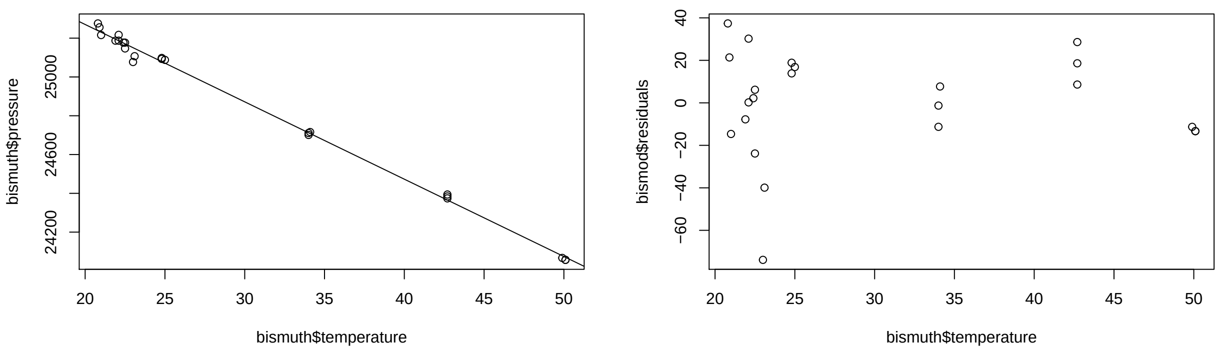
M440B HW 12

Erick Castillo

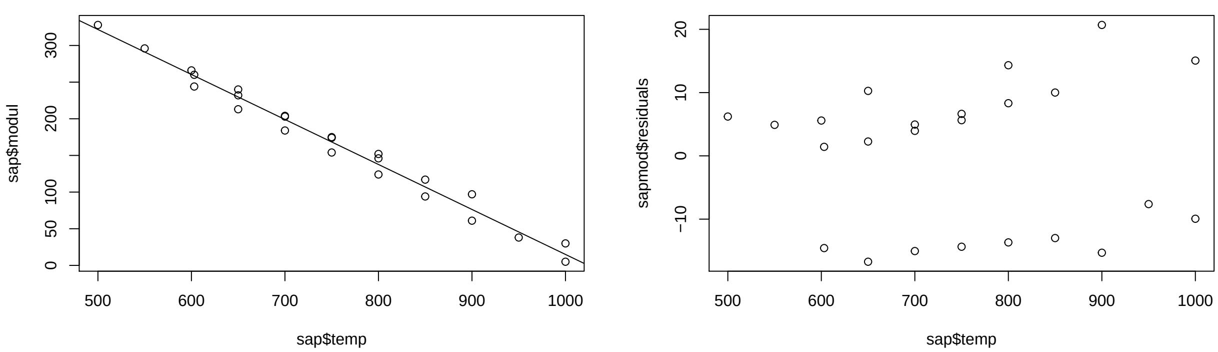
4/7/2021

Problem 36: Notice that the question states that pressure is a function of temperature meaning that pressure is the response variable.



The left plot exhibits the OLS fit of the data while the right plot is the residual plot from the OLS model against the predictor variable. There appears to be heteroskedasticity present in the model because the residuals at one point are very extreme when compared to the rest of the other observations .

Problem 38: The response variable in this case is the Young's modulus (g) while temperature is a predictor variable.



On the left is the scatter plot of the data along with the fitted line. On the right is the residual plot of the data. Notice that there appears to be a trend. That is, the errors do not appear to be random. It appears that there are two seperate collections of points, one on the top, the other on the bottom.

