# Homework 5

Matthew Bentz

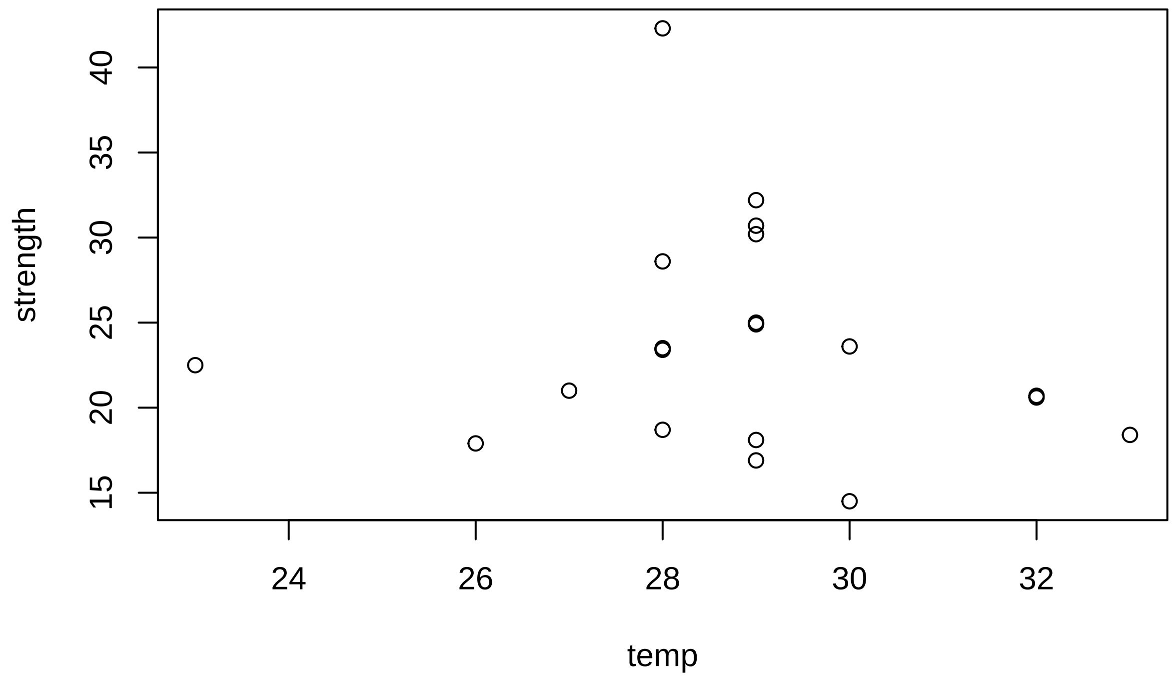
4/25/2022

## 1.

$$
\begin{aligned}
r\frac{S_y}{S_x} &= \frac{S_{xy}}{S_{xx}} \\
r &= \frac{S_{xy}\sqrt{S_{xx}}}{S_{xx}\sqrt{S_{yy}}} \\
r &= \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \\
S_{xy} &= r\sqrt{S_{xx}}\sqrt{S_{yy}} \\
b_1 &= \frac{r\sqrt{S_{xx}}\sqrt{S_{yy}}}{S_{xx}} \\
b_1 &= r\frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}} \\
b_1 &= r\frac{S_y}{S_x}
\end{aligned}
$$

