.4      0.18
Coamano      2  0.05       77      1.9      1.9     903.82
```

Figure 1: read table and output

From Fig.1 we could know that the header of this table is Species, Area, Elevation, Nearest, Scrutz, Adjacent.

Task 1

In this task, we use the step-down method to solve the problem for the response variable Species.

First code and output are shown in Fig.2.

```
> summary(lm(Species~Area+Elevation+Nearest+Scrutz+Adjacent,data=galadata))
```

Call:

```
lm(formula = Species ~ Area + Elevation + Nearest + Scrutz + Adjacent,
    data = galadata)
```

Residuals:

Min	1Q	Median	3Q	Max
-111.679	-34.898	-7.862	33.460	182.584

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.068221	19.154198	0.369	0.715351
Area	-0.023938	0.022422	-1.068	0.296318
Elevation	0.319465	0.053663	5.953	3.82e-06 ***
Nearest	0.009144	1.054136	0.009	0.993151
Scrutz	-0.240524	0.215402	-1.117	0.275208
Adjacent	-0.074805	0.017700	-4.226	0.000297 ***

signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 60.98 on 24 degrees of freedom

Multiple R-squared: 0.7658, Adjusted R-squared: 0.7171

F-statistic: 15.7 on 5 and 24 DF, p-value: 6.838e-07

Figure 2: First step of step-down method

We could know that the p-value of “Nearest” is highest (0.993151) and is larger than 0.05, so we remove it. The next step code and output are shown in Fig.3.

```
> summary(lm(Species~Area+Elevation+Scruz+Adjacent,data=galadata))
```

Call:
lm(formula = Species ~ Area + Elevation + Scruz + Adjacent, data = galadata)

Residuals:

	Min	1Q	Median	3Q	Max
	-111.637	-34.930	-7.864	33.432	182.524

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.07538	18.74982	0.377	0.709093
Area	-0.02398	0.02151	-1.115	0.275554
Elevation	0.31957	0.05113	6.250	1.54e-06 ***
Scruz	-0.23936	0.16464	-1.454	0.158434
Adjacent	-0.07485	0.01663	-4.501	0.000136 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59.74 on 25 degrees of freedom
Multiple R-squared: 0.7658, Adjusted R-squared: 0.7284
F-statistic: 20.44 on 4 and 25 DF, p-value: 1.39e-07

Figure 3: Second step of step-down method

We could know that the p-value of “Area” is highest (0.275554) and is larger than 0.05, so we remove it. The next step code and output are shown in Fig.4.

```
> summary(lm(Species~Elevation+Scruz+Adjacent,data=galadata))
```

Call:
lm(formula = Species ~ Elevation + Scruz + Adjacent, data = galadata)

Residuals:

	Min	1Q	Median	3Q	Max
	-88.70	-29.68	-11.12	24.25	192.67

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.25183	17.69201	0.806	0.427808
Elevation	0.27444	0.03137	8.749	3.17e-09 ***
Scruz	-0.21784	0.16426	-1.326	0.196320
Adjacent	-0.06744	0.01532	-4.404	0.000162 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60.02 on 26 degrees of freedom
Multiple R-squared: 0.7542, Adjusted R-squared: 0.7258
F-statistic: 26.59 on 3 and 26 DF, p-value: 4.393e-08

Figure 4: Third step of step-down method

We could know that the p-value of “Scruz” is highest (0.196320) and is larger than 0.05, so we remove it. The next step code and output are shown in Fig.5.

```
> summary(lm(Species~Elevation+Adjacent,data=galadata))

Call:
lm(formula = Species ~ Elevation + Adjacent, data = galadata)

Residuals:
    Min       1Q   Median       3Q      Max
-103.41  -34.33  -11.43   22.57  203.65

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.43287    15.02469   0.095 0.924727
Elevation     0.27657     0.03176   8.707 2.53e-09 ***
Adjacent     -0.06889     0.01549  -4.447 0.000134 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60.86 on 27 degrees of freedom
Multiple R-squared:  0.7376,    Adjusted R-squared:  0.7181
F-statistic: 37.94 on 2 and 27 DF,  p-value: 1.434e-08
```

Figure 5: Fourth step of step-down method

We could know that all the p-value is smaller than 0.05, So all explanatory variables in the model are significant.

The resulting of the step-up method is:

Species = 1.43287 + 0.27657*Elevation – 0.06889*Adjacent + error

Task 2

For the transformed response variable sqrt(Species) we do the similar steps.

First code and output are shown in Fig.6.

```
> summary(lm(sqrt(Species)~Area+Elevation+Nearest+Scruz+Adjacent,data=galadata))

Call:
lm(formula = sqrt(Species) ~ Area + Elevation + Nearest + Scruz +
    Adjacent, data = galadata)

Residuals:
    Min       1Q   Median       3Q      Max
-4.5572 -1.4969 -0.3031  1.3527  5.2110

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.3919243  0.8712678   3.893 0.000690 ***
Area          -0.0019718  0.0010199  -1.933 0.065080 .
Elevation     0.0164784  0.0024410   6.751 5.55e-07 ***
Nearest       0.0249326  0.0479495   0.520 0.607844
Scruz        -0.0134826  0.0097980  -1.376 0.181509
Adjacent     -0.0033669  0.0008051  -4.182 0.000333 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.774 on 24 degrees of freedom
Multiple R-squared:  0.7827,    Adjusted R-squared:  0.7374
F-statistic: 17.29 on 5 and 24 DF,  p-value: 2.874e-07
```

Figure 6: First step of step-down method

We could know that the p-value of “Nearest” is highest (0.607844) and is larger than 0.05, so we remove it. The next step code and output are shown in Fig.7.

```
> summary(lm(sqrt(Species)~Area+Elevation+Scruz+Adjacent,data=galadata))

Call:
lm(formula = sqrt(Species) ~ Area + Elevation + Scruz + Adjacent,
    data = galadata)

Residuals:
    Min       1Q   Median       3Q      Max
-4.4407 -1.6642 -0.3444  1.2431  5.0482

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.4114369   0.8576631    3.978 0.000525 ***
Area        -0.0020798   0.0009839   -2.114 0.044681 *
Elevation     0.0167745   0.0023387    7.173 1.62e-07 ***
Scruz        -0.0102950   0.0075309   -1.367 0.183784
Adjacent     -0.0034857   0.0007607   -4.583 0.000110 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.733 on 25 degrees of freedom
Multiple R-squared:  0.7802,    Adjusted R-squared:  0.7451
F-statistic: 22.19 on 4 and 25 DF,  p-value: 6.392e-08
```

Figure 7: Second step of step-down method

We could know that the p-value of “Scruz” is highest (0.183784) and is larger than 0.05, so we remove it. The next step code and output are shown in Fig.8.

