

AML-HW6

Nidia Bucarelli (nidiaib2), Sunny Katiyar (sunnyk2), Wang Xiang (wangx2)

26/04/2020

QUESTION 10.9

Loading the dataset

```
brunhild_data = read.csv("brunhild.csv", sep = "\t")
```

Building the Linear Regression Model

```
lin_mod_brunhild = lm(formula = log(Sulfate) ~ log(Hours), data = brunhild_data)
```

Producing the summary of the linear model:

```
summary(lin_mod_brunhild)
```

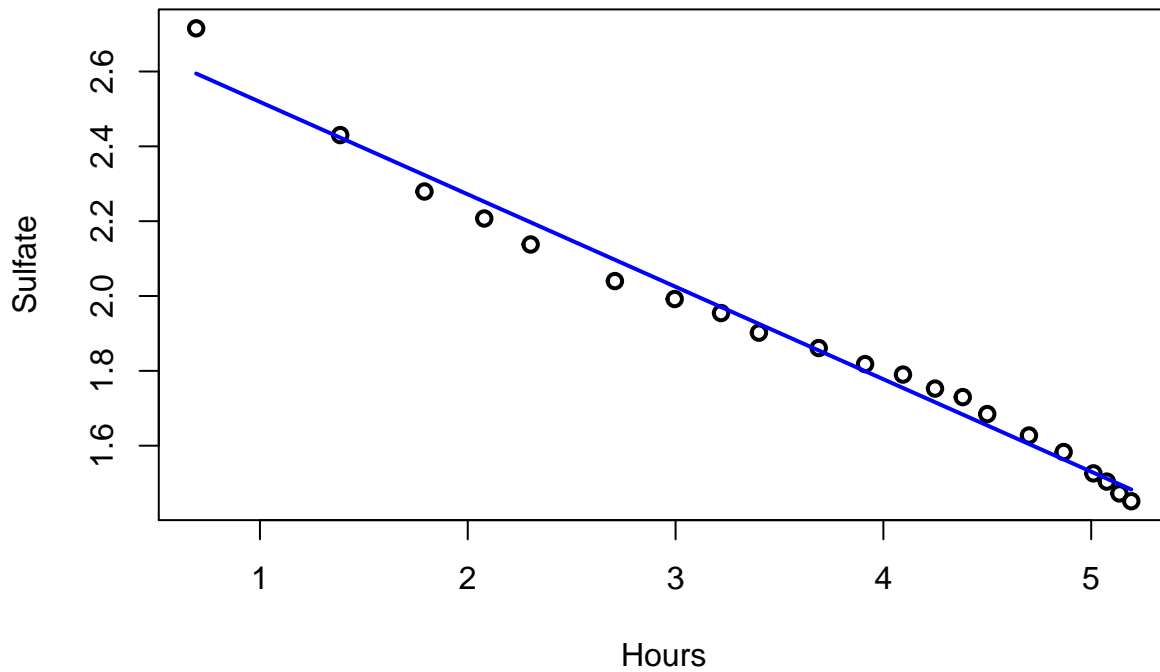
```
##
## Call:
## lm(formula = log(Sulfate) ~ log(Hours), data = brunhild_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.059282 -0.031323 -0.001922  0.022676  0.120759
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.765837   0.027682   99.91  <2e-16 ***
## log(Hours)  -0.247046   0.007244  -34.10  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04348 on 19 degrees of freedom
## Multiple R-squared:  0.9839, Adjusted R-squared:  0.9831
## F-statistic: 1163 on 1 and 19 DF, p-value: < 2.2e-16
```

(A)

A plot showing the data points and the regression line in log-log coordinates

```
plot(log(brunhild_data), main = "Data Points & Regression Line (Log-Log Coordinates)",  
     col.main = "dark orange", lwd=2)  
lines(log(brunhild_data$Hours), fitted(lin_mod_brunhild), col="blue", lwd=2)
```

Data Points & Regression Line (Log-Log Coordinates)