## 2.

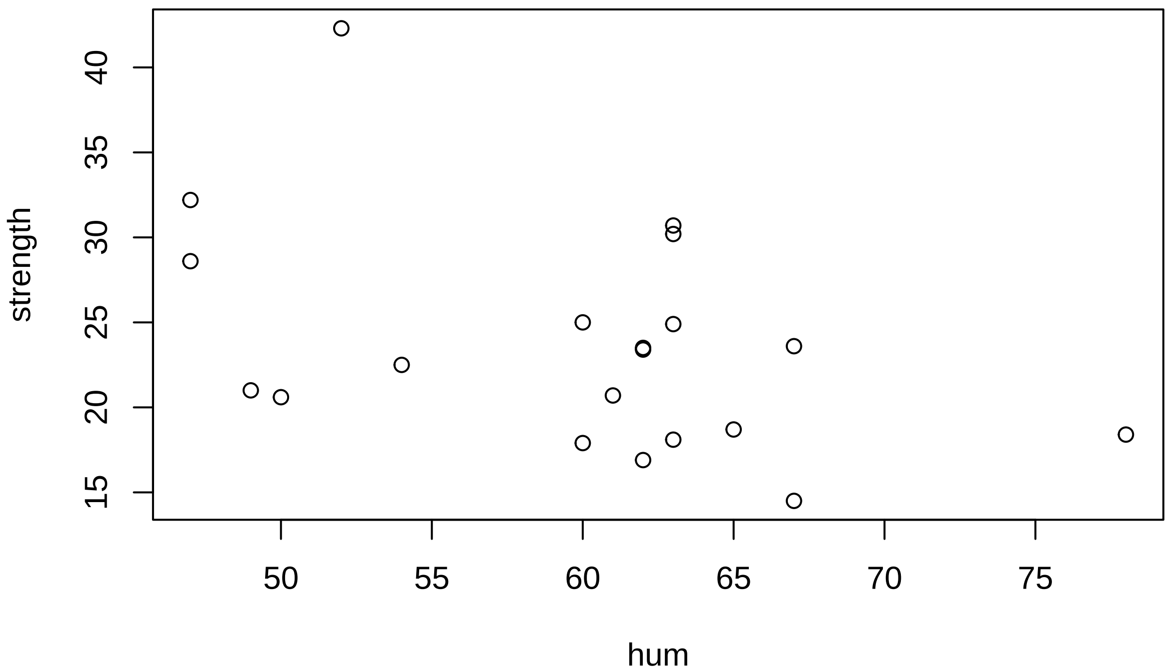
```r
temp <- c(30,29,33,29,28,29,29,32,32,29,29,28,30,28,28,27,26,28,29,23)
hum <- c(67,63,78,62,47,47,60,61,50,63,63,52,67,62,65,49,60,62,63,54)
strength <- c(23.6, 18.1,18.4,16.9,28.6,32.2,25.0,20.7,20.6,30.7,30.2,42.3,14.5,23.4,18.7,21.0,17.9,23.5
weight <- c(11.58, 11.61,12.04,12.98,10.77,11.23,11.91,12.05,11.83,11.26,11.75,12.03,13.40,12.82,12.63,

plot(strength ~ temp)
```




```
cor(strength, temp)
```

```
## [1] -0.1368723
```

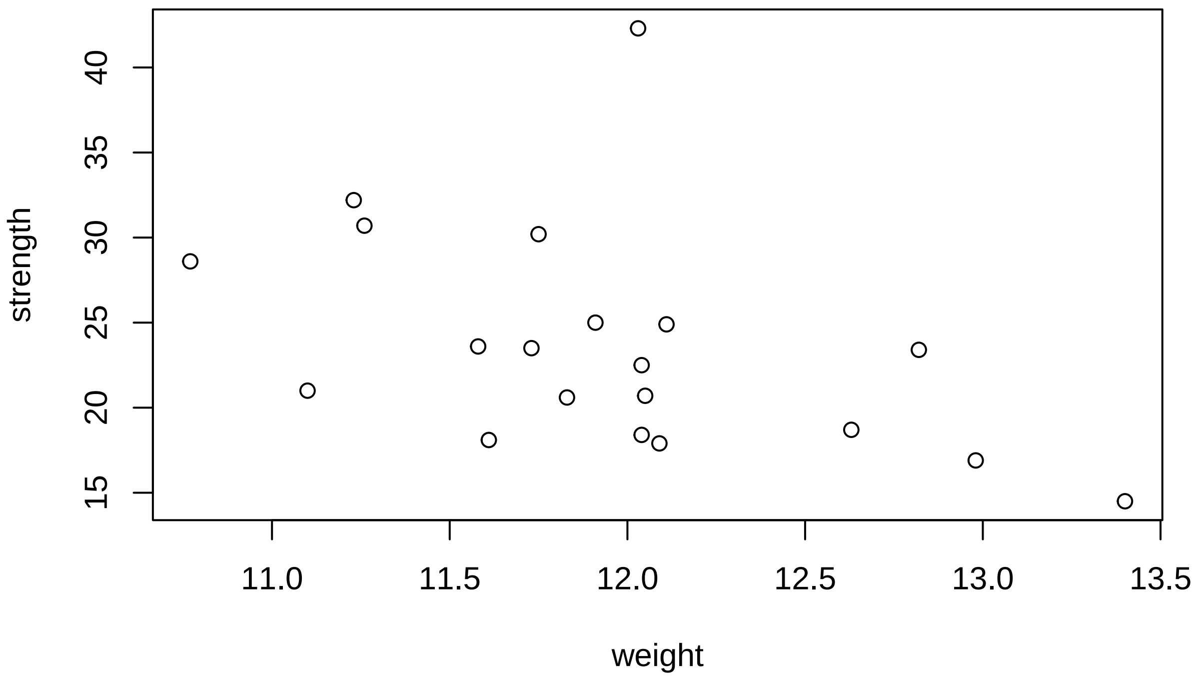
```
plot(strength ~ hum)
```




```
cor(strength, hum)
```

```
## [1] -0.4449783
```

```
plot(strength ~ weight)
```



