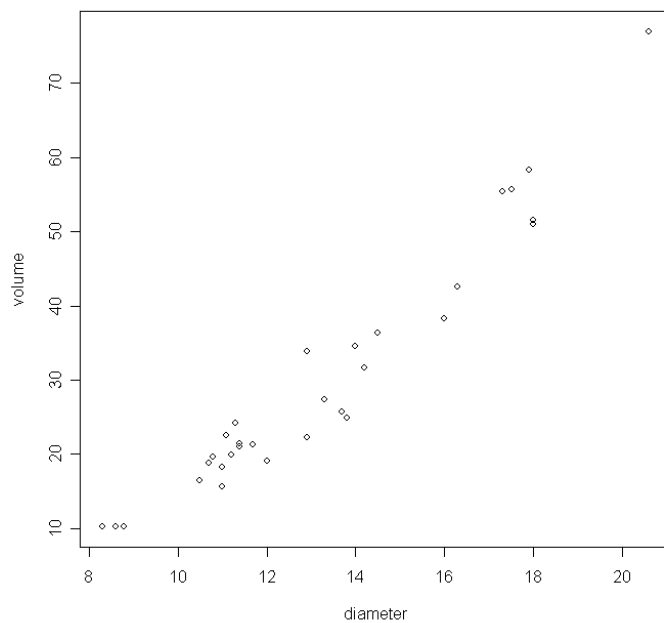
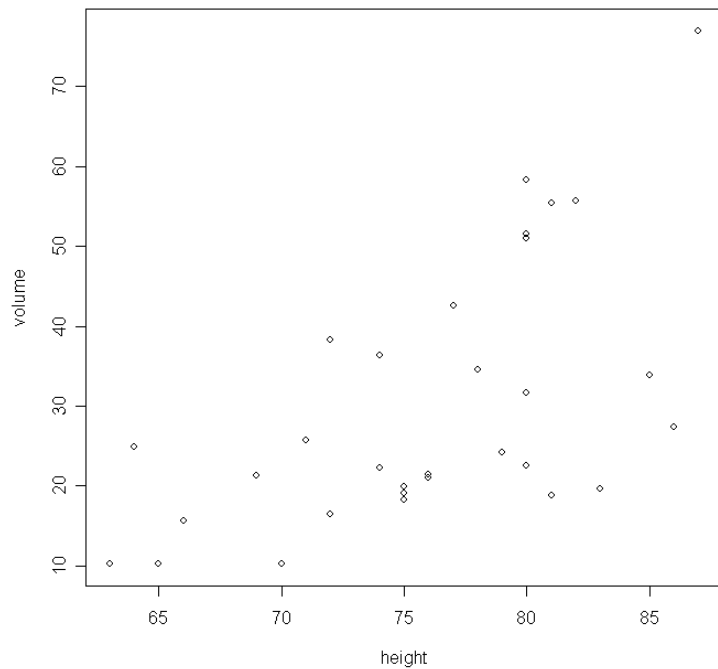


Cherry.R

```
> #####  
> library(MASS)  
> library(car)  
>  
> cherry <- read.table(file="N:\\courses\\stat8620\\Fall 08\\trees.dat",header=T)  
> cherry  
  d h v  
1  8.3 70 10.3  
2 10.7 81 18.8  
<portion omitted>  
31 17.9 80 58.3  
>  
> # Step 1.  
>  
> plot(cherry$d,cherry$v,xlab="diameter",ylab="volume")
```



```
> plot(cherry$h,cherry$v,xlab="height",ylab="volume")
```



```
> lm1 <- lm(v~ h + d,data=cherry)
> summary(lm1)
```