```
> summary(lm(sqrt(Species)~Area+Elevation+Adjacent,data=galadata))

Call:
lm(formula = sqrt(Species) ~ Area + Elevation + Adjacent, data = galadata)

Residuals:
    Min       1Q   Median       3Q      Max
-4.9752 -1.7377 -0.4927  1.4342  5.6267

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.8611345   0.7698707    3.716 0.000975 ***
Area        -0.0019221   0.0009933   -1.935 0.063931 .
Elevation     0.0165770   0.0023729    6.986 2.04e-07 ***
Adjacent     -0.0035043   0.0007731   -4.533 0.000115 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.778 on 26 degrees of freedom
Multiple R-squared:  0.7638,    Adjusted R-squared:  0.7366
F-statistic: 28.03 on 3 and 26 DF,  p-value: 2.63e-08
```

Figure 8: Third step of step-down method

We could know that the p-value of “Area” is highest (0.063931) and is larger than 0.05, so we remove it. The next step code and output are shown in Fig.9.

```
> summary(lm(sqrt(Species)~Elevation+Adjacent,data=galadata))
```

Call:
lm(formula = sqrt(Species) ~ Elevation + Adjacent, data = galadata)

Residuals:

Min	1Q	Median	3Q	Max
-5.1626	-2.1405	-0.1246	1.6201	6.3530

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.5378950	0.7198603	4.915	3.83e-05 ***
Elevation	0.0129421	0.0015218	8.504	4.06e-09 ***
Adjacent	-0.0028992	0.0007421	-3.907	0.000566 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.916 on 27 degrees of freedom
Multiple R-squared: 0.7298, Adjusted R-squared: 0.7098
F-statistic: 36.46 on 2 and 27 DF, p-value: 2.126e-08

Figure 9: Third step of step-down method

We could know that all the p-value is smaller than 0.05, So all explanatory variables in the model are significant.

The resulting of the step-up method is:

$\sqrt{\text{Species}} = 3.5378950 + 0.0129421 \cdot \text{Elevation} - 0.0028992 \cdot \text{Adjacent} + \text{error}$

Task 3

The codes making plots are shown below:

```
> gala_lm = lm(Species~Elevation+Adjacent,data=galadata)
> gala_lm2 = lm(sqrt(Species)~Elevation+Adjacent,data=galadata)
> plot(fitted(gala_lm),residuals(gala_lm))
> plot(fitted(gala_lm2),residuals(gala_lm2))
> qqnorm(residuals(gala_lm))
> qqnorm(residuals(gala_lm2))
```

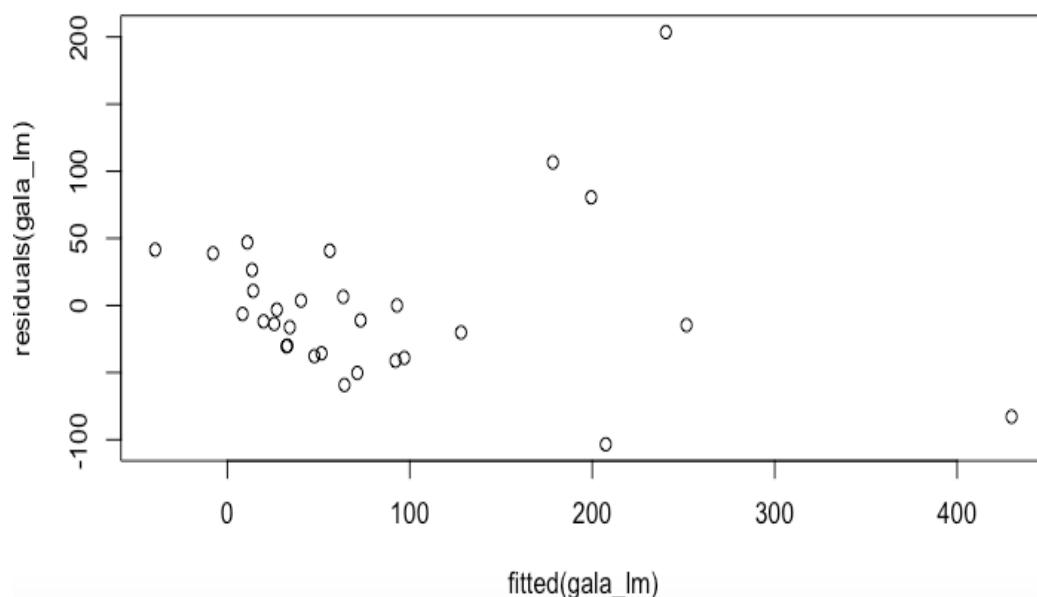


Figure 10: fitted(gala_lm)

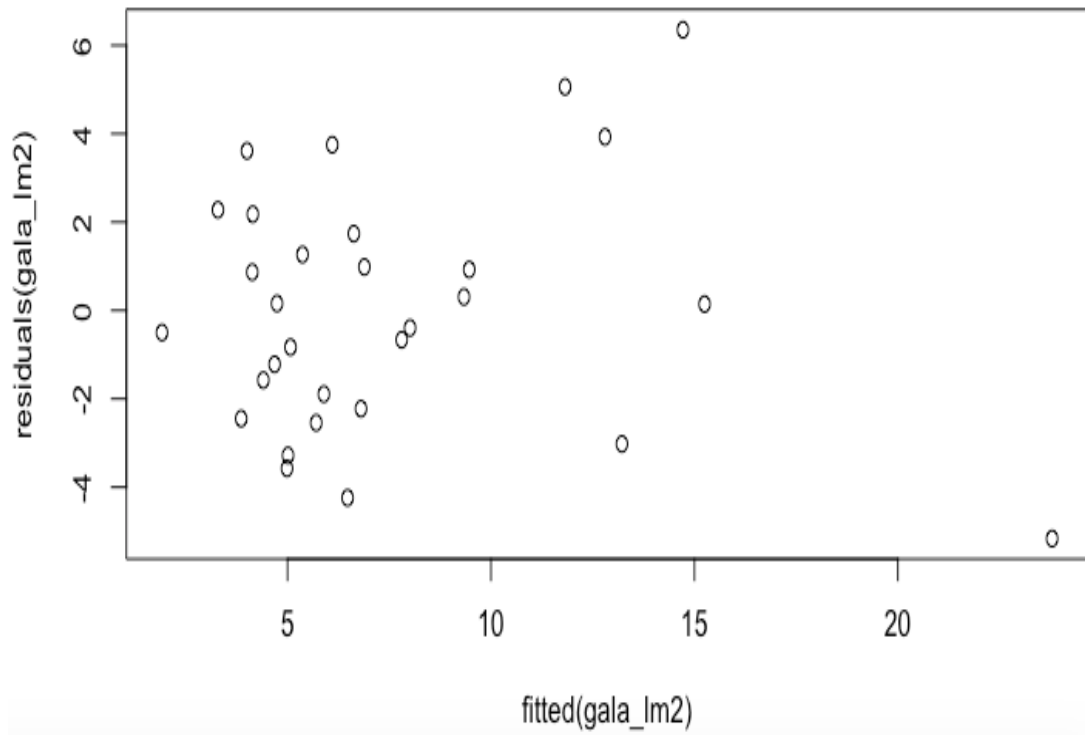


Figure 11: `fitted(gala_lm2)`

Normal Q-Q Plot

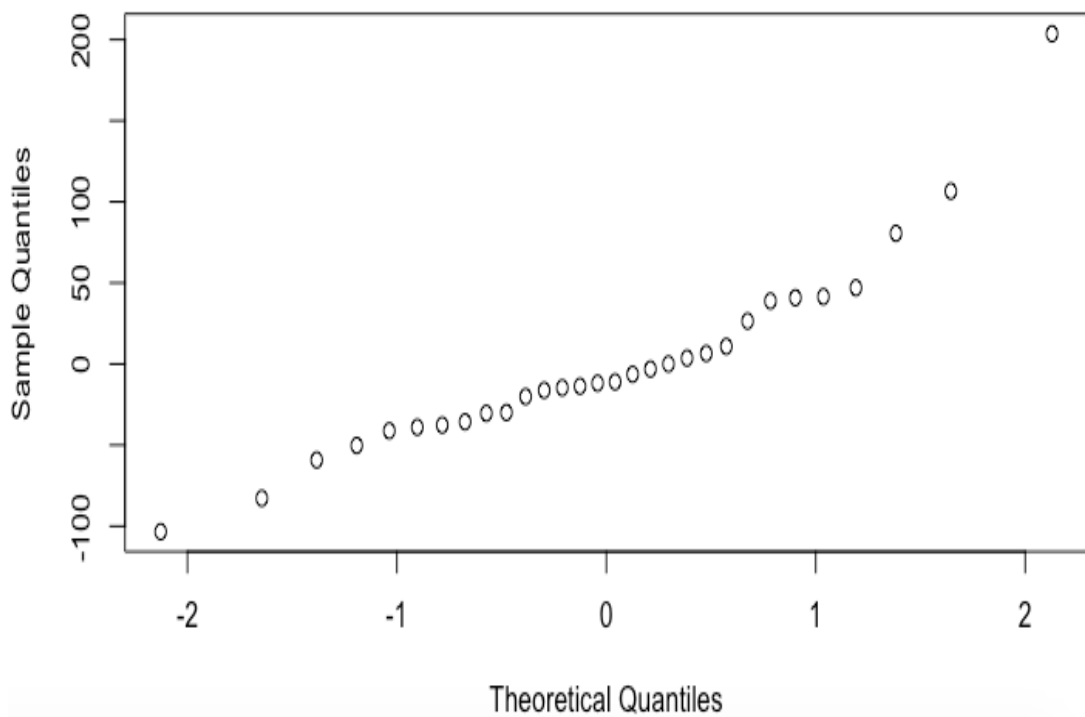


Figure 12: `qqnorm(gala_lm)`

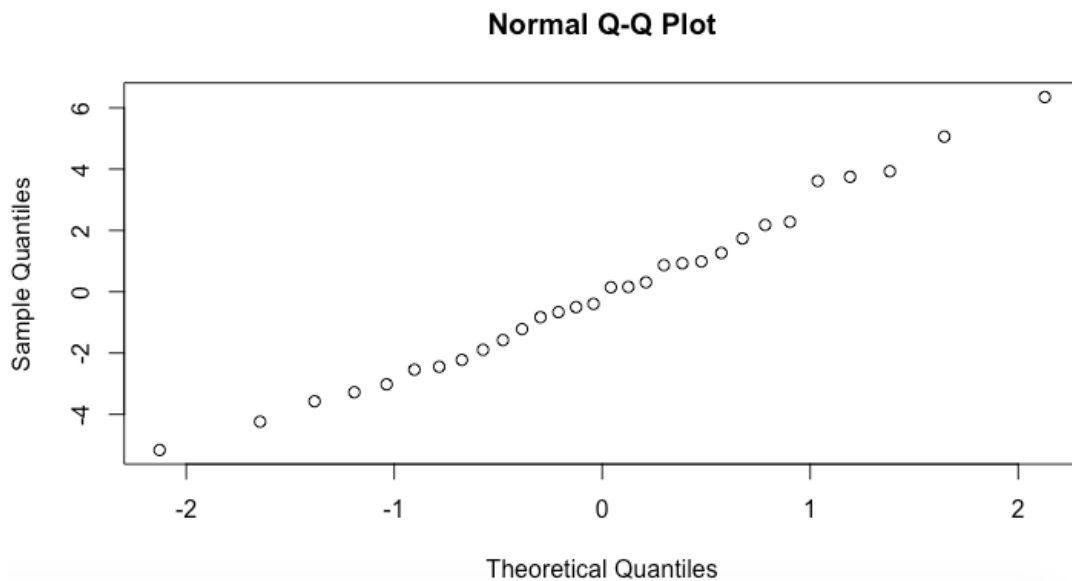


Figure 13: qqnorm(gala_lm2)

```
> shapiro.test(residuals(gala_lm))
```

Shapiro-Wilk normality test

```
data: residuals(gala_lm)
W = 0.89306, p-value = 0.005706
```

```
> shapiro.test(residuals(gala_lm2))
```

Shapiro-Wilk normality test

```
data: residuals(gala_lm2)
W = 0.98489, p-value = 0.9353
```

Figure 14: Shapiro-test

The QQ-plot of residuals and fitted plot are shown as Figure 10 to 13.

From the QQ-plot of “gala_lm” we can see it have some curves and the second QQ-plot for sqrt model is approximating a straight line.

From the Shapiro.test we can see the p-value of residuals(gala_lm) is smaller than 0.05 and p-value of residuals(gala_lm2) is greater than 0.05. So the first sample is not normal while the second sqrt sample can be considered as normal.

So we believe the second resulting model using sqrt is better.

Task 4

We have two explanatory variables in our formula, which is “Elevation” and “Adjacent”. So we perform the Cook’s distance respectively.

The code for the “Elevation” and output is shown in Fig.15 and Fig.16.