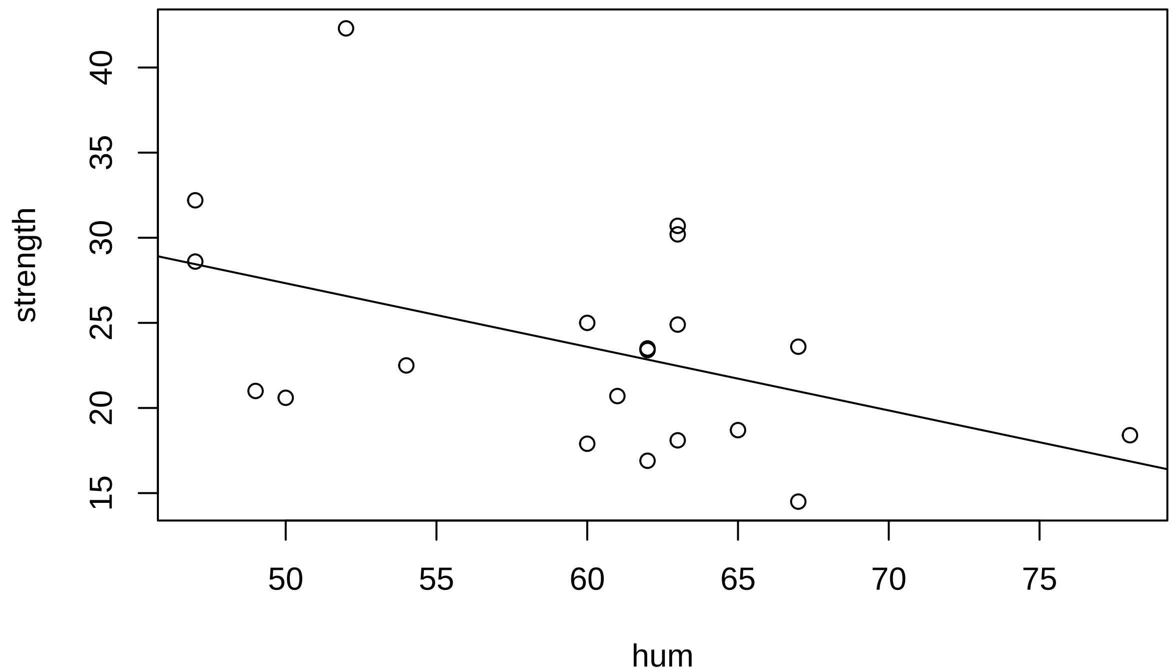

```
cor(strength, weight)
```

```
## [1] -0.466553
```

There is little to no relationship between temperature and cone strength. The linear correlation coefficient shows a weak negative correlation of -0.137. The relationship between humidity and cone weight on cone strength shows a similarly stronger correlation of -0.445 and -0.467 respectively. A decrease in humidity and cone weight will give a noticeably positive effect on cone strength.