Call:

```
lm(formula = v ~ h + d, data = cherry)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.4065	-2.6493	-0.2876	2.2003	8.4847

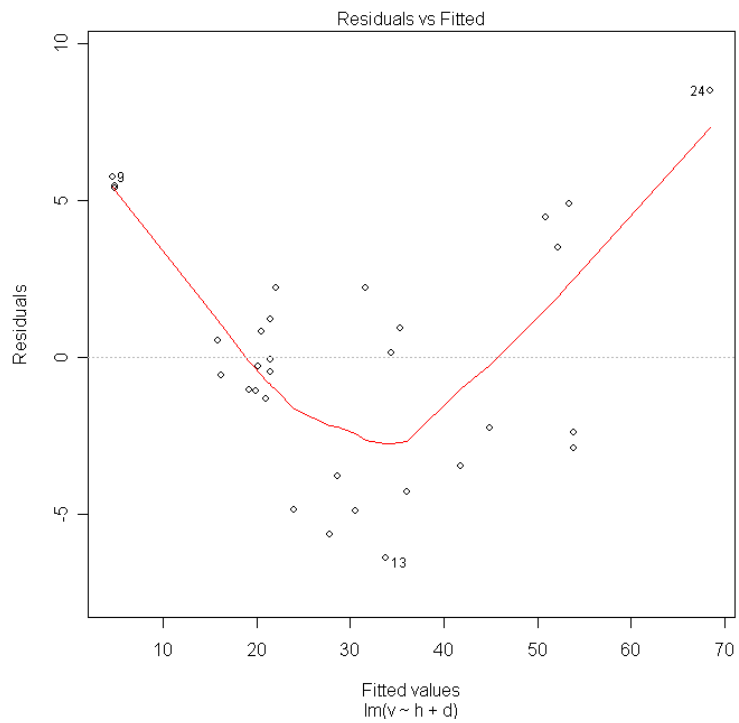
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-57.9877	8.6382	-6.713	2.75e-07 ***
h	0.3393	0.1302	2.607	0.0145 *
d	4.7082	0.2643	17.816	< 2e-16 ***

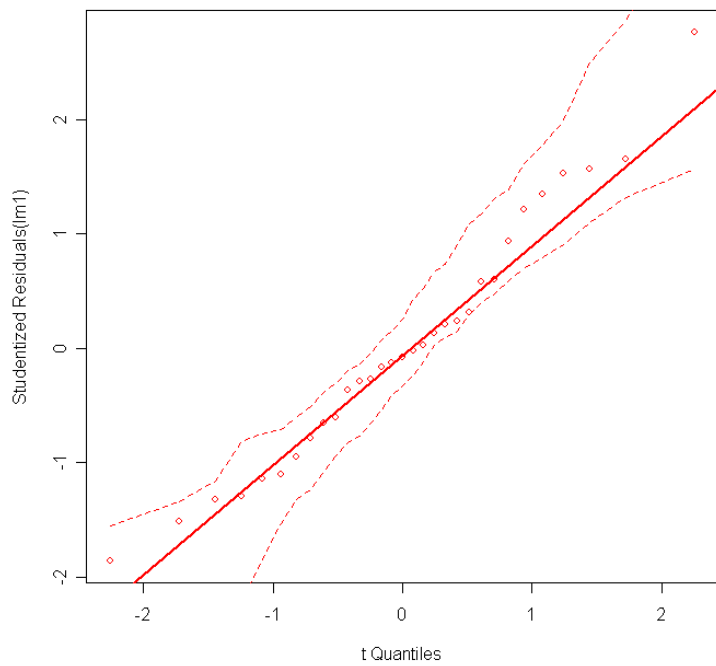
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.882 on 28 degrees of freedom
 Multiple R-squared: 0.948, Adjusted R-squared: 0.9442
 F-statistic: 255 on 2 and 28 DF, p-value: < 2.2e-16

```
> plot(lm1,which=1)
```



```
> qq.plot(lm1,sim=T)
```



```
>
> infl <- influence.measures(lm1)
> summary(infl)
Potentially influential observations of
```

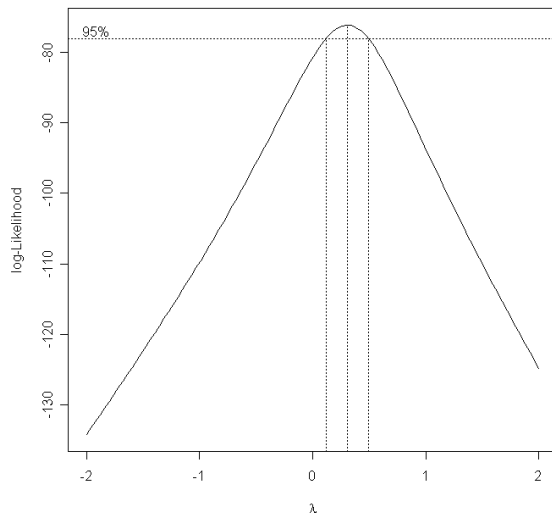
```
lm(formula = v ~ h + d, data = cherry) :
```