```
> sqrtlm=lm(sqrt(Species)~Elevation+Adjacent,data=gala)
> round(cooks.distance(sqrtlm),2)
```

Baltra	Bartolome	Caldwell	Champion	Coamano	Daphne.Major
0.00	0.02	0.02	0.00	0.00	0.00
Daphne.Minor	Darwin	Eden	Enderby	Espanola	Fernandina
0.00	0.01	0.01	0.03	0.02	0.71
Gardner1	Gardner2	Genovesa	Isabela	Marchena	Onslow
0.03	0.03	0.01	1.64	0.00	0.02
Pinta	Pinzon	Las.Plazas	Rabida	SanCristobal	SanSalvador
0.04	0.00	0.00	0.01	0.06	0.00
SantaCruz	SantaFe	SantaMaria	Seymour	Tortuga	Wolf
0.25	0.00	0.08	0.00	0.01	0.01

```
> plot(1:30,cooks.distance(sqrtlm))
```

Figure 15: Code and output for Cook’s distance

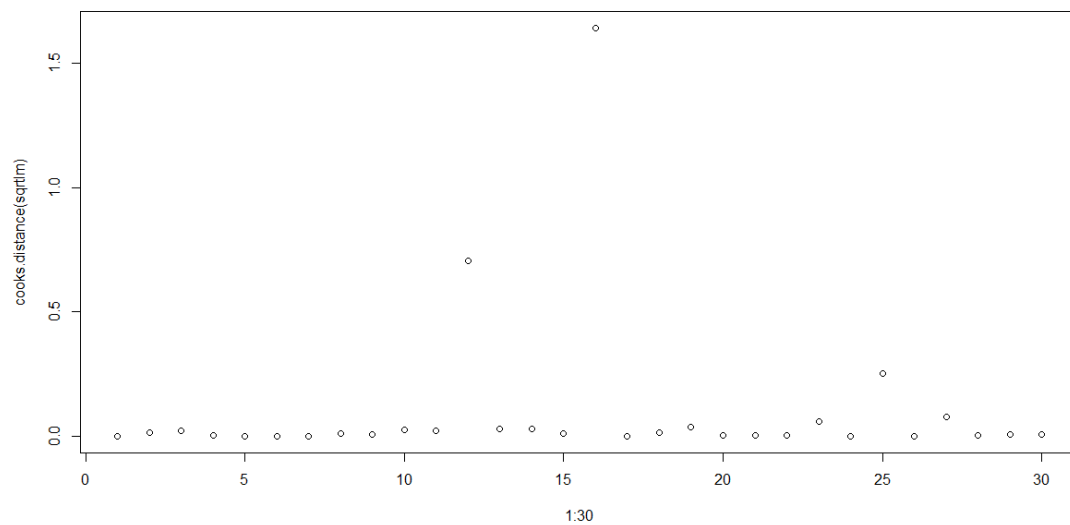


Figure 16: Output for Cook's distance

From the output we could know that the value of “Isabela” is 1.64 larger than 1, so it is the influence point in this model.

We then remove the “Isabela”, from the data we could it is the 16th point. The code and output are shown in Fig.17 and Fig.18.

```
> newgala=galadata[-16,]
> newsqrtlm=lm(sqrt(Species)~Elevation+Adjacent,data=newgala)
> round(cooks.distance(newsqrtlm),2)
```

Baltra	Bartolome	Caldwell	Champion	Coamano	Daphne.Major
0.00	0.04	0.02	0.01	0.00	0.00
Daphne.Minor	Darwin	Eden	Enderby	Espanola	Fernandina
0.00	0.01	0.00	0.03	0.03	7.25
Gardner1	Gardner2	Genovesa	Marchena	onslow	Pinta
0.06	0.04	0.02	0.00	0.01	0.22
Pinzon	Las.Plazas	Rabida	SanCristobal	sanSalvador	SantaCruz
0.00	0.00	0.01	0.05	0.08	0.30
SantaFe	SantaMaria	Seymour	Tortuga	wolf	
0.00	0.09	0.01	0.01	0.01	

```
> plot(1:29,cooks.distance(newsqrtlm))
```

Figure 17: Code and output for Cook's distance 2

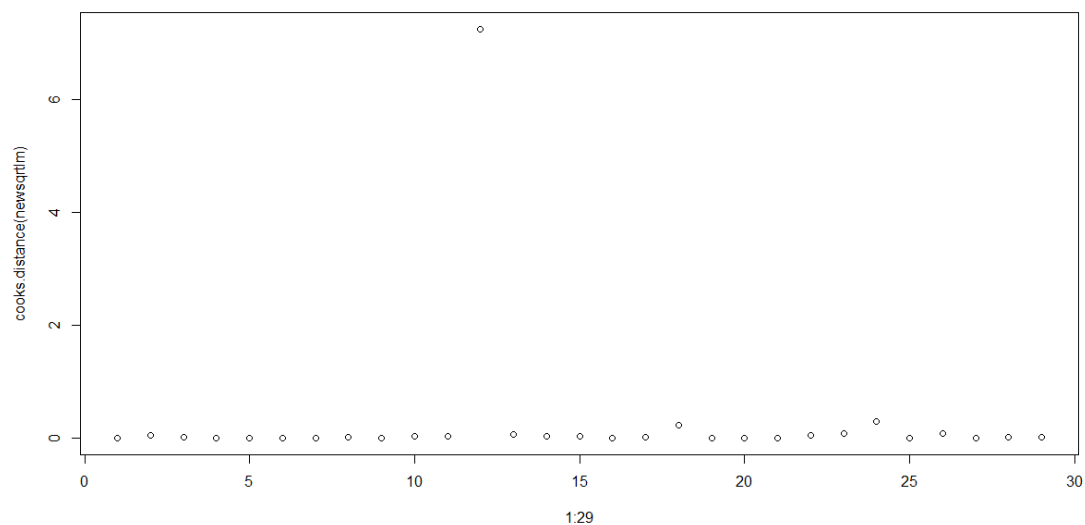


Figure 18: Output for Cook's distance 2

From the output we could know that the value of “Fernandina” is 7.25 larger than 1, so it is still influence point.

We then remove the “Fernandina”, from the data we could it is the 12th point. The code and output are shown in Fig.19 and Fig.20.

```
> newgala2=galadata[-c(12,16),]
> newsqrtlm2=lm(sqrt(Species)~Elevation+Adjacent,data=newgala2)
> round(cooks.distance(newsqrtlm2),2)
```

Baltra	Bartolome	Caldwell	Champion	Coamano	Daphne.Major
0.00	0.07	0.02	0.01	0.22	0.00
Daphne.Minor	Darwin	Eden	Enderby	Espanola	Gardner1
0.00	0.01	0.00	0.03	0.04	0.07
Gardner2	Genovesa	Marchena	Onslow	Pinta	Pinzon
0.04	0.02	0.00	0.01	0.34	0.00
Las.Plazas	Rabida	SanCristobal	SanSalvador	SantaCruz	SantaFe
0.00	0.00	0.04	0.11	0.28	0.00
SantaMaria	Seymour	Tortuga	wolf		
0.09	0.01	0.01	0.01		

