Comparison of Human Chest Diameter and Height

Code ▾

Hide

```
library(dplyr)
```

```
Registered S3 method overwritten by 'dplyr':
  method           from
  print.rowwise_df

Attaching package: 恸牾dplyr恸牸

The following objects are masked from 恸牾package:stats恸牸:

    filter, lag

The following objects are masked from 恸牾package:base恸牸:

    intersect, setdiff, setequal, union
```
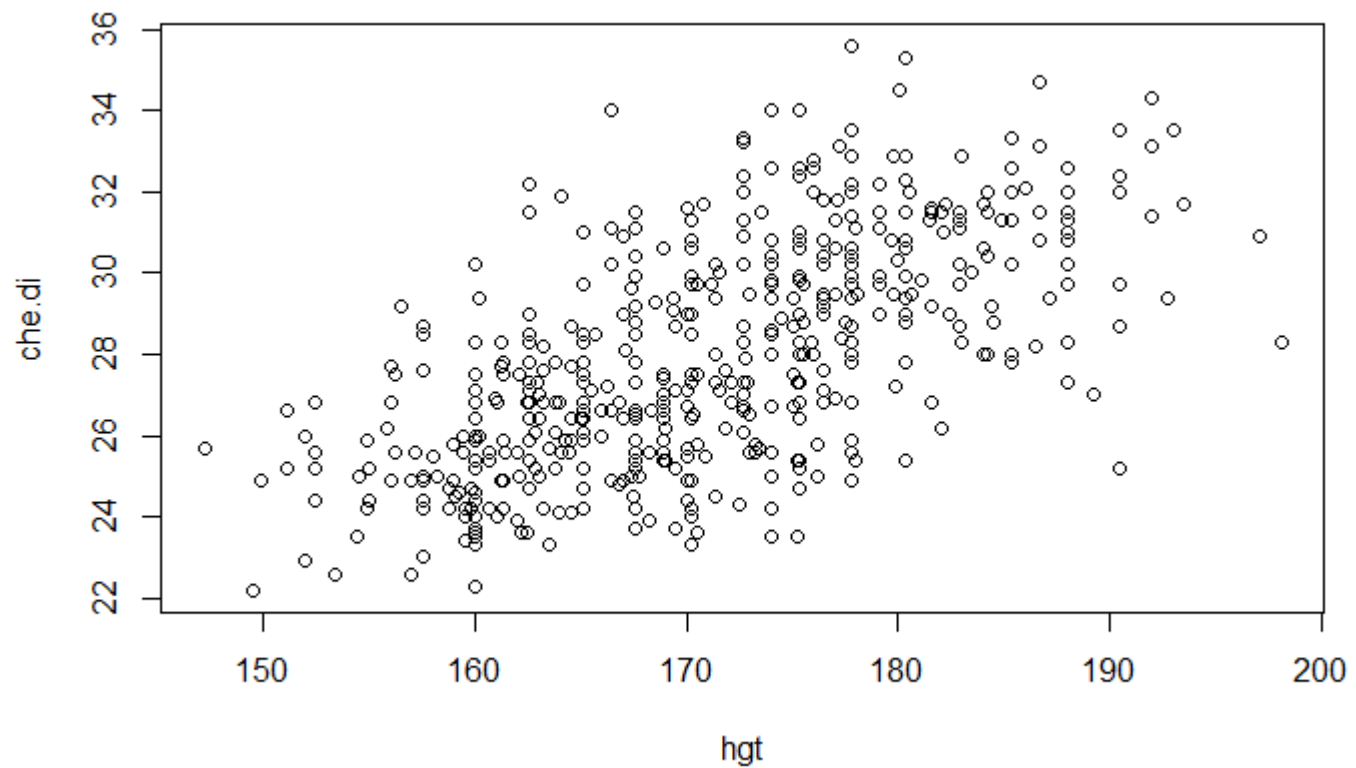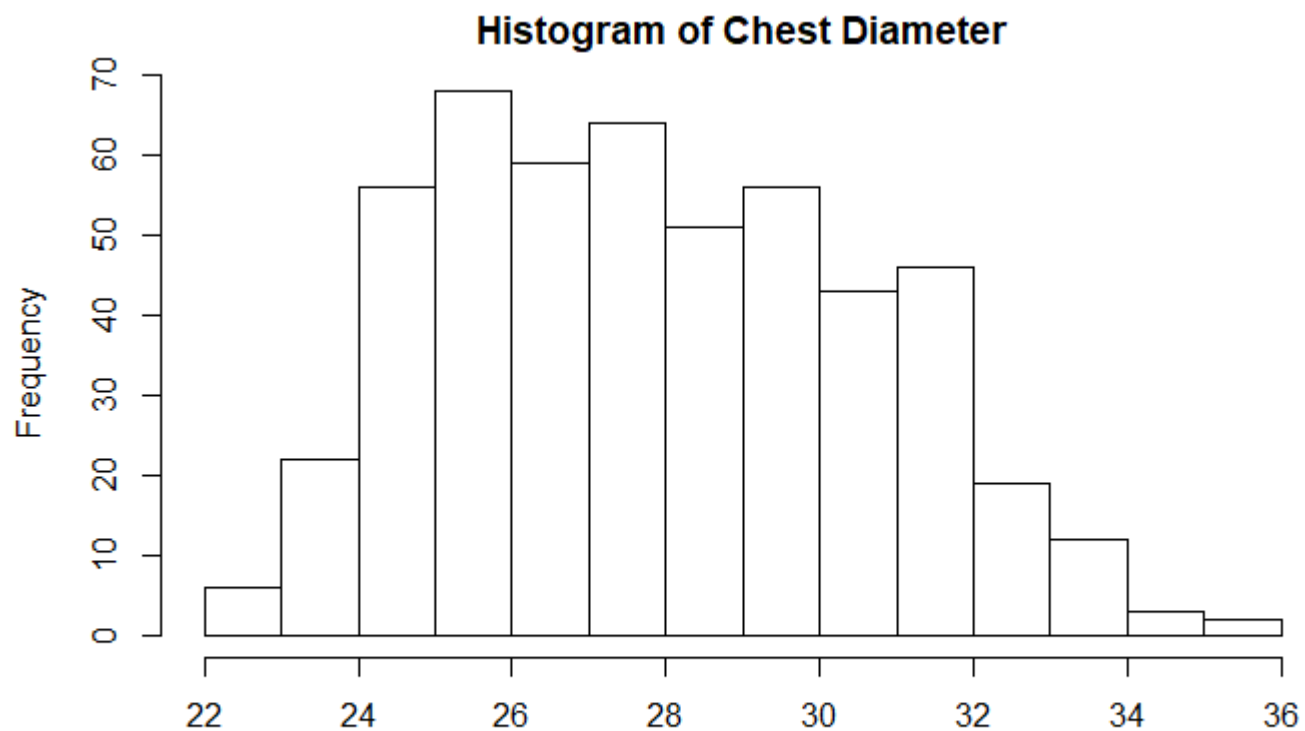
Hide

```
bdims_filtered <- select(bdims_csv,che.di,hgt)

plot(che.di ~ hgt, data = bdims_filtered)
```

```
bdims_filtered$che.di %>% hist(main="Histogram of Chest Diameter")
```
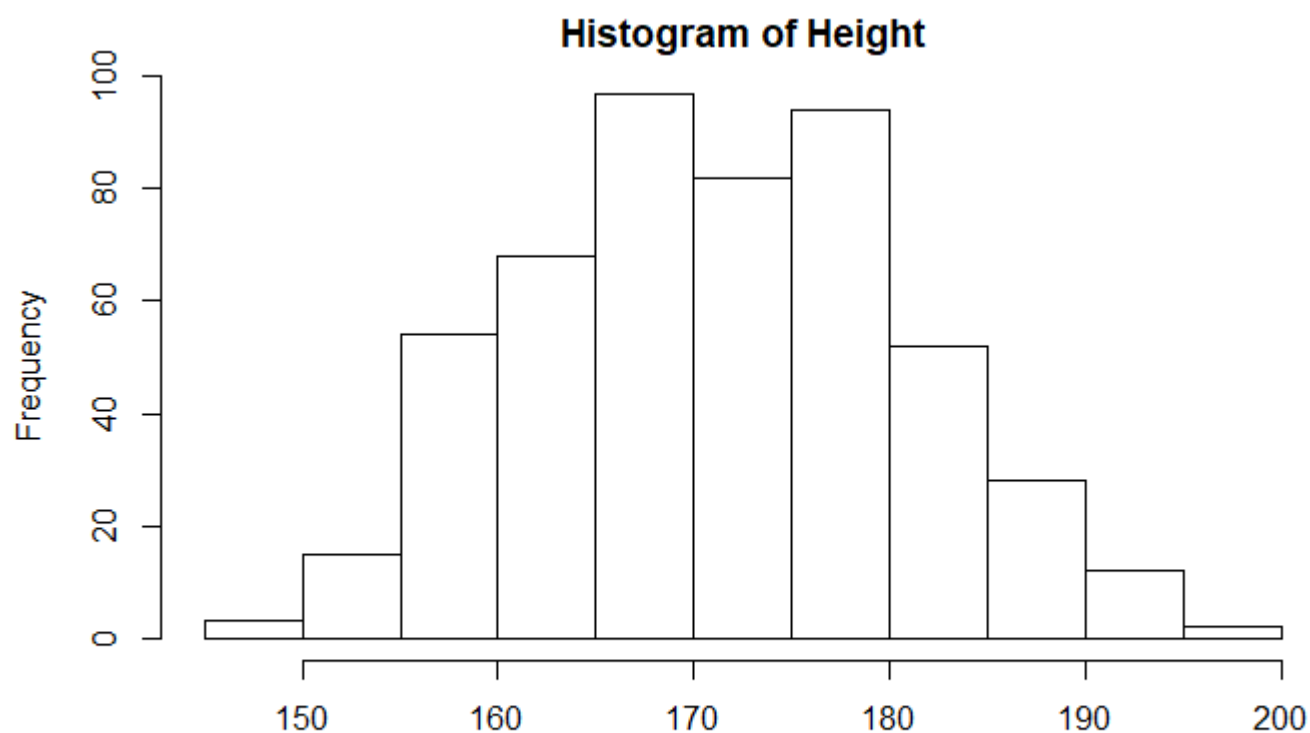
Hide

```
bdims_filtered$hgt %>% hist(main="Histogram of Height")
```