## 3.

```
plot(strength ~ hum)
abline(lm(strength ~ hum))
```




```r
fit = lm(strength ~ hum) # Fits the regression model
summary(fit) # To see all the good stuff
```

```
##
## Call:
## lm(formula = strength ~ hum)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7270 -4.7011  0.3539  2.4776 15.7200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46.0042    10.6726   4.311 0.000421 ***
## hum          -0.3735     0.1772  -2.108 0.049302 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.019 on 18 degrees of freedom
## Multiple R-squared:  0.198,  Adjusted R-squared:  0.1535
## F-statistic: 4.444 on 1 and 18 DF,  p-value: 0.0493
```

```r
mean(residuals(fit)^2)
```

```
## [1] 32.60289
```

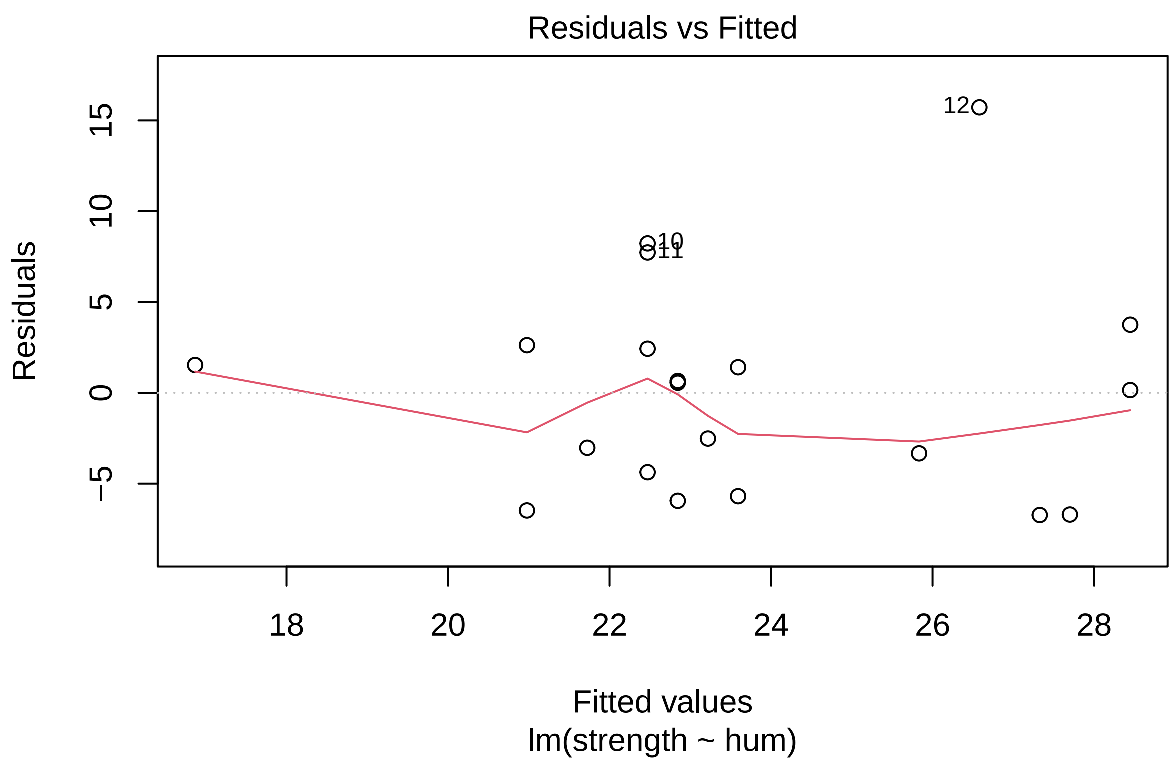
Humidity is statistically significant in modeling cone strength. We can see that from the p-value of 0.0493 at a confidence level of $\alpha = 0.05$ that we would reject the null hypothesis that there is no significant effect of humidity on cone strength. The coefficient $b_1$ of the model shows that with a one unit increase in humidity, there is a -0.3735 unit decrease in cone strength.