```
> plot(1:28,cooks.distance(newsqrtlm2))
```

Figure 19: Code and output for Cook’s distance 3

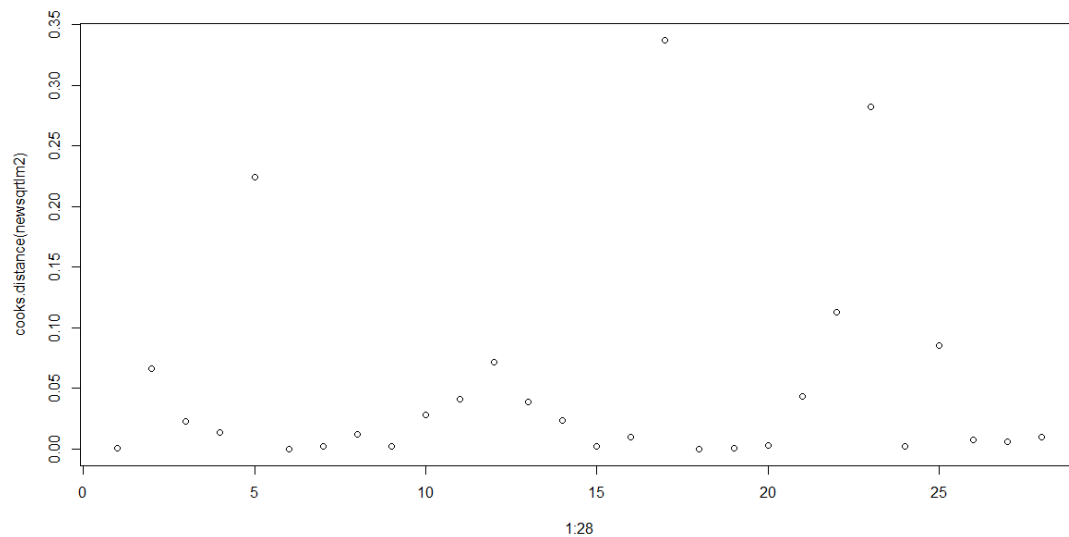


Figure 20: Output for Cook’s distance 3

From the output we could know that all the output is smaller than 1 so there is no influence points now.

For the model, the code and output are shown in Fig21.