Histogram of Height

Hide

```
plot(log(che.di) ~ log(hgt), data = bdims_filtered)
```
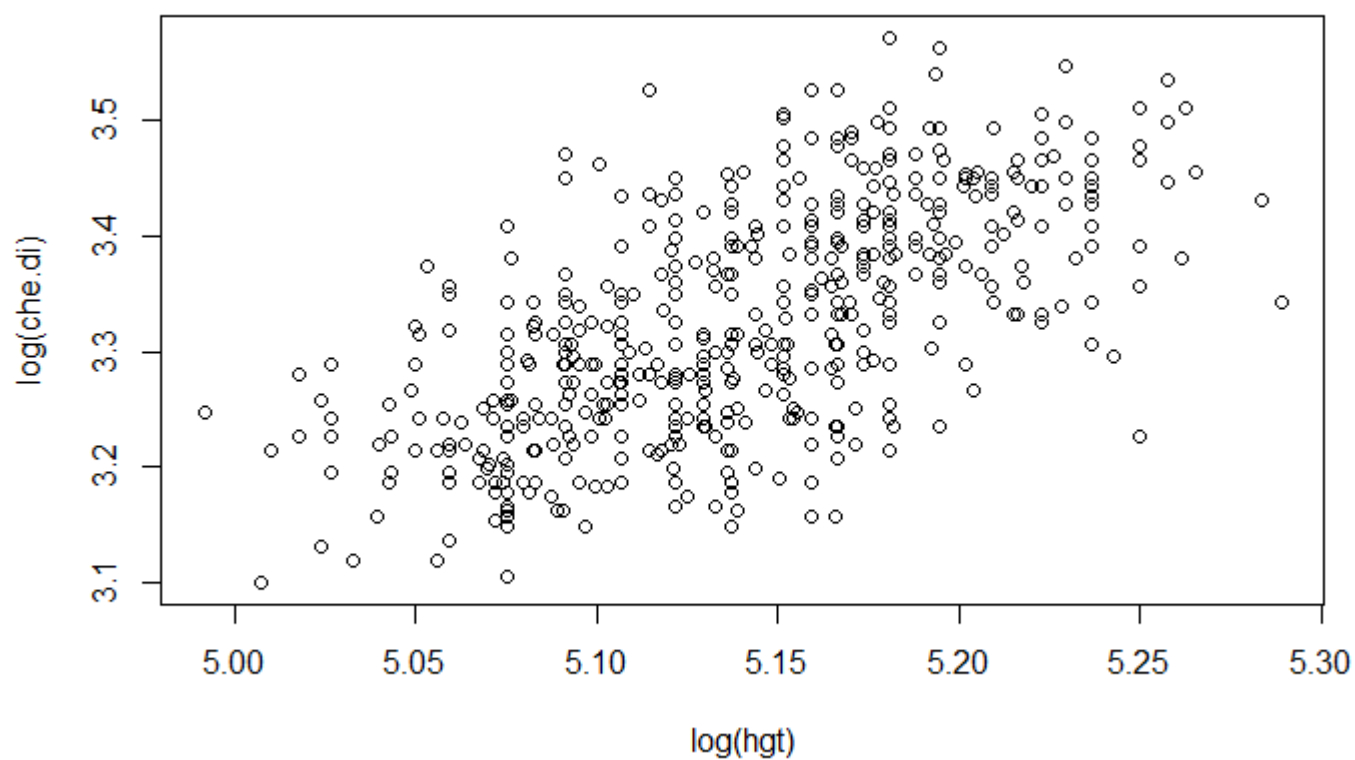
Hide

```
plot(sqrt(che.di) ~ sqrt(hgt), data = bdims_filtered)
```
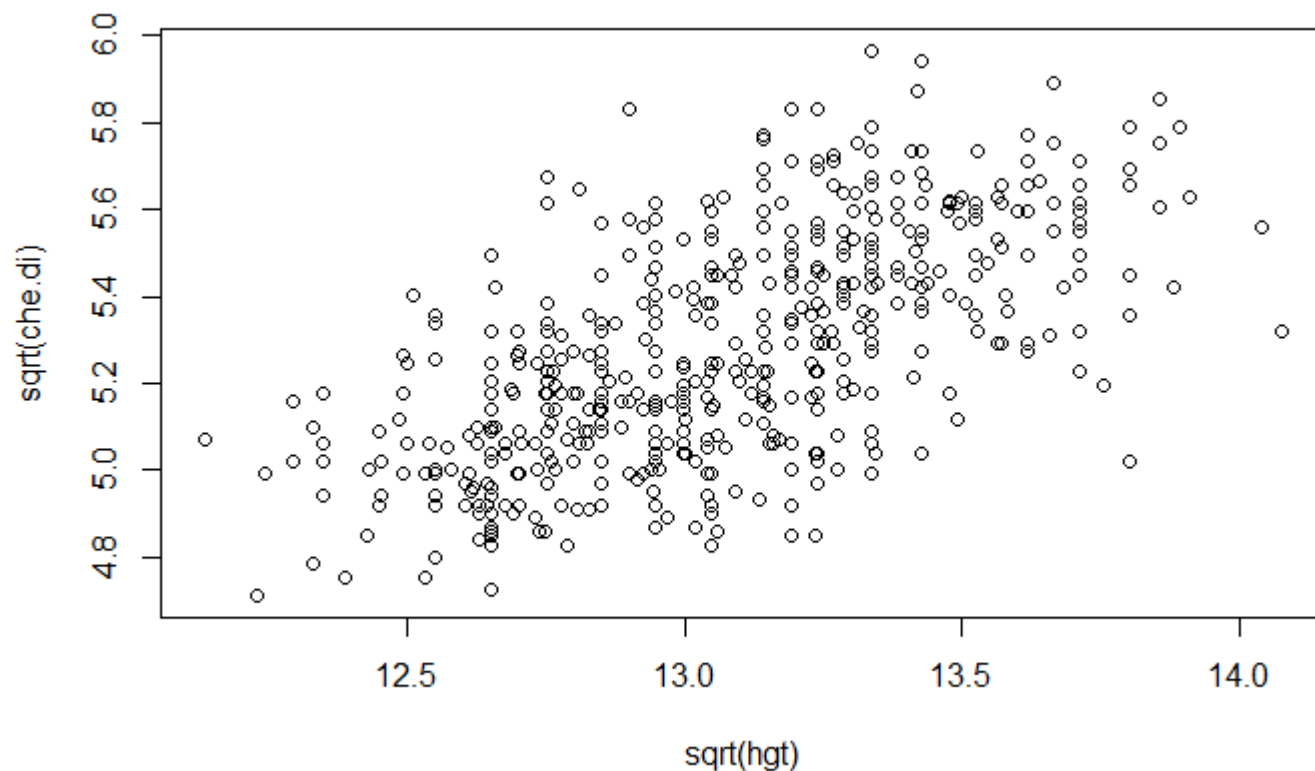