The question further asks to construct confidence intervals for the fitted $\hat{\beta}_i$, $i = \{0, 1\}$. This can be easily done with the following code:

```
confint(sapmod)
```

```
##                   2.5 %      97.5 %
## (Intercept) 600.7816841 656.4758665
## temp         -0.6502121  -0.5771666
```

but that would be cheating. Instead, let's use the summary table provided by R and the following information:

$$CI : \hat{\beta}_i \pm t_{n-2,\frac{\alpha}{2}}(SE_{\hat{\beta}_i}),\ i = \{0, 1\}$$

```
##
## Call:
## lm(formula = modul ~ temp, data = sap)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.731 -13.164   4.427   7.059  20.692
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 628.62878   13.42758   46.82   <2e-16 ***
## temp         -0.61369    0.01761  -34.85   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.72 on 22 degrees of freedom
## Multiple R-squared:  0.9822, Adjusted R-squared:  0.9814
## F-statistic:  1214 on 1 and 22 DF,  p-value: < 2.2e-16
```

Using the above output, I can then calculate the 95% confidence interval for $\hat{\beta}_0$ as follows.

```
quantiles1 = qt(c(0.025,0.975), df = length(sap$temp)-2)
se0 = 13.428
se1 = 0.0176

#confidence interval for beta_0
sapmod$coefficients[1]+quantiles1*se0
```

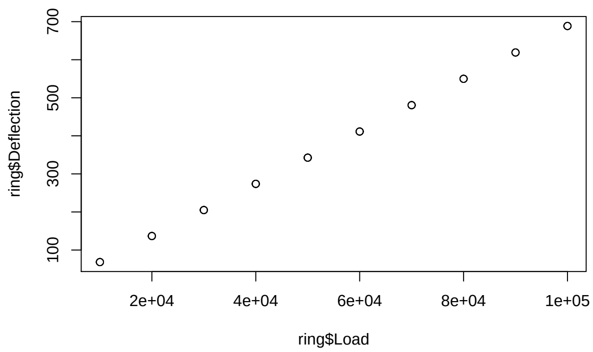
```
## [1] 600.7808 656.4767
```

```
#confidence interval for beta_1
sapmod$coefficients[2]+quantiles1*se1
```

```
## [1] -0.6501895 -0.5771892
```

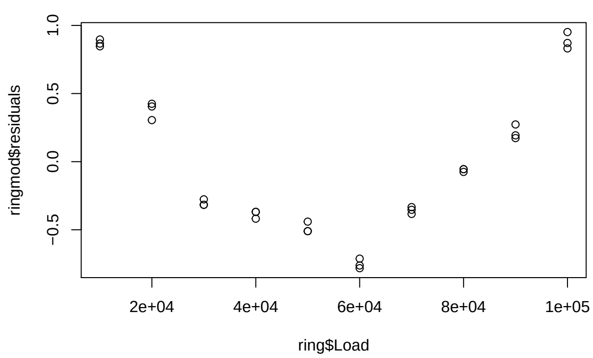
Notice these values are close to the ones produced by the confint() function.

Problem 40A: Plot load against deflection. Does the plot appear linear?



The plot appears to be linear at first glance, however, notice that the x-axis units are increasing at a much faster rate when compared to the y-axis units. This is the first indicator that this fit may not be linear in nature.

B. Run a linear regression and plot the residuals. Examine and comment on the fit.

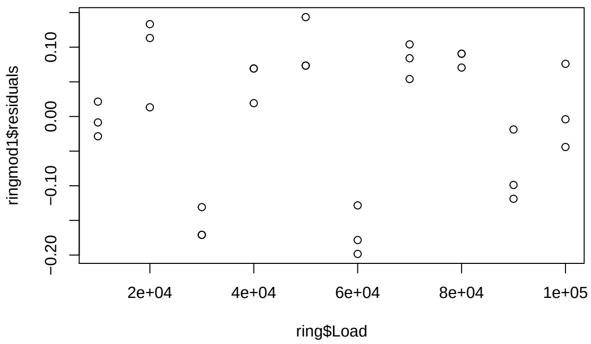


Notice that the residuals appear quadratic in nature. This means that a linear function would not be the best model for this data. Instead, maybe a quadratic fit would work better.