```
> summary(newsqrt1m2)
```

Call:
lm(formula = sqrt(Species) ~ Elevation + Adjacent, data = newgala2)

Residuals:

Min	1Q	Median	3Q	Max
-5.2124	-1.7094	-0.3825	1.8110	4.4016

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.455151	0.784143	3.131	0.0044 **
Elevation	0.016866	0.001933	8.724	4.68e-09 ***
Adjacent	-0.001157	0.002336	-0.495	0.6247

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.624 on 25 degrees of freedom
Multiple R-squared: 0.7597, Adjusted R-squared: 0.7405
F-statistic: 39.52 on 2 and 25 DF, p-value: 1.817e-08

Figure 21: New model output

We could know that the new model is:

$$\sqrt{\text{Species}} = 2.455151 + 0.016866 \cdot \text{Elevation} - 0.001157 \cdot \text{Adjacent} + \text{error}$$

Which changed to some extent after removing the influence points.

Task 5

Using the given codes we can make 4 diagrams shown as below:

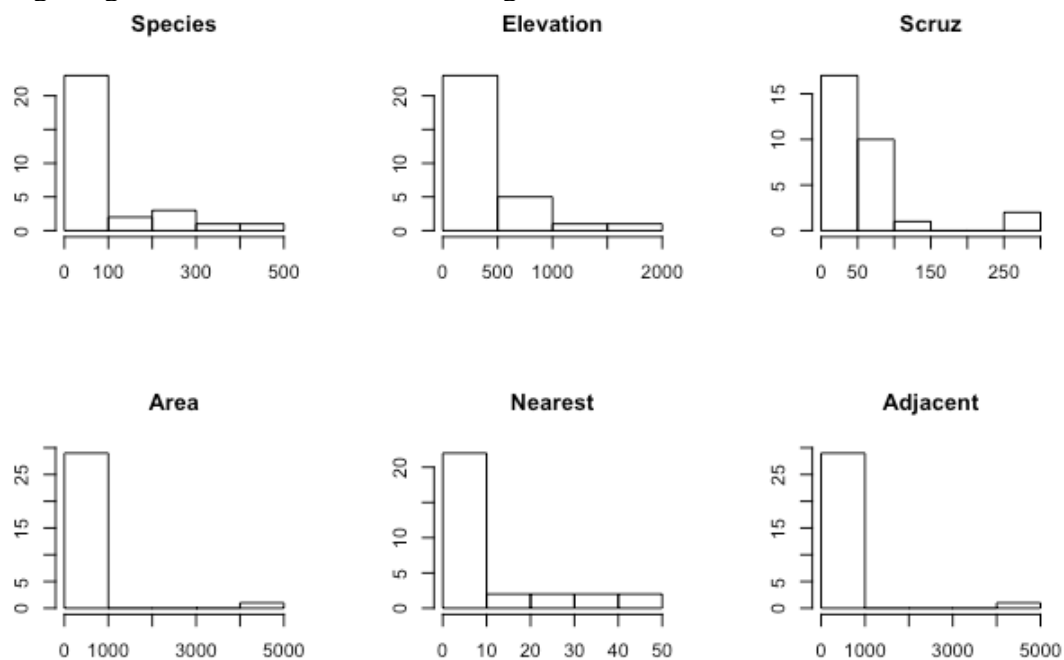


Figure 22 histogram for each variables

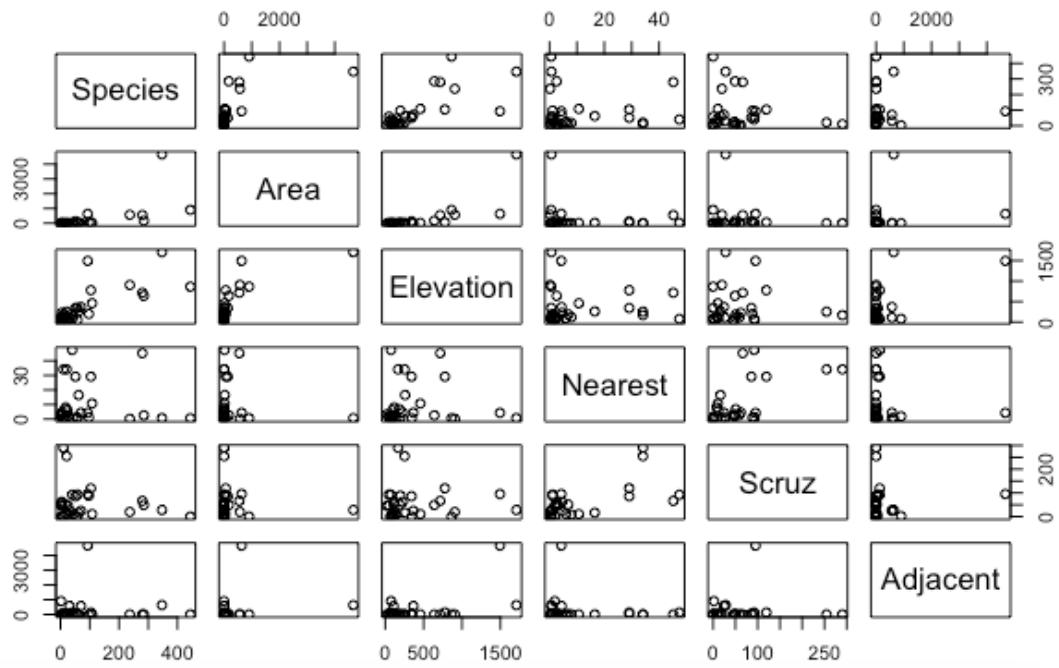


Figure 23: Scatter plot for each variables

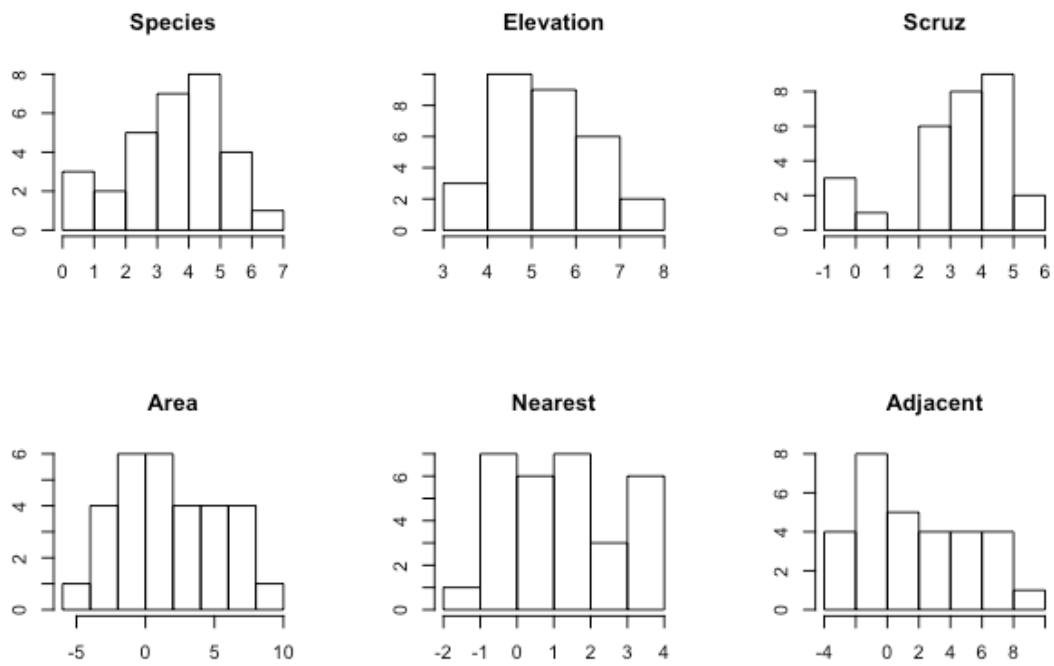


Figure 24 histogram for each variables

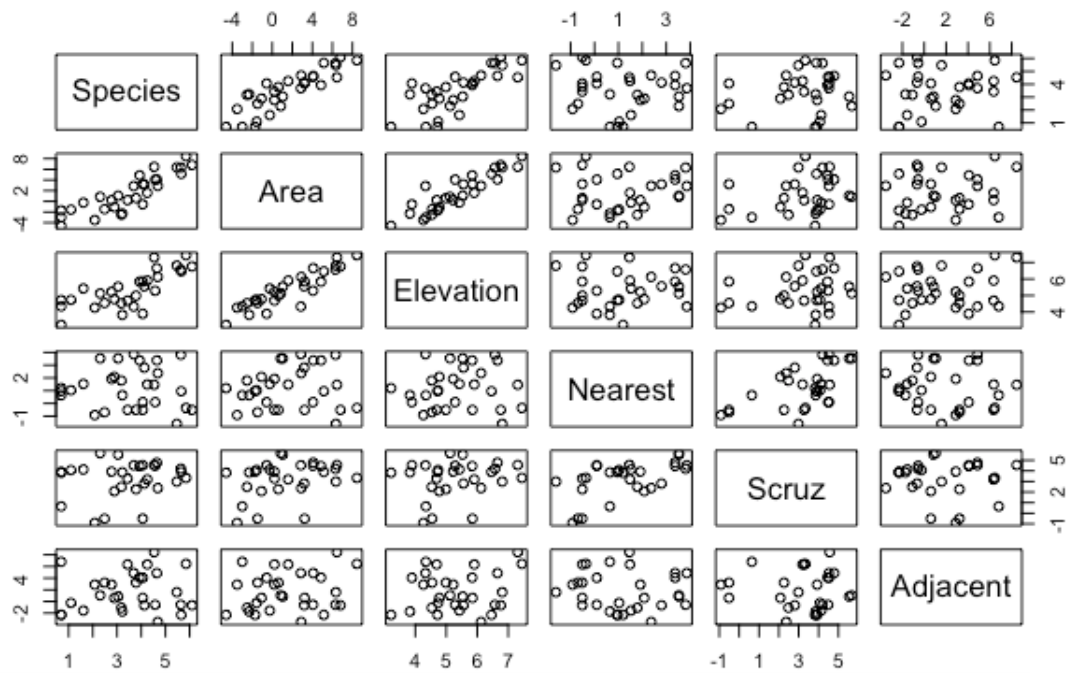


Figure 25: Scatter plot for each variables

For the histograms, we can see the former ones without log show a Poisson distribution. After applying logarithm the distribution shown by the histograms is close to the normal distribution.

And for the pairs, in the first diagram most of the data stack together, which make it difficult to observe and give a conclusion. But in the second diagram, we can find that the data get much scattered. And we can also find that there are some correlation between Species~Area, Species~Elevation, Area~Elevation.

In general we can get a normal distribution, have a clearer view to the data and can find some relationship between the data.

Task 6

According to the instruction, the code and output are shown in Fig.26 and Fig.27.