```
dfb.1_ dfb.h dfb.d dffit cov.r cook.d hat  
24 -0.74 0.34 0.97 1.50_* 0.68 0.61 0.23  
> shapiro.test(residuals(lm1))
```

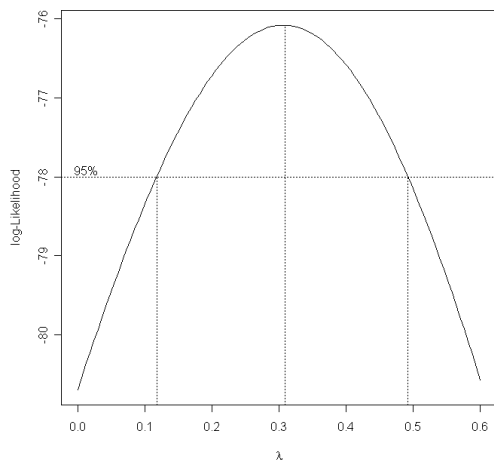
Shapiro-Wilk normality test

```
data: residuals(lm1)  
W = 0.9743, p-value = 0.644
```

```
>  
> # Step 2.  
>  
> bc1 <- boxcox(v ~ h + d, data = cherry,  
+ lambda = seq(-2, 2, length = 100), plotit=T)
```



```
>  
> bc2 <- boxcox(v ~ h + d, data = cherry,  
+ lambda = seq(0, .6, length = 100), plotit=T)
```



```

>
> bc2$x[bc2$y==max(bc2$y)]
[1] 0.3090909
> max(bc2$y)
[1] -76.08043
>
> prof1 <- function(lambda,y,X) {
+ n <- length(y)
+ z <- y^lambda
+ # pl <- -(n/2)*log( sum( (z- X%%solve(t(X)%%X)%%t(X)%%z )^2 ) ) +
+ # n*log(abs(lambda))+(lambda-1)*sum(log(y))-(n/2)*(1+log(2*pi/n))
+ pl <- -(n/2)*log( sum( (z- X%%solve(t(X)%%X)%%t(X)%%z )^2 ) ) +
+ n*log(abs(lambda))+(lambda-1)*sum(log(y))
+ pl
+ }
> X <- model.matrix(~h+d,data=cherry)
> prof1(.30909,cherry$v,X)
[1] -76.08043
>
> lim <- max(bc2$y)-0.5*qchisq(0.95,1)
> cbind(bc2$x,bc2$y,lim)
      lim
[1,] 0.000000000 -80.70206 -78.00116
<portion omitted>
[19,] 0.109090909 -78.16673 -78.00116
[20,] 0.115151515 -78.04850 -78.00116
[21,] 0.121212121 -77.93310 -78.00116
<portion omitted>
[50,] 0.296969697 -76.08541 -78.00116
[51,] 0.303030303 -76.08080 -78.00116
[52,] 0.309090909 -76.08043 -78.00116
[53,] 0.315151515 -76.08432 -78.00116
<portion omitted>

```

```

[82,] 0.490909091 -77.97579 -78.00116
[83,] 0.496969697 -78.09657 -78.00116
<portion omitted>
[100,] 0.600000000 -80.57769 -78.00116
>
> lm2 <- lm(l(v^(1/3))~ h + d,data=cherry)
> summary(lm2)

```

Call:

```
lm(formula = l(v^(1/3)) ~ h + d, data = cherry)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-0.159602 -0.050200 -0.006827  0.069649  0.133981

```

Coefficients:

```

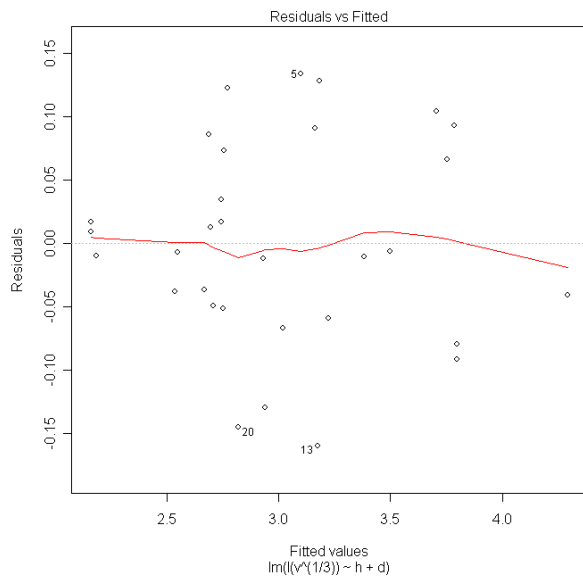
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.085388  0.184315  -0.463  0.647
h            0.014472  0.002777   5.211 1.56e-05 ***
d            0.151516  0.005639  26.871 < 2e-16 ***
---

```

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

Residual standard error: 0.08283 on 28 degrees of freedom
Multiple R-squared: 0.9777, Adjusted R-squared: 0.9761
F-statistic: 612.5 on 2 and 28 DF, p-value: < 2.2e-16

```
> plot(lm2,which=1)
```



```

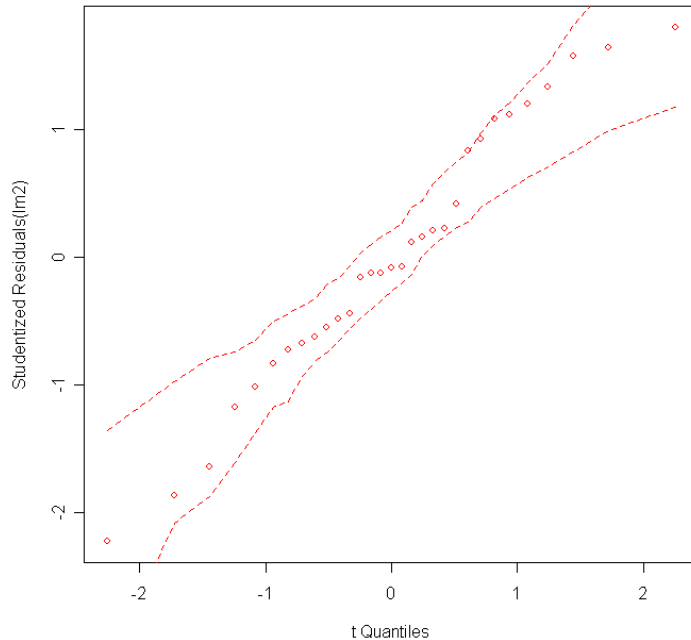
> qq.plot(lm2,sim=T)
integer(0)

```

Warning message:

In object\$coefficients :

\$ operator is invalid for atomic vectors, returning NULL



>

```
> names(summary(lm2))
```

```
[1] "call"      "terms"     "residuals" "coefficients"
```

```
[5] "aliases"   "sigma"     "df"        "r.squared"
```

```
[9] "adj.r.squared" "fstatistic" "cov.unscaled"
```

```
> summary(lm2)$sigma
```

```
[1] 0.08282697
```

>

```
> medvhat <- fitted(lm2)^3
```

```
> meanvhat <- medvhat * (1 + (3 * summary(lm2)$sigma^2) / (fitted(lm2)^2))
```