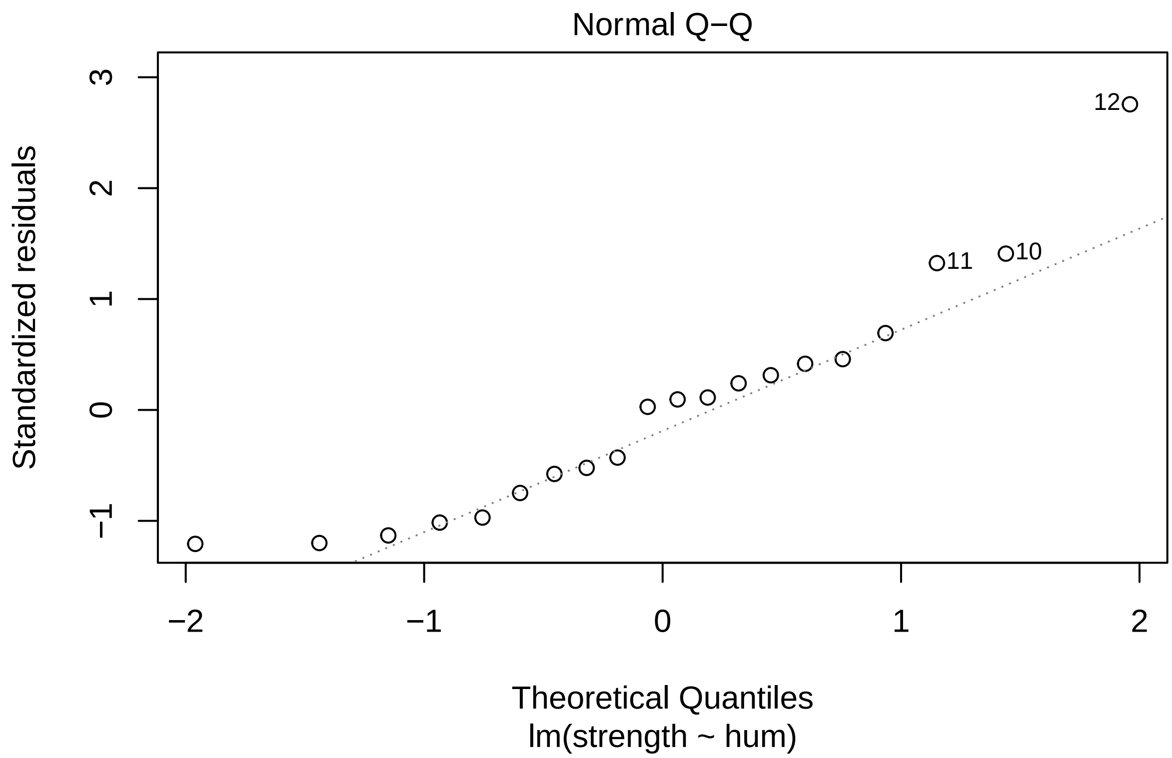
## 4.

19.8% of the total variability of cone strength is explained by relative humidity. The mean square error of the model is $6.019^2$ or 36.23.