Start linear regression model fitting test Ho Data does not fit Ha Data fits

Hide

```
lmtest <- lm(log(che.di) ~ log(hgt), data = bdims_filtered)
lmtest %>% summary()
```

```
Call:
lm(formula = log(che.di) ~ log(hgt), data = bdims_filtered)

Residuals:
     Min      1Q   Median      3Q      Max
-0.22123 -0.05121  0.00044  0.05057  0.22962

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.42566    0.31592  -7.678 8.43e-14 ***
log(hgt)     1.11888    0.06145  18.208  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0759 on 505 degrees of freedom
Multiple R-squared:  0.3963,    Adjusted R-squared:  0.3951
F-statistic: 331.5 on 1 and 505 DF,  p-value: < 2.2e-16
```

Hide

```
lmtest %>% confint()
```

```
                 2.5 %     97.5 %
(Intercept) -3.0463489 -1.804974
log(hgt)     0.9981551  1.239607
```

P valuse is very small as p < 0.001. The model is statistically significant F Statistic is F(1,505) = 331.5

Model coeffecients are as follows

alpha (intercept) = -2.42566 == -2.43 beta (slope) = 1.11888 == 1.12

The slope represents the average change in y following a one unit increase in x. As log(hgt) increases by 1 unit, log(che.di) changes on average by 1.12

Not required to perform as the p value is < 0.001 (Hypothesis testing for the model coeffecients

Intercept Ho alpha = 0 Ha alpha is not equal to 0

Slope Ho beta = 0 Ha beta is not equal to 0)

