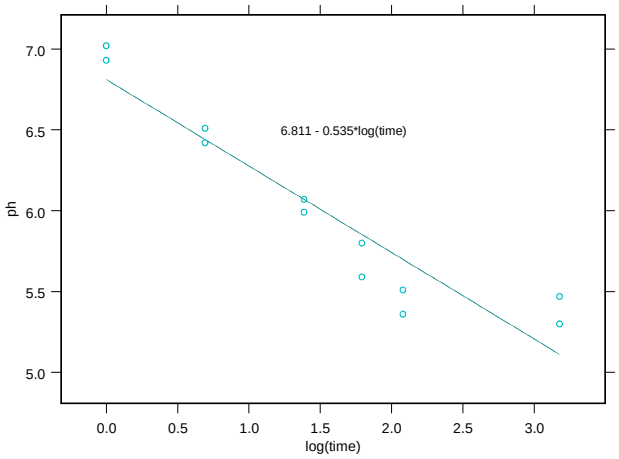
Assessing Model Fit
Continuation of Meat Processing Example: Ex 8.16

Same data as we analyzed previously plus 2 additional measurements taken at 24 hours. Scatterplot plus OLS regression line using log(time) as X and pH as Y.



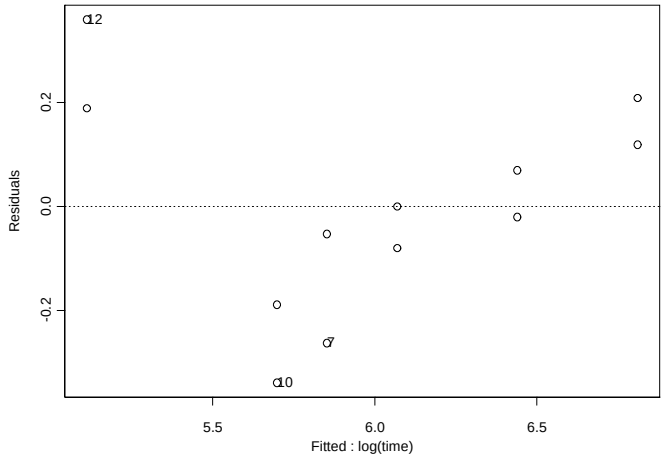
Is the model OK? Can we proceed with confidence intervals and prediction intervals?

S-PLUS OUTPUT:

```
Call: lm(formula = ph ~ log.time, data = Ex0816, na.action = na.omit)
Residuals:
        Min          1Q      Median          3Q         Max
 -0.3389731 -0.1070994 -0.0102258 0.1360878 0.3587868

Coefficients:
                  Value  Std. Error     t value    Pr(>|t|)
(Intercept)   6.8114785   0.1112888  61.2054319   0.0000000
   log.time  -0.5350021   0.0608994  -8.7850099   0.0000051

Residual standard error: 0.2135464 on 10 degrees of freedom
Multiple R-Squared: 0.8852901
F-statistic: 77.1764 on 1 and 10 degrees of freedom,
the p-value is 5.140138e-006

Analysis of Variance Table

Response: ph

Terms added sequentially (first to last)
          Df   Sum of Sq     Mean Sq     F Value            Pr(F)
 log.time  1 3.519404215 3.519404215 77.17639914 5.140137971e-006
Residuals 10 0.456020785 0.045602079
```

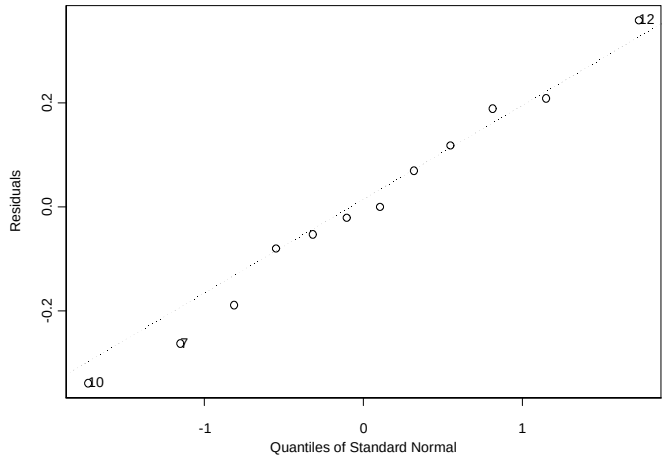
Does the straight line model fit the data? We can't tell if the model is appropriate based on the above output.

The most useful tool is a plot of residuals versus fitted values



Lack of fit more obvious here!

Normal quantiles versus quantiles of residuals does not look too bad. Normality assumption is required for testing hypotheses and forming Confidence and Prediction intervals. Because of the lack of fit above, no sense worrying about the normality assumption just yet.



Because we have replicate observations, we can conduct a lack of test to see if there is significant lack of fit.