```
> # Step 3.
```

>

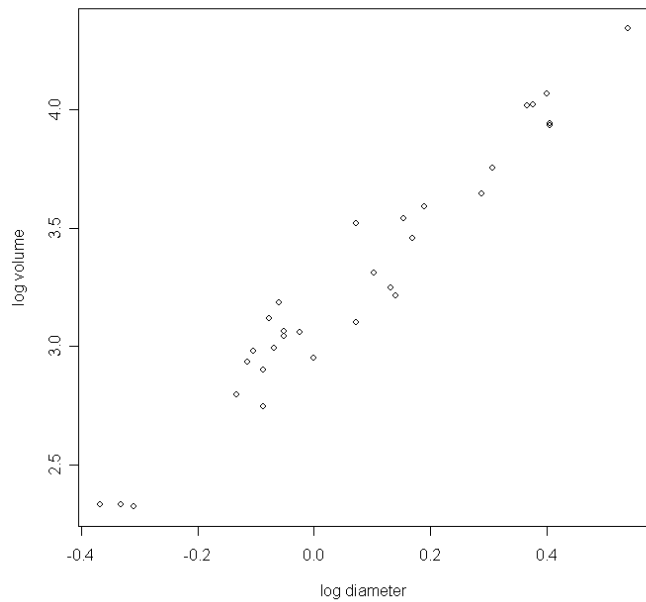
```
> cherry$logv <- log(cherry$v)
```

```
> cherry$logd <- log(cherry$d/12)
```

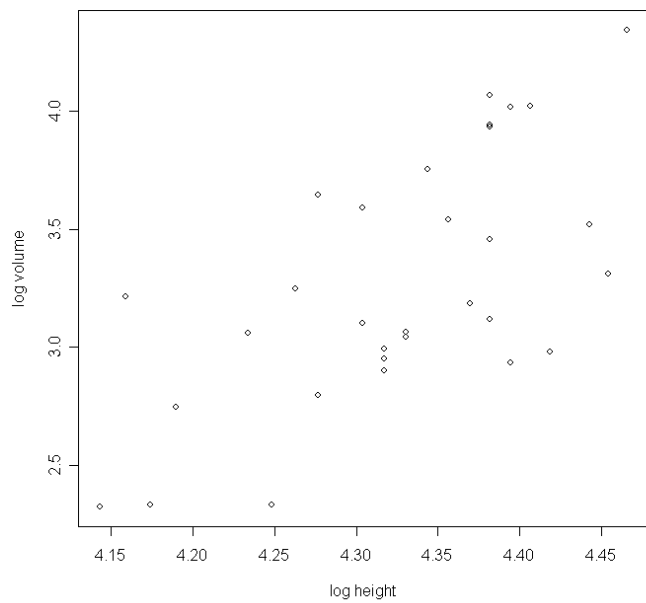
```
> cherry$logh <- log(cherry$h)
```

>

```
> plot(cherry$logd, cherry$logv, xlab="log diameter", ylab="log volume")
```



```
> plot(cherry$logh,cherry$logv,xlab="log height",ylab="log volume")
```



```
> lm3 <- lm(logv~ logh + logd,data=cherry)
> summary(lm3)
```

Call:

```
lm(formula = logv ~ logh + logd, data = cherry)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.168561	-0.048488	0.002431	0.063637	0.129223