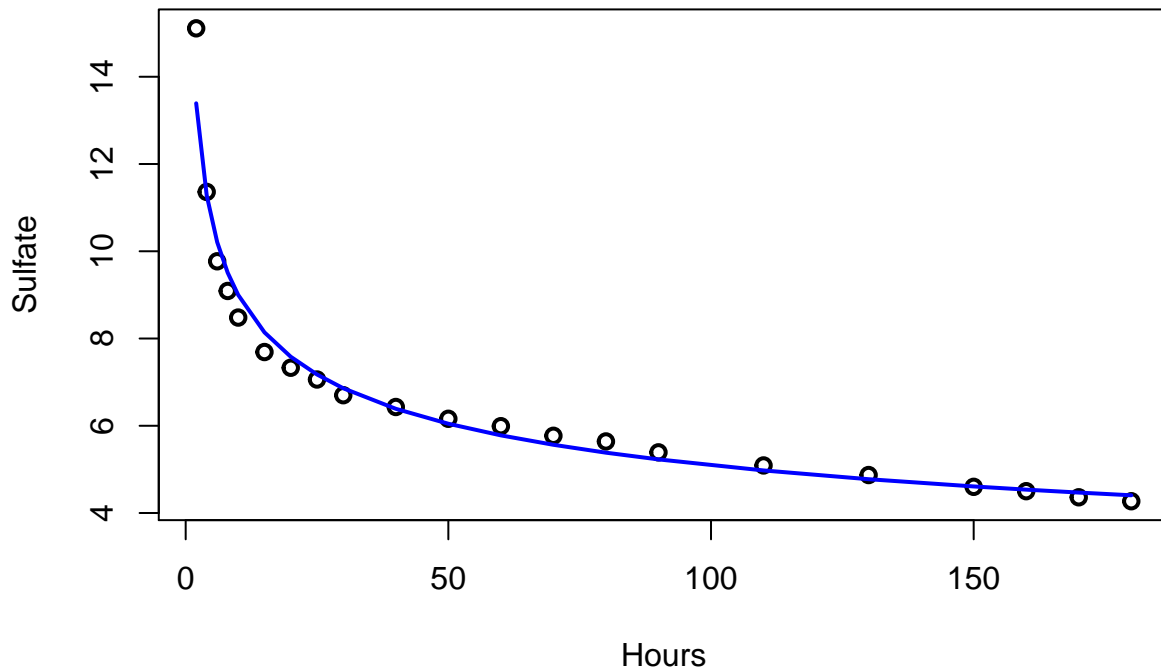


(B)

A plot showing the data points and the regression line in the original coordinates

```
plot(brunhild_data, main = "Data Points & Regression Curve (Original Coordinates)",  
     col.main = "dark orange", lwd=2)  
lines(brunhild_data$Hours, exp(fitted(lin_mod_brunhild)), col="blue", lwd=2)
```

Data Points & Regression Curve (Original Coordinates)