```
> modlog=lm(log(Species)~log(Area)+log(Elevation)+log(Nearest)+log(Scrutz+1)+log(Adjacent)
),data=galdadata)
> modlog1=step(modlog)
Start: AIC=-9.33
log(Species) ~ log(Area) + log(Elevation) + log(Nearest) + log(Scrutz +
1) + log(Adjacent)

Df Sum of Sq RSS AIC
- log(Nearest) 1 0.1360 14.872 -11.0521
- log(Adjacent) 1 0.1913 14.927 -10.9407
- log(Scrutz + 1) 1 0.6238 15.360 -10.0839
- log(Elevation) 1 0.8992 15.635 -9.5506
<none> 14.736 -9.3276
- log(Area) 1 16.0642 30.800 10.7895

Step: AIC=-11.05
log(Species) ~ log(Area) + log(Elevation) + log(Scrutz + 1) +
log(Adjacent)

Df Sum of Sq RSS AIC
- log(Adjacent) 1 0.1547 15.026 -12.7416
- log(Elevation) 1 0.9713 15.843 -11.1541
<none> 14.872 -11.0521
- log(Scrutz + 1) 1 1.5889 16.461 -10.0069
- log(Area) 1 16.3631 31.235 9.2101

Step: AIC=-12.74
log(Species) ~ log(Area) + log(Elevation) + log(Scrutz + 1)

Df Sum of Sq RSS AIC
- log(Elevation) 1 0.9848 16.011 -12.8373
<none> 15.026 -12.7416
- log(Scrutz + 1) 1 1.5722 16.599 -11.7563
- log(Area) 1 16.2288 31.255 7.2297
```

Figure 26: Code and output 1

```

Step:  AIC=-12.84
log(Species) ~ log(Area) + log(Scruz + 1)

              Df Sum of Sq    RSS    AIC
<none>                 16.011 -12.837
- log(Scruz + 1)    1      1.207 17.218 -12.657
- log(Area)         1     54.875 70.886  29.796
~ |

```

Figure 27: Code and output 2 (continue)

From the AIC-criterion method, we know that there are finally 2 explanatory values, namely $\log(\text{Scruz}+1)$ and $\log(\text{Area})$. We then summary the new modlog1, the code and output are shown in Fig.28.

```

> summary(modlog1)

Call:
lm(formula = log(Species) ~ log(Area) + log(Scruz + 1), data = galadata)

Residuals:
    Min       1Q   Median       3Q      Max
-1.42781 -0.46073  0.03252  0.43723  1.56320

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.32961    0.33602   9.909 1.73e-10 ***
log(Area)       0.39721    0.04129   9.620 3.24e-10 ***
log(Scruz + 1) -0.13403    0.09394  -1.427   0.165
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7701 on 27 degrees of freedom
Multiple R-squared:  0.7741,    Adjusted R-squared:  0.7574
F-statistic: 46.27 on 2 and 27 DF,  p-value: 1.893e-09

```

Figure 28: Summary for the model based on AIC-criterion

From the output we could know that the p-value of $\log(\text{Scruz}+1)$ is 0.165 larger than 0.05 and $\log(\text{Area})$ is much smaller than 0.05. Hence, we know that only the $\log(\text{Area})$ is significant.

Task 7

In task 5 we found the $\log(\text{galadata})$ is from a normal distribution and some collinearities between $\log(\text{Species})$ and $\log(\text{Area})$. But we cannot find any collinearities between the variables without logarithm. Also we know the variable Species is follow a Poisson distribution, so it's reasonable to use logarithm to model a linear function between $\log(\text{Species})$ and $\log(\text{Area})$.

Task 8

From Task 6 we could get the result of modlog1. We first plot Cook's distance. The code and result are shown in Fig.29 and Fig.30.

```

> round(cooks.distance(modlog1), 2)
      Baltra  Bartolome  Caldwell  Champion  Coamano  Daphne.Major
      0.03      0.01      0.06      0.10      0.19      0.00
Daphne.Minor  Darwin    Eden      Enderby  Espanola  Fernandina
      0.08      0.03      0.01      0.10      0.00      0.05
Gardner1      Gardner2  Genovesa  Isabela   Marchena  Onslow
      0.12      0.04      0.00      0.02      0.03      0.01
Pinta         Pinzon    Las.Plazas  Rabida  SanCristobal  SanSalvador
      0.01      0.01      0.01      0.01      0.01      0.00
SantaCruz     SantaFe   SantaMaria  Seymour  Tortuga      Wolf
      0.00      0.00      0.03      0.01      0.00      0.00
> plot(cooks.distance(modlog1))

```

Figure 29: Code and output for Cook's distance of modlog1

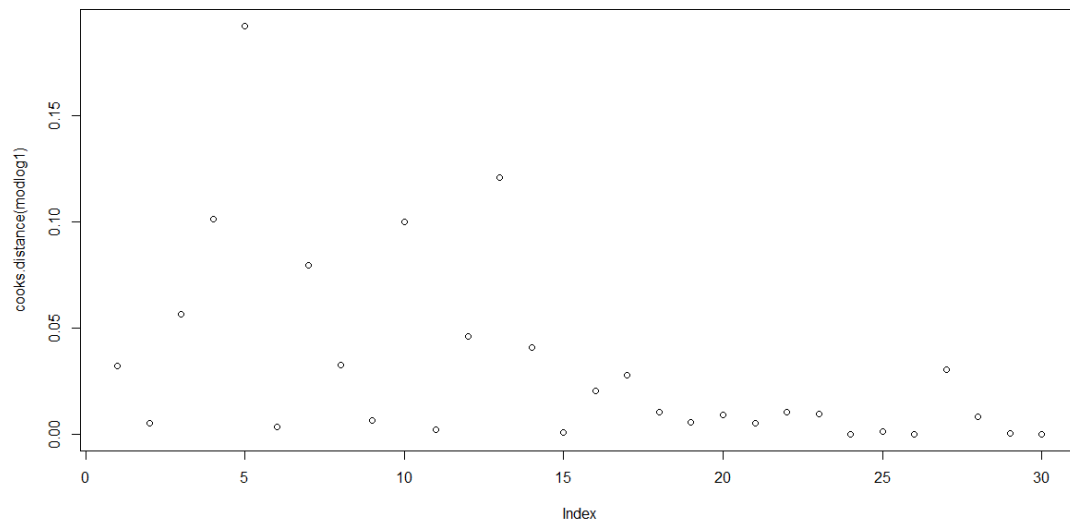


Figure 30: Output for Cook's distance

From the output we could know that all values of cook distance are smaller than 1, there is no influence point in this model.

We then draw the Q-Q plot of residuals and a plot of residuals versus fitted values via the following code as shown in Fig.31.