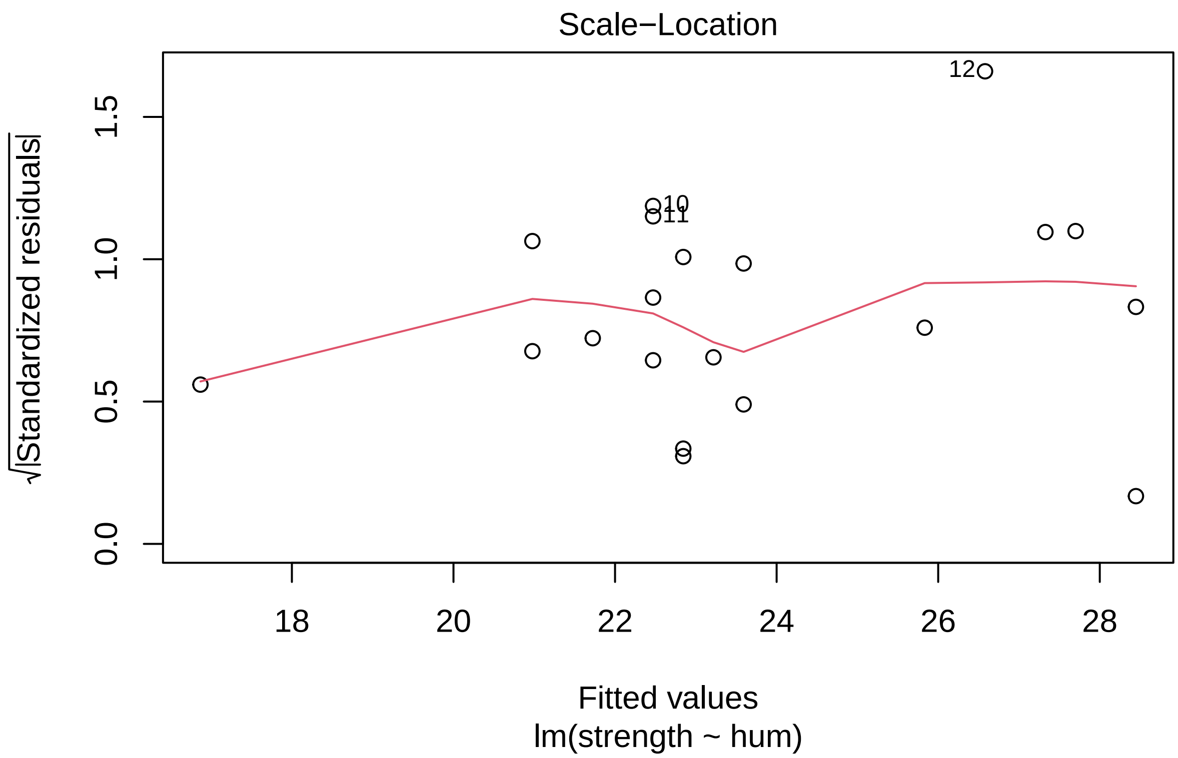
## 5.