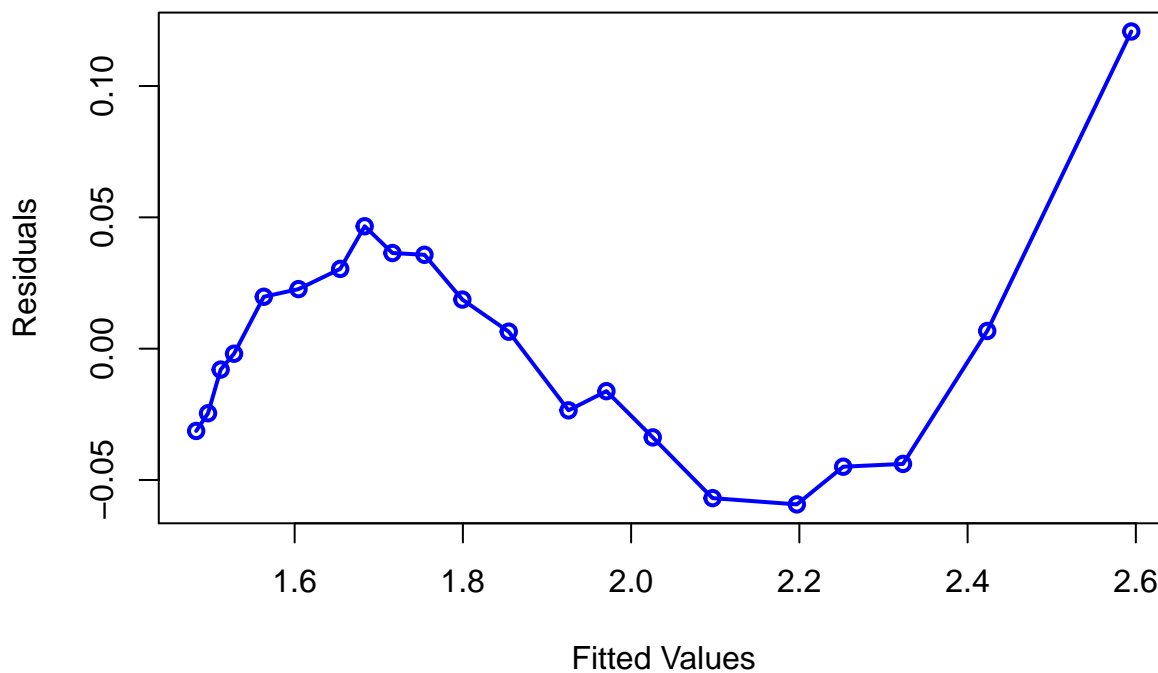


(C) Residuals against fitted values

In log-log coordinates

```
plot(fitted.values(lin_mod_brunhild), residuals(lin_mod_brunhild),  
     type = "o", col="blue", lwd=2, xlab = "Fitted Values", ylab = "Residuals",  
     main = "Residuals vs Fitted Values (Log-Log Coordinates)", col.main = "dark orange")
```

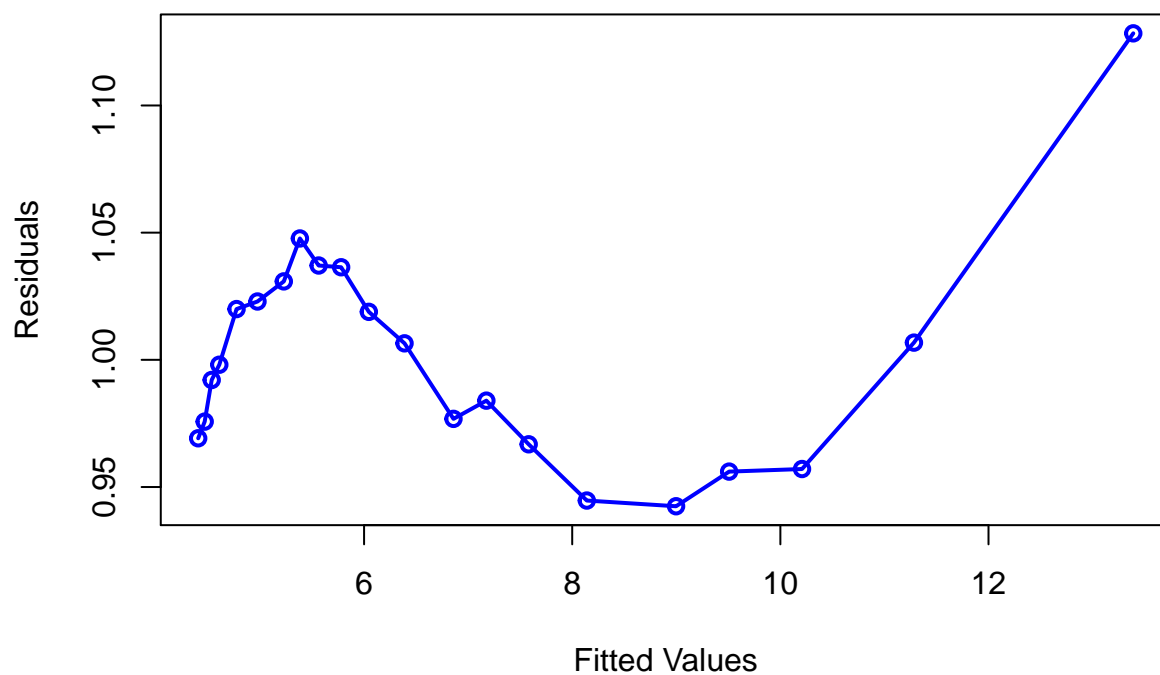
Residuals vs Fitted Values (Log-Log Coordinates)



In original coordinates

```
plot(exp(fitted.values(lin_mod_brunhild)), exp(residuals(lin_mod_brunhild)),  
     type = "o", col="blue", lwd=2, xlab = "Fitted Values", ylab = "Residuals",  
     main = "Residuals vs Fitted Values (Original Coordinates)", col.main = "dark orange")
```

Residuals vs Fitted Values (Original Coordinates)



(D) Comments on the regression fit

- From the regression line fit and data points plot, we can see that the regression line is very close to the data points. This indicates that there are not many outliers in the dataset.
- The curve trend of the Residuals Vs Fitted Values plot indicates that a non-linear fit (Spline) could give better results.