C. Fit deflection as a quadratic function of load. Estimate coefficients and standard errors. Plot the residuals. Comment on the fit.

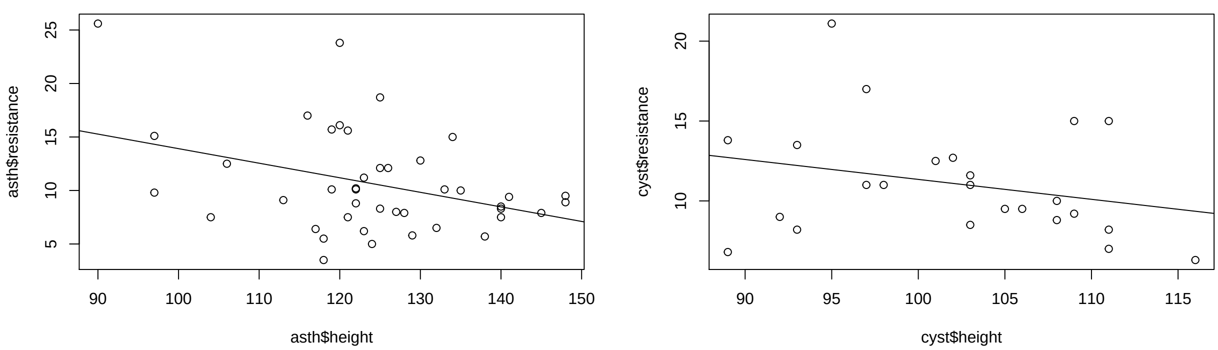
```
##
## Call:
## lm(formula = Deflection ~ Load + I(Load^2), data = ring)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19832 -0.08509  0.02033  0.07533  0.14339
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 1.363e-01  7.284e-02    1.871   0.0722 .
## Load        6.812e-03  3.042e-06 2239.355   <2e-16 ***
```

```
## I(Load^2)   7.294e-10  2.695e-11   27.065   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1073 on 27 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 5.11e+07 on 2 and 27 DF,  p-value: < 2.2e-16
```



Notice that the residuals in the above plot do not follow some sort of pattern. This is an indication that a quadratic fit was necessary for the data. The coefficients and the standard errors are provided in the above summary table.

Problem 44: The problem asks to read in both the asthma and cystfibr files, then to consider if there is a statistically significant relation between respiratory resistance and height in either group.



The plot on the left shows the LS fit for the asthma dataset while the plot on the right shows the LS fit for the cystfibr dataset. There are several ways to see if height is a statistically significant variable in determining resistance. The easisest method is to use the summary function:

```
summary(asthmod)
```

```
##
## Call:
## lm(formula = resistance ~ height, data = asth)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.9608 -3.0408 -0.3806  1.6936 12.6113
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.5157     6.6363   4.146 0.000171 ***
## height       -0.1361     0.0531  -2.562 0.014264 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.461 on 40 degrees of freedom
## Multiple R-squared:  0.141,  Adjusted R-squared:  0.1195
## F-statistic: 6.566 on 1 and 40 DF,  p-value: 0.01426
```

The p-value corresponding with height shows that there is decent evidence to suggest that the taller an individual is, the more likely they are to be susceptible to asthma.

```
summary(cystmod)
```

```
##
## Call:
## lm(formula = resistance ~ height, data = cyst)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9168 -1.9496 -0.6576  1.2793  9.1309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.80704    9.62928   2.472   0.0216 *
## height      -0.12461    0.09411  -1.324   0.1991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.462 on 22 degrees of freedom
## Multiple R-squared:  0.07381,    Adjusted R-squared:  0.03171
## F-statistic: 1.753 on 1 and 22 DF,  p-value: 0.1991
```

This p-value indicates that there is very weak evidence to suggest that the taller an individual is, the less resistance they develop towards cyst in the lungs.