Error residual is as follows

y = yi - ycap

yi for che.di = 30.3 is log(30.3) = 3.4111

ycap (predicted value) for che.di = 30.3 and hgt = 180.0 3.3861

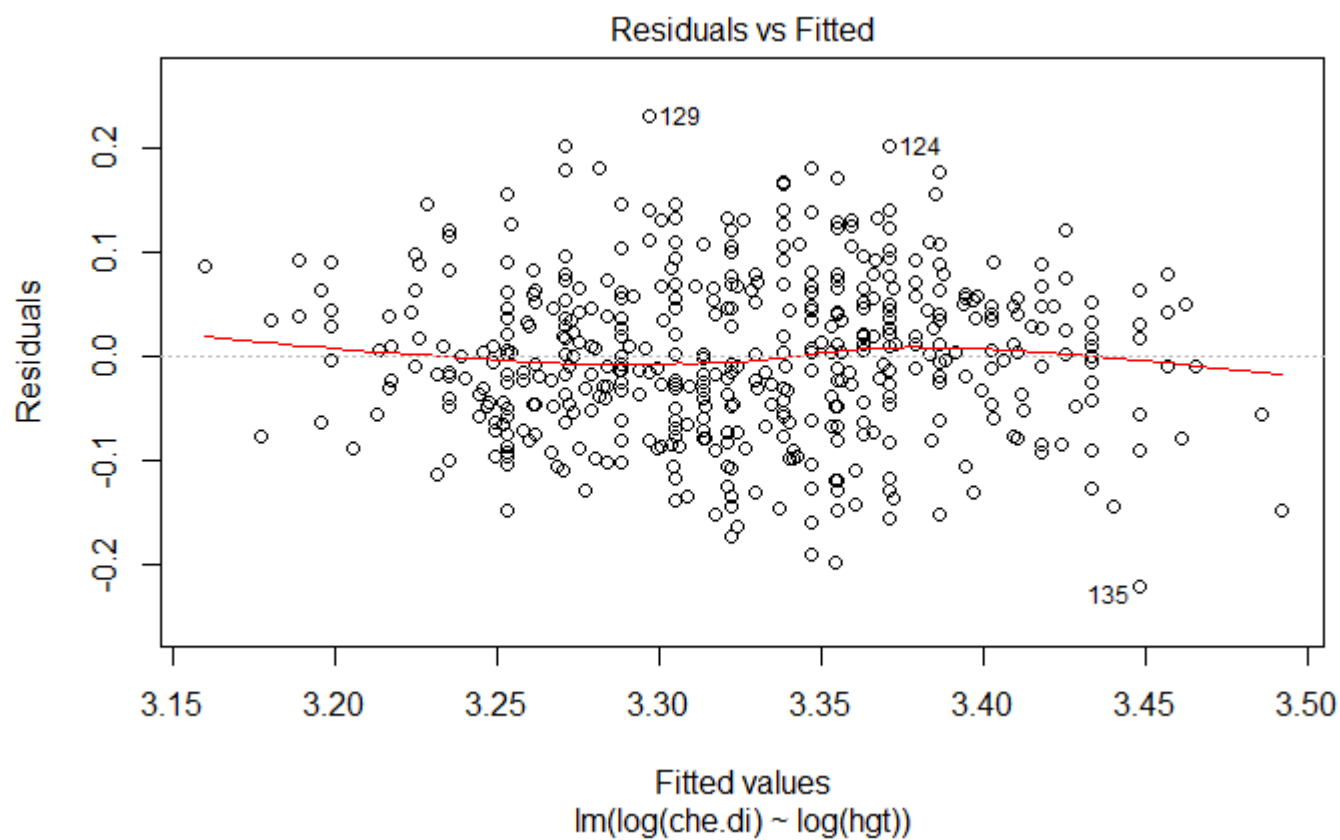hence y = 3.4111 - 3.3861 = 0.025

Testing assumptions :-

Independence - Proved because of the data

Linearity - Not perfectly linear (plot 1) Normality of residuals - Proved from lot 2 (most points lie on the black line)
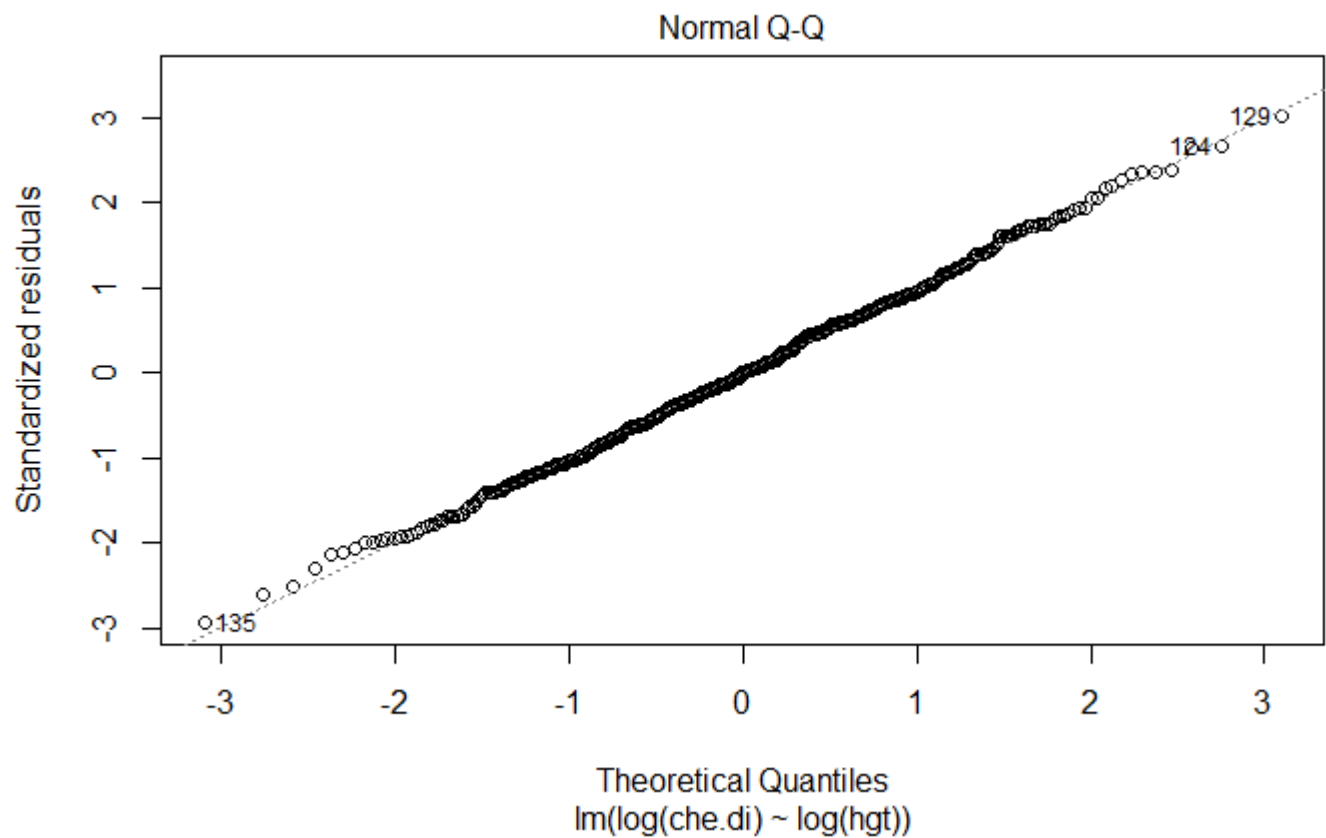Homoscedasticity - Proved from plot 3 as the red line is almost flat

Hide

```
lmtest %>% plot(which = 1)
```

Residuals vs Fitted

Fitted values
lm(log(che.di) ~ log(hgt))

Hide

```
lmtest %>% plot(which = 2)
```

Normal Q-Q

lm(log(che.di) ~ log(hgt))