To do this we need to get the ANOVA table for a 1-way Analysis of variance in addition to the ANOVA table from the regression output. See the lab web page for HW 2 on how to fit the AOV model.

```
*** Analysis of Variance Model ***

Call:
   aov(formula = ph ~ as.factor(log.time), data = Ex0816, na.action = na.omit)

Terms:
                as.factor(log.time) Residuals
 Sum of Squares            3.916375  0.059050
Deg. of Freedom                   5         6

Residual standard error: 0.0992051746
Estimated effects are balanced

                     Df Sum of Sq      Mean Sq    F Value     Pr(F)
as.factor(log.time)  5  3.916375 0.7832750000 79.5876376 0.00002114860428
          Residuals  6  0.059050 0.0098416667
```

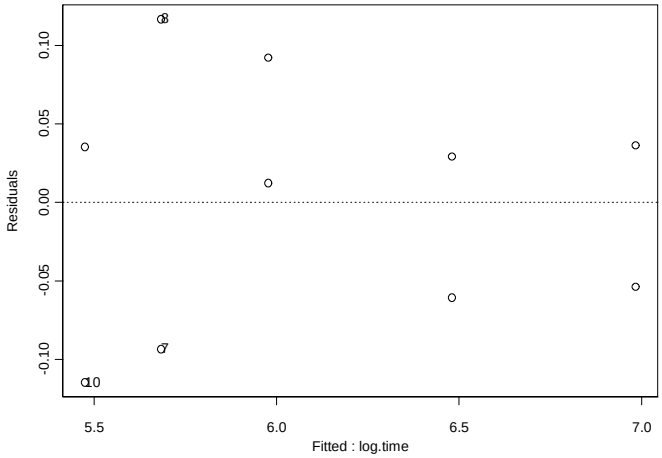
Construct the ANOVA table for the Lack of Fit test.

Source	***df***	***Sum of Squares***	***Mean Square***	***F***	***Pr(F) (p-value)***
Regression[1]					
Residual[2]:					
Lack of Fit[3]					
Pure Error[4]					

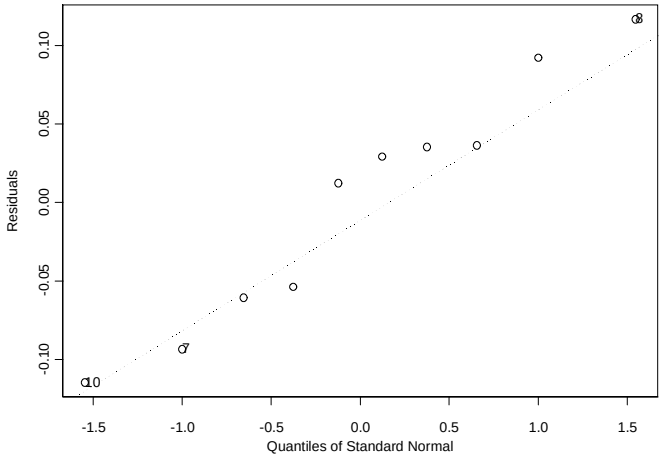
1 From the Regression ANOVA table line (listed as log.time here)
2 From the Regression ANOVA table residual line
3 Lack of Fit df and SS are obtained by subtracting Pure Error from the Residual line
4 Pure error df and SS comes from the Residual line of the 1-way AOV model (called "Within Groups" in the text.

Conclusion?
Go back to model without the last 24 hour data and check fit.

Without the last 24 hour data: residual vs fitted values look ok



Normal quantile-quantile plots: OK



Summaries without the 24 hour data

```
(Note to fit the model using a subset of observations 1:10 set subset to 1:10 in
the regression dialog)

       *** Linear Model ***

Analysis of Variance Table

Terms added sequentially (first to last)
          Df   Sum of Sq     Mean Sq     F Value            Pr(F)
 log.time  1 3.006466745 3.006466745 444.3060709 2.695158219e-008
Residuals  8 0.054133255 0.006766657



       *** One Anova Model ***

                    Df Sum of Sq Mean Sq     F Value            Pr(F)
as.factor(log.time)  4    3.0160 0.75400 84.52914798 0.00008878601939
          Residuals  5    0.0446 0.00892
```

Construct the ANOVA table for the Lack of Fit test for subset.

Source	***df***	***Sum of Squares***	***Mean Square***	***F***	***Pr(F) (p-value)***
Regression[5]					
Residual[6]:					
Lack of Fit[7]					
Pure Error[8]					

Conclusions?

5 From the Regression ANOVA table line (listed as log.time here)
6 From the Regression ANOVA table residual line
7 Lack of Fit df and SS are obtained by subtracting Pure Error from the Residual line
8 Pure error df and SS comes from the Residual line of the 1-way AOV model (called "Within Groups" in the text.