QUESTION 10.10

Loading the dataset

```
physical_data = read.csv("physical.csv", sep = "\t")
```

Building the Linear Regression Model

```
lin_mod_physical = lm(formula = Mass ~ ., data = physical_data)
```

Producing the summary of the linear model:

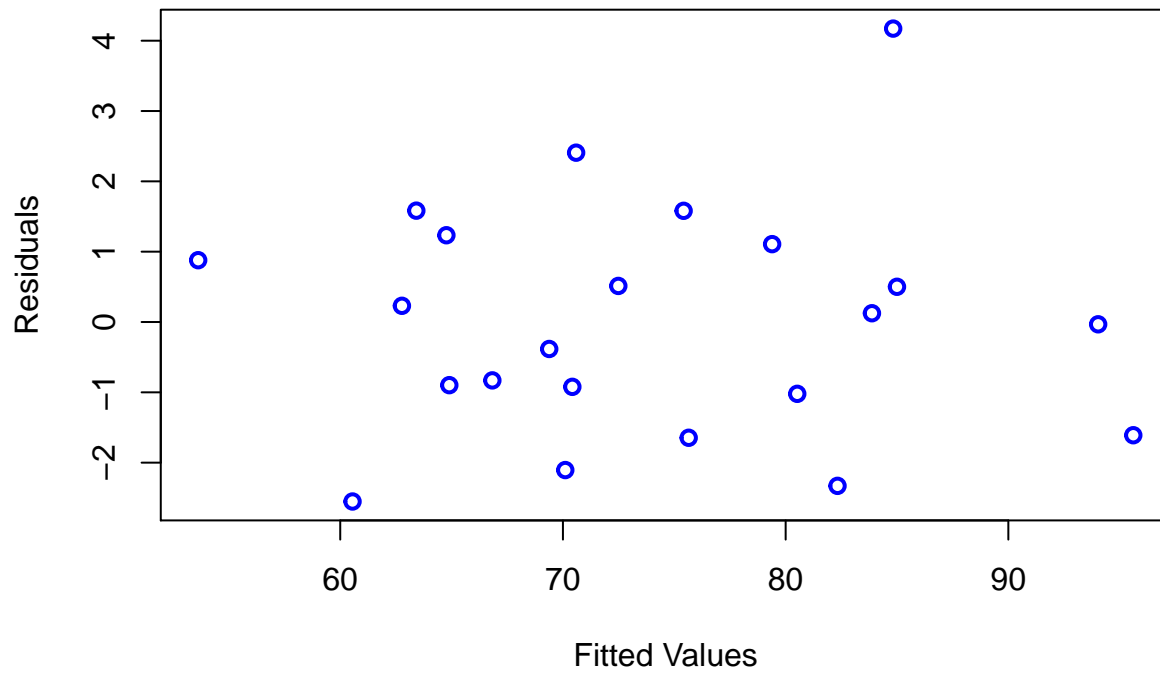
```
summary(lin_mod_physical)
```

```
##
## Call:
## lm(formula = Mass ~ ., data = physical_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5523 -0.9965  0.0461  1.0499  4.1719
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -69.51714   29.03739  -2.394  0.035605 *
## Fore         1.78182    0.85473   2.085  0.061204 .
## Bicep         0.15509    0.48530   0.320  0.755275
## Chest         0.18914    0.22583   0.838  0.420132
## Neck        -0.48184    0.72067  -0.669  0.517537
## Shoulder    -0.02931    0.23943  -0.122  0.904769
## Waist         0.66144    0.11648   5.679  0.000143 ***
## Height        0.31785    0.13037   2.438  0.032935 *
## Calf          0.44589    0.41251   1.081  0.302865
## Thigh         0.29721    0.30510   0.974  0.350917
## Head        -0.91956    0.52009  -1.768  0.104735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.287 on 11 degrees of freedom
## Multiple R-squared:  0.9772, Adjusted R-squared:  0.9565
## F-statistic: 47.17 on 10 and 11 DF, p-value: 1.408e-07
```