Hide

```
lmtest %>% plot(which = 3)
```

Scale-Location

lm(log(che.di) ~ log(hgt))
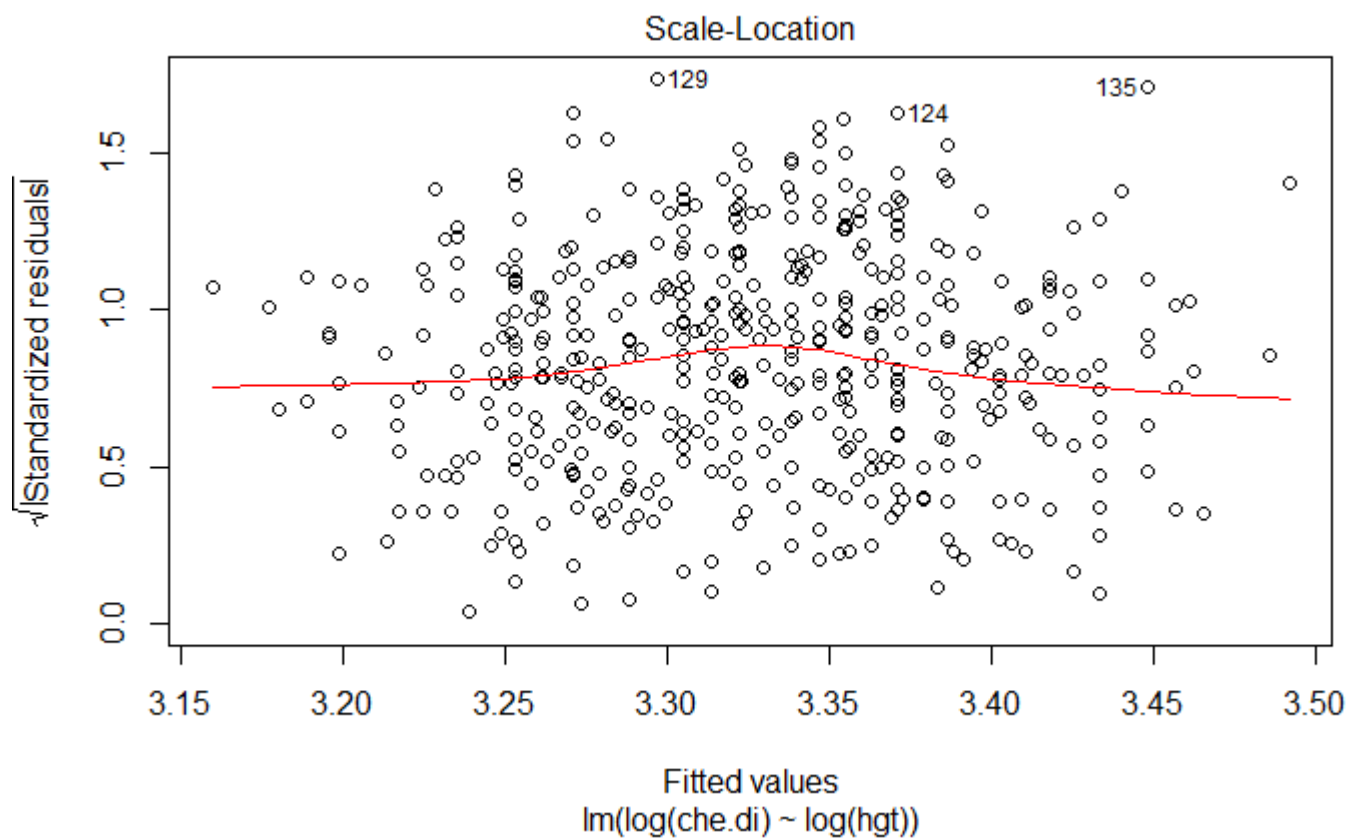
```
lmtest %>% plot(which = 4)
```