```
qqnorm(residuals(modlog1))
plot(residuals(modlog1), fitted(modlog1))
```

Figure 31: Code for the rest of task 8

The outputs are shown in Fig.32 and Fig.33 respectively.

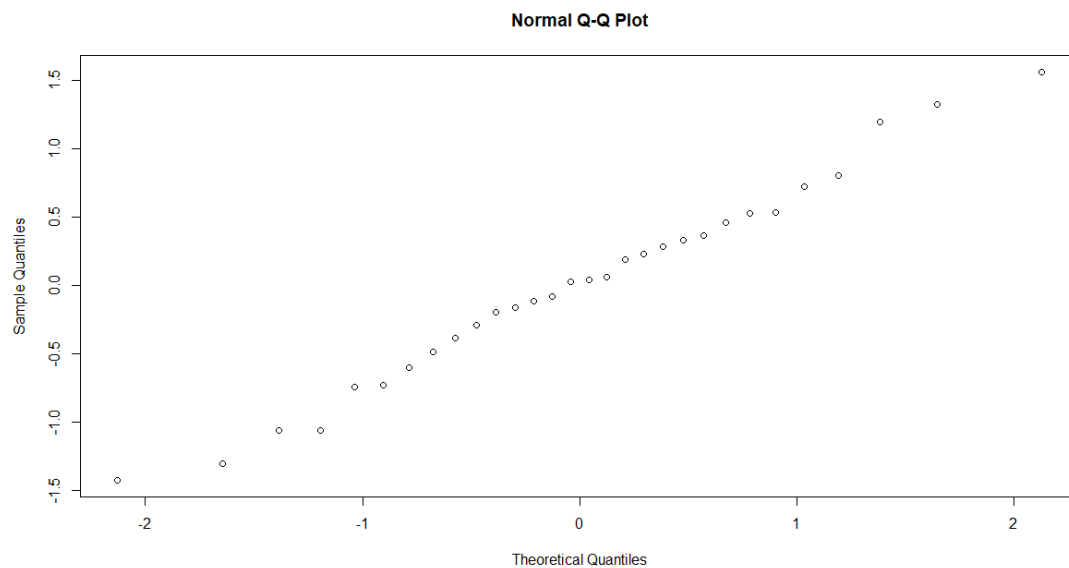


Figure 32: Q-Q plot of residuals

As the Q-Q plot shown, the distribution of residuals is normal distribution.

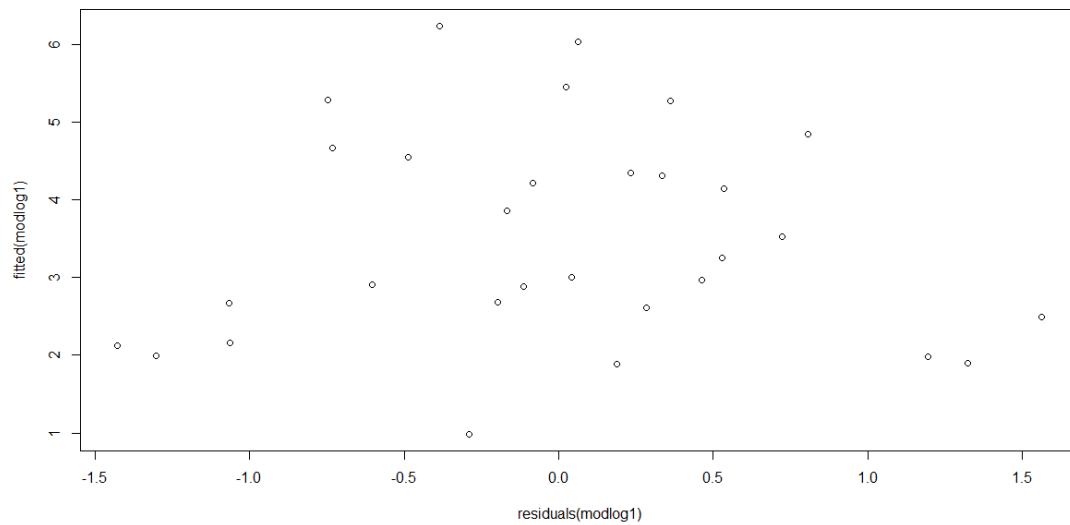


Figure 33: Q-Q plot of fitted values

From the plot of residuals versus fitted values, we can see there is scattered and no obvious structure, so there is no significant relation between the residuals and fitted values.

Task 9

After using the given codes to add columns to the `gala_data`, we use `pairs` to make a plot as figure 34.

```
> pairs(gala_data)
```

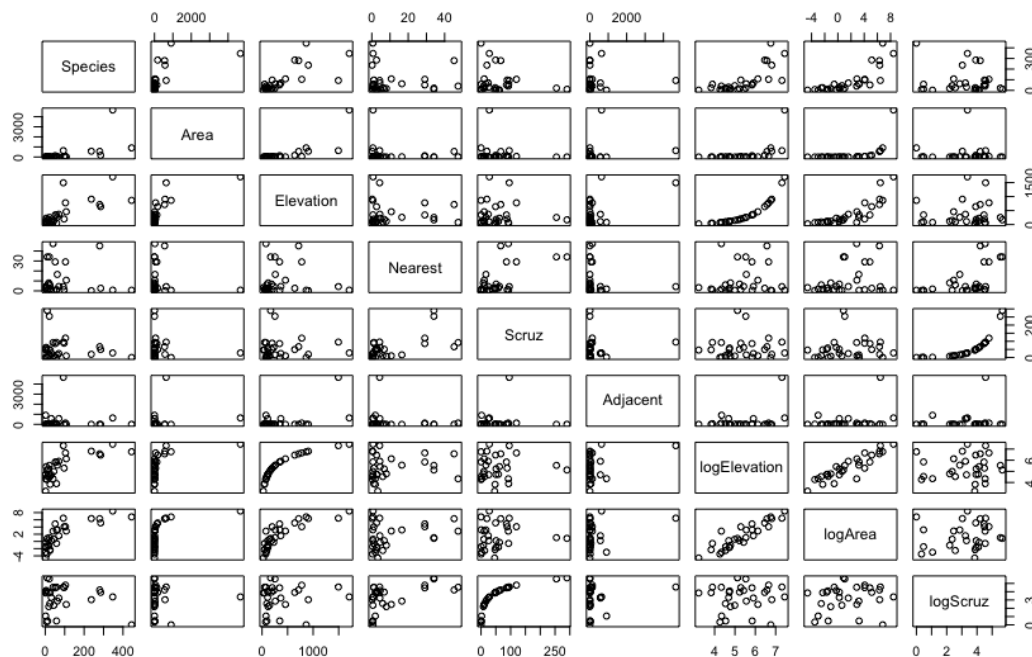


Figure 34 Scatter plot for each variables

From the diagram we can see that there might have possible collinearities between “logElevation” and “logArea”.

To check whether it is useful to include interactions between “logArea” and “logElevation”, we check the linear correlations between these two variables, the result is shown as figure 35

```
> round(cor(galadata[,7:8]),2)
```

	logElevation	logArea
logElevation	1.0	0.9
logArea	0.9	1.0

Figure 35 round for galadata

We can find that both of variables have a 0.9 correlations coefficients to the other in this pair. So it's not useful to include interactions between “logElevation” and “logArea”.

Task 10

In task 3 we have compared two models and the model using $\sqrt{\text{Species}}$ is better than the model using Species.

For modlog1, there is scattered and no obvious structure, which means there's no significant relation between residuals and fitted values. Therefore, we also discard this model.

And for modlog2, it contains two variables “logElevation” and “logArea” which have collinearities. Since it's not useful to fit them in the same model, modlog2 is not the best choice.

To conclude, we prefer the model using $\sqrt{\text{Species}}$. Because it contains less variables and most of the variables in the model are relevant and easy to observe.