(A) Plot of Residuals against the Fitted Values

```
plot(fitted.values(lin_mod_physical), residuals(lin_mod_physical),  
     col="blue", lwd=2, xlab = "Fitted Values", ylab = "Residuals",  
     main = "Residuals vs Fitted Values (Original Coordinates)", col.main = "dark orange")
```

Residuals vs Fitted Values (Original Coordinates)



(B)

Fitting a linear model to regress the cube root of mass against the diameters.

```
lin_mod_physical_cube = lm(formula = Mass^(1/3)~., data = physical_data)
```

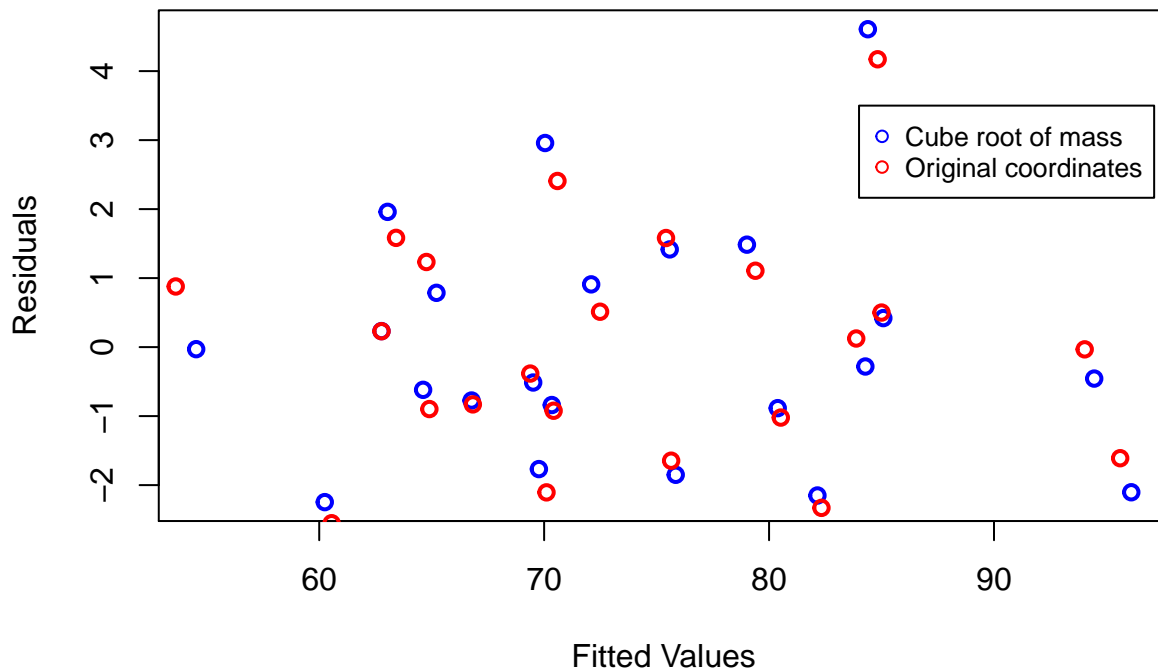
Plotting the residuals against the fitted values

```
#par(mfrow = c(1, 2))
plot(fitted.values(lin_mod_physical_cube)^3,
     (physical_data$Mass - fitted.values(lin_mod_physical_cube)^3),
     col="blue", lwd=2, xlab = "Fitted Values", ylab = "Residuals",
     main = "Residuals vs Fitted Values", col.main = "dark green")

points(fitted.values(lin_mod_physical), residuals(lin_mod_physical),
       col="red", lwd=2)
#xlab = "Fitted Values", ylab = "Residuals",
#main = "Original Coordinates", col.main = "dark orange")

# Add a legend
legend(84, 3.5, legend=c("Cube root of mass", "Original coordinates"),
      col=c("blue", "red"), pch=1, cex=0.8)
```

Residuals vs Fitted Values



(C) Comments on the regression fit

- From the Residuals vs Fitted Values plot, we can see that the residual values of both the models are very close to each other. However, the overall magnitude of the residuals in the model with original coordinates is slightly lower than that in the model with cube root of mass.
- Moreover, the model with linear coordinates is comparatively easier to interpret. Therefore, the regression model in the original coordinates is better than the other model with cube root of mass.