Cook's distance
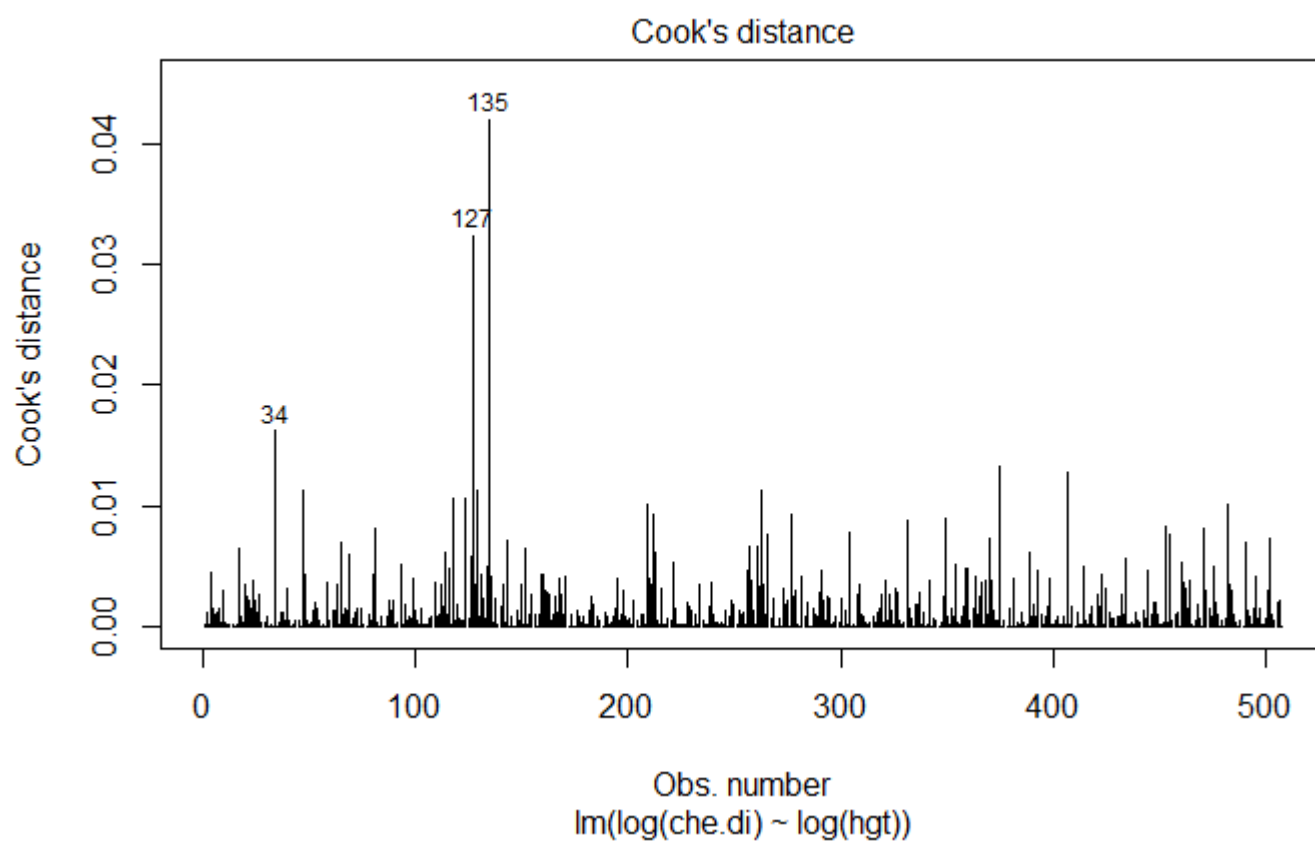
```
lmtest %>% plot(which = 5)
```
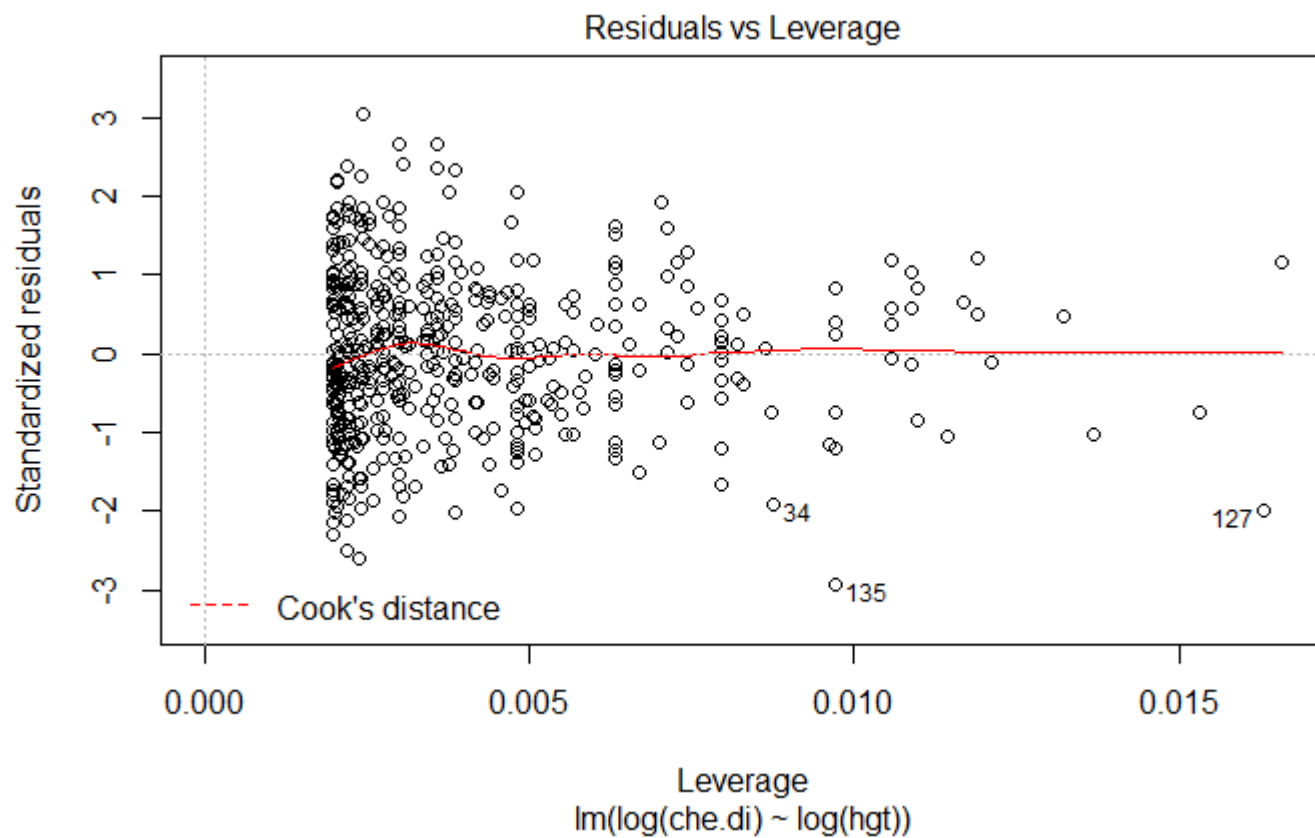
Residual vs Leverage

In the example below, no values fall close to 0.5 band. In fact, the bands are not even visible