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.70492   0.88190  -1.933  0.0634 .
logh         1.11712   0.20444   5.464 7.8e-06 ***
logd         1.98265   0.07501  26.432 < 2e-16 ***
```

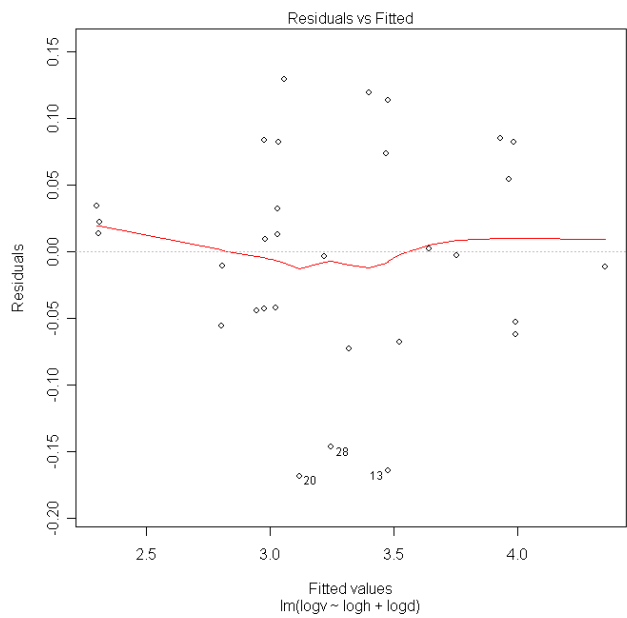
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08139 on 28 degrees of freedom

Multiple R-squared: 0.9777, Adjusted R-squared: 0.9761

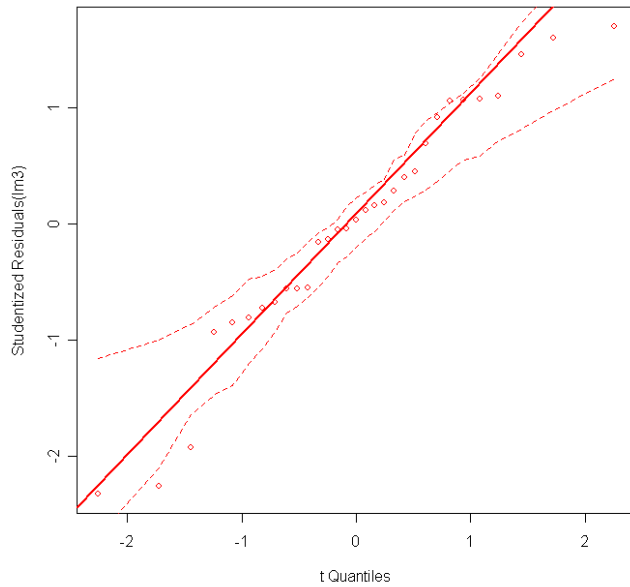
F-statistic: 613.2 on 2 and 28 DF, p-value: < 2.2e-16

```
> plot(lm3,which=1)
```



```
> qq.plot(lm3,sim=T)
```

```
integer(0)
```

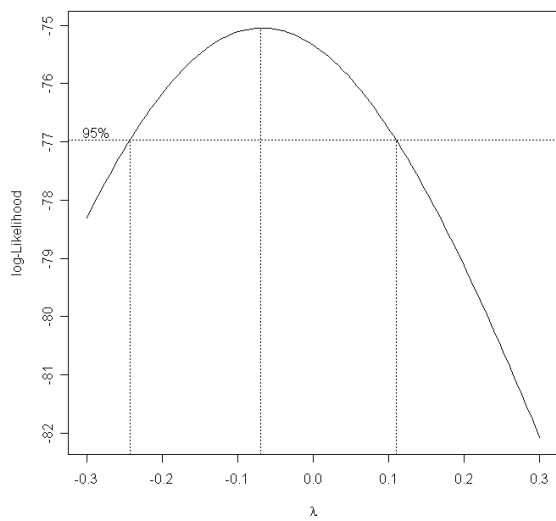


```
>
> shapiro.test(residuals(lm3))
```

Shapiro-Wilk normality test

```
data: residuals(lm3)
W = 0.9592, p-value = 0.2782
```

```
>
> bc3 <- boxcox(v ~ logh + logd, data = cherry,
+ lambda = seq(-.3, .3, length = 100), plotit=T)
```



```
>
> bc3$x[bc3$y==max(bc3$y)]
```

```

[1] -0.06969697
> max(bc3$y)
[1] -75.04532
> lim <- max(bc3$y)-0.5*qchisq(0.95,1)
> cbind(bc3$x,bc3$y,lim)
      lim
[1,] -0.300000000 -78.31252 -76.96605
<portion omitted>
[10,] -0.245454545 -77.02905 -76.96605
[11,] -0.239393939 -76.90270 -76.96605
<portion omitted>
[38,] -0.075757576 -75.04963 -76.96605
[39,] -0.069696970 -75.04532 -76.96605
[40,] -0.063636364 -75.04584 -76.96605
<portion omitted>
[68,] 0.106060606 -76.89529 -76.96605
[69,] 0.112121212 -77.01898 -76.96605
<portion omitted>
[100,] 0.300000000 -82.07841 -76.96605
>
> cherry$off <- cherry$logh+2*cherry$logd
>
> glm4 <- glm(logv~ 1,data=cherry,
+ offset=off,family=gaussian(link="identity"))
> summary(glm4)

```

Call:

```

glm(formula = logv ~ 1, family = gaussian(link = "identity"),
    data = cherry, offset = off)

```

Deviance Residuals:

```

      Min      1Q  Median      3Q      Max
-0.168446 -0.047355 -0.003518  0.066308  0.136467

```

Coefficients:

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.19935    0.01421  -84.42 <2e-16 ***
---

```

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

(Dispersion parameter for gaussian family taken to be 0.006256193)

```

Null deviance: 0.18769 on 30 degrees of freedom
Residual deviance: 0.18769 on 30 degrees of freedom
AIC: -66.342

```

Number of Fisher Scoring iterations: 2

```
>
>
> # Step 4.
>
> glm5 <- glm(v~ h+d,data=cherry,
+ family=gaussian(link=power(1/3)))
> summary(glm5)
```

Call:
glm(formula = v ~ h + d, family = gaussian(link = power(1/3)),
data = cherry)

Deviance Residuals:
Min 1Q Median 3Q Max
-4.6590 -1.5744 -0.2949 1.6653 4.4737

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.051322 0.224095 -0.229 0.820518
h 0.014287 0.003342 4.274 0.000201 ***
d 0.150331 0.005838 25.749 < 2e-16 ***

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

(Dispersion parameter for gaussian family taken to be 6.577063)

Null deviance: 8106.08 on 30 degrees of freedom
Residual deviance: 184.16 on 28 degrees of freedom
AIC: 151.21

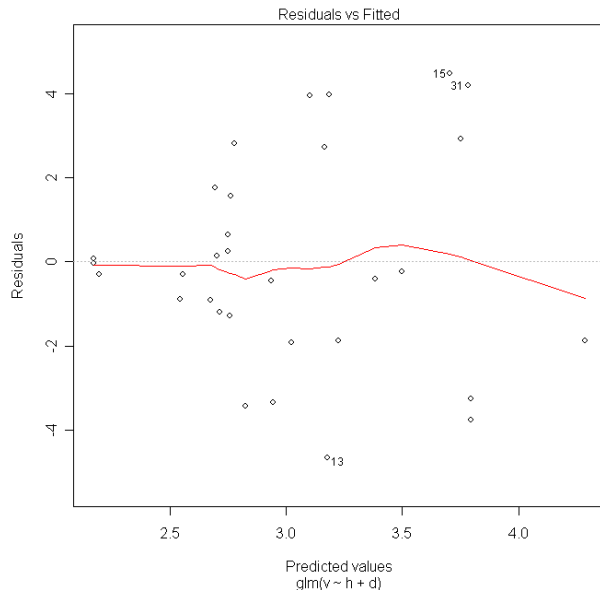
Number of Fisher Scoring iterations: 4

```
> shapiro.test(residuals(glm5))
```

Shapiro-Wilk normality test

data: residuals(glm5)
W = 0.9592, p-value = 0.2778

```
> library(qpcR)
> Rsq(glm5)
[1] 0.9771597
> plot(glm5,which=1)
```



```
>
>
>
> glm6 <- glm(v ~ logh+logd,data=cherry,
+ family=gaussian(link="log"))
> summary(glm6)
```

Call:
 glm(formula = v ~ logh + logd, family = gaussian(link = "log"),
 data = cherry)

Deviance Residuals:
 Min 1Q Median 3Q Max
 -4.9080 -1.1817 -0.2101 1.7014 4.2551

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
 (Intercept) -1.57484 1.04613 -1.505 0.143422
 logh 1.08765 0.24216 4.491 0.000111 ***
 logd 1.99692 0.08208 24.330 < 2e-16 ***

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

(Dispersion parameter for gaussian family taken to be 6.41642)

Null deviance: 8106.08 on 30 degrees of freedom
 Residual deviance: 179.66 on 28 degrees of freedom
 AIC: 150.44

Number of Fisher Scoring iterations: 4

```
> shapiro.test(residuals(glm6))
```

Shapiro-Wilk normality test

```
data: residuals(glm6)
```

```
W = 0.9653, p-value = 0.3996
```

```
> Rsq(glm6)
```

```
[1] 0.9777677
```

```
> plot(glm6,which=1)
```