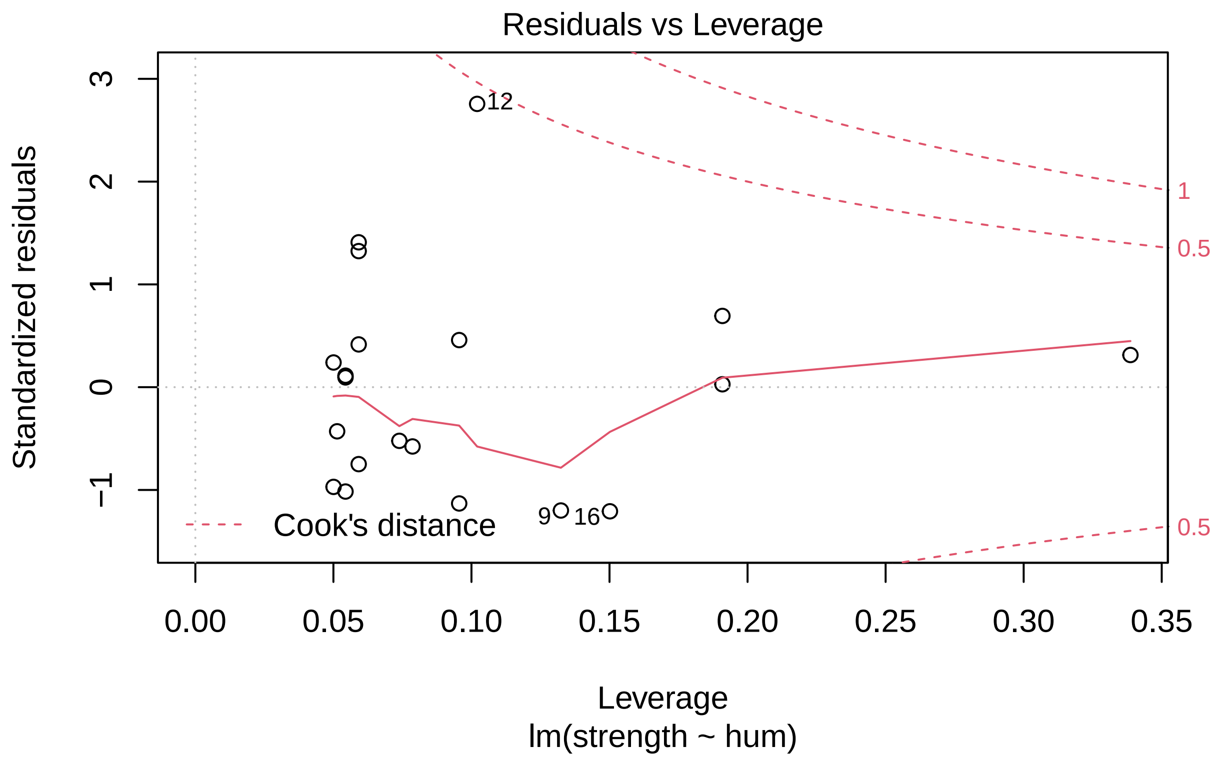
```
plot(lm(strength ~ hum))
```





Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(strength ~ hum)
−2
−1
0
1
2
3
12
11
10


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
12
10
11
18
20
22
24
26
28
Fitted values
lm(strength ~ hum)




The errors are normally distributed, as shown in the Normal Q-Q plot with points generally on the line. There is a linear relationship between humidity and cone strength, as shown in the Residuals vs Fitted plot with the line having a relatively small amount of variability and influx. Point 12 could be considered influential, since it is close to the 0.5 line on the leverage graph. It can also be considered an outlier based on its high redidual error from the Residuals vs Fitted graph.

### 6.

```
allFit <- lm(strength ~ hum + temp + weight)
summary(allFit)
```

```
##
## Call:
## lm(formula = strength ~ hum + temp + weight)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -8.062 -2.690 -1.046  2.669 17.035
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 77.5448372 32.2927425   2.401   0.0288 *
## hum         -0.2388803  0.2251560  -1.061   0.3045
## temp        -0.0005644  0.7075167  -0.001   0.9994
```

```
## weight      -3.3118920  2.4970630  -1.326   0.2034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.052 on 16 degrees of freedom
## Multiple R-squared:  0.2792, Adjusted R-squared:  0.1441
## F-statistic: 2.066 on 3 and 16 DF,  p-value: 0.1451
```

When all the variables are included in the model, none of them are statistically significant. We can see this from the p-values being much greater than our $\alpha$ at 0.05.

## 7.

```
strFit <- lm(strength ~ weight)
summary(strFit)
```

```
##
## Call:
## lm(formula = strength ~ weight)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.182 -2.788 -1.489  2.426 19.003
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   80.141     25.262   3.172  0.00527 **
## weight        -4.725      2.111  -2.238  0.03811 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.945 on 18 degrees of freedom
## Multiple R-squared:  0.2177, Adjusted R-squared:  0.1742
## F-statistic: 5.008 on 1 and 18 DF,  p-value: 0.03811
```

This model's $R^2$ is similar to the $R^2$ of the previous question - even less. Regardless, I would choose the model with only the cone weight as a predictor because it has statistical significance as well as a stronger and more predictable relationship with cone strength.

## 8.

From the previous model, we can use the values of 80.141 as our $b_0$ and -4.725 as our $b_1$. This will provide us with the equation $\widehat{strength} = b_0 + b_1 weight$ or $\widehat{strength} = 80.141 - 4.725 weight$. Plugging in 12g for our value of weight, we find that the predicted value of cone strength is 23.441 N.