Hide

```
lmtest %>% plot(which = 5)
```

Residuals vs Leverage
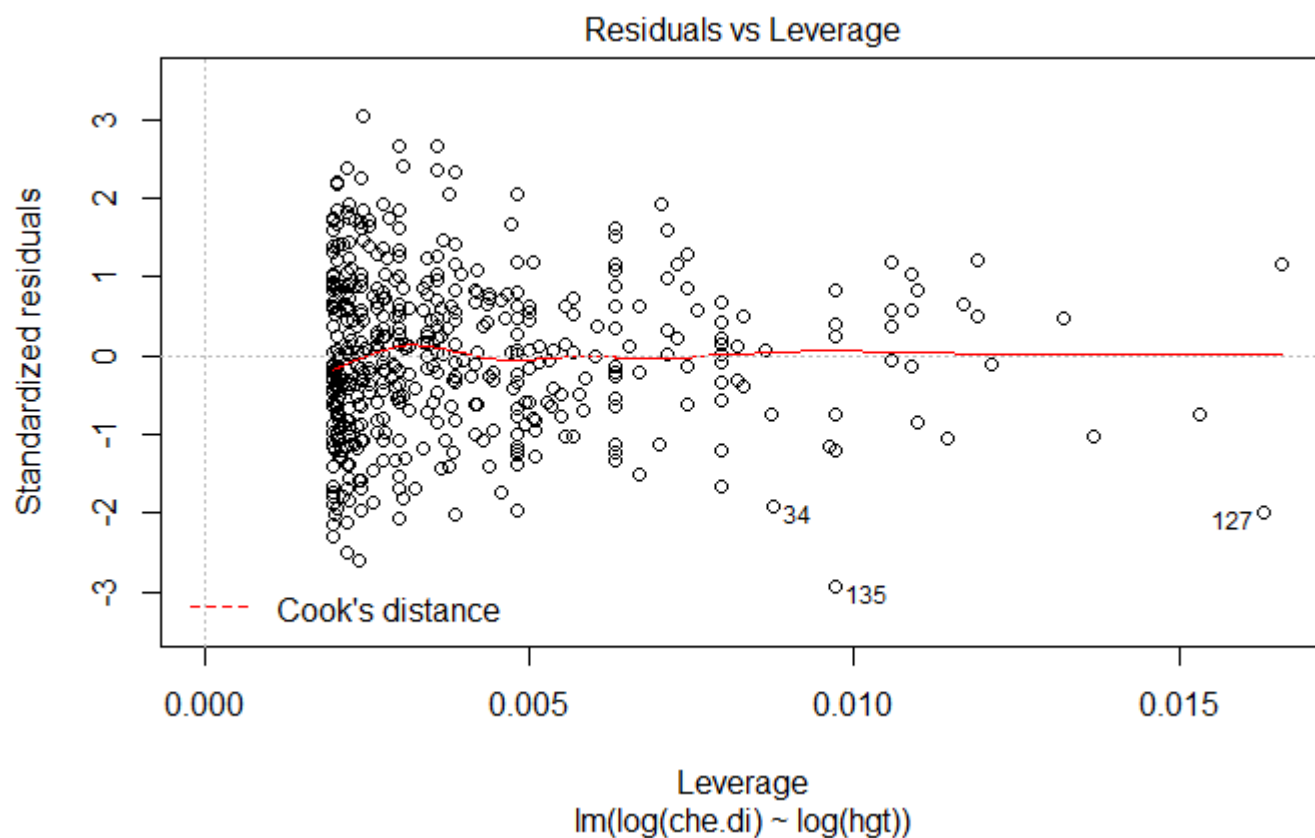
Linear Regression - Strength and Direction of Linear Relationships

Correlation coeffecient

Hide

```
r <- cor(log(bdims_filtered$che.di), log(bdims_filtered$hgt), use = "complete.obs")

r
```

```
[1] 0.6295466
```

CI for Correlation coeffecient

Hide

```
install.packages("psychometric")
```

```
WARNING: Rtools is required to build R packages but is not currently installed. Please download
and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 慟牰C:/Users/smart/OneDrive/Documents/R/win-library/3.6慟牰
(as 慟牰lib慟牮 is unspecified)
also installing the dependency 慟牰multilevel慟牮

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/multilevel_2.6.zip'
Content type 'application/zip' length 433488 bytes (423 KB)
downloaded 423 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/psychometric_2.2.zip'
Content type 'application/zip' length 194016 bytes (189 KB)
downloaded 189 KB
```

```
package 'multilevel' successfully unpacked and MD5 sums checked
package 'psychometric' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\smart\AppData\Local\Temp\RtmpgTOgno\downloaded_packages
```

Hide

```
library(psychometric)
```

```
Loading required package: multilevel
Loading required package: nlme

Attaching package: 慟牰nlme慟牮

The following object is masked from 慟牰package:dplyr慟牮:

    collapse

Loading required package: MASS

Attaching package: 慟牰MASS慟牮

The following object is masked from 慟牰package:dplyr慟牮:

    select
```

Hide

```
CIr(r = r, n = length(bdims_filtered$hgt), level = .95)
```

```
[1] 0.5739283 0.6793837
```

Hide

```
detach("package:psychometric", unload=TRUE)
```

Simple linear regression summary:

Linearity was assumed, normality of residuals OK, homoscedasticity OK, no influential cases. r = 0.6296, r_sqr = 0.3964 Model ANOVA, F(1,505) = 331.5, p< 0.001 a = -2.43, p< 0.001, 95% CI (-3.0463489 -1.804974) b = 1.12, p< 0.001, 95% CI (0.9981551 1.239607)

Decision:

Overall model: Reject Ho. Intercept: Reject Ho. Slope: Reject Ho. log(chest diameter) = -2.43 + 1.12log(Height) What do we conclude?: There was a statistically significant positive linear relationship between chest diameter and height.

Making predicitions Take a random example from the dataset and make predictions, refer to the slide

Hide

```
bdims_filtered %>% summarise(Min = min(hgt,na.rm = TRUE),
          Q1 = quantile(hgt,probs = .25,na.rm = TRUE),
          Median = median(hgt, na.rm = TRUE),
          Q3 = quantile(hgt,probs = .75,na.rm = TRUE),
          Max = max(hgt,na.rm = TRUE),
          Mean = mean(hgt, na.rm = TRUE),
          SD = sd(hgt, na.rm = TRUE),
          n = n(),
          Missing = sum(is.na(hgt)))
```

Min	Q1	Median	Q3	Max	Mean	SD	n	Missing
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
147.2	163.8	170.3	177.8	198.1	171.1438	9.407205	507	0

1 row

Hide

```
bdims_filtered %>% summarise(Min = min(che.di,na.rm = TRUE),
              Q1 = quantile(che.di,probs = .25,na.rm = TRUE),
              Median = median(che.di, na.rm = TRUE),
              Q3 = quantile(che.di,probs = .75,na.rm = TRUE),
              Max = max(che.di,na.rm = TRUE),
              Mean = mean(che.di, na.rm = TRUE),
              SD = sd(che.di, na.rm = TRUE),
              n = n(),
              Missing = sum(is.na(che.di)))
```

Min	Q1	Median	Q3	Max	Mean	SD	n	Missing
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
22.2	25.65	27.8	29.95	35.6	27.97377	2.74165	507	0

1 row