Statistical Modeling I
Practical in R – Output

Practical in R

This practical reminds you how to load the data and how to run a linear regression in R. Moreover, it gives you the opportunity to learn how to work with standardized residuals and fitted values.

To calibrate an instrument which measures the lactic acid concentration in blood, an investigator used twenty samples of known concentration (x) and then obtained the concentration as measured on the instrument (y). The results were

x	1	1	1	1	3	3	3	3
y	1.1	0.7	1.8	0.4	3.0	2.4	4.9	4.4
x	5	5	5	5	10	10	10	10
y	4.5	7.3	8.2	6.2	12.0	13.1	12.6	13.2
x	15	15	15	15				
y	18.7	19.7	17.4	17.1				

Enter the x and y values as vectors in R.

1. Produce a scatterplot of the data. Does the relationship between y and x seem to be linear?

```
> x<- c(1,1,1,1,3,3,3,3,5,5,5,5,10,10,10,10,15,15,15,15)
> y<- c(1.1,0.7,1.8,0.4,3.0,2.4,4.9,4.4,4.5,7.3,8.2,6.2,12,
13.1,12.6,13.2,18.7,19.7,17.4,17.1)
> plot(x,y, main="Plot of Y versus X")
```

See Figure 1.1 and we have a linear relation between the x and the y.

2. Fit the model to the data by `conc <- lm(y ~ x)`
we use `conc` as the name of our model as the data is about concentration of lactic acid.

To see the details of our fitted model we use
`summary(conc)`

```
> conc<-lm(y~x)
> summary(conc)

Call:
lm(formula = y ~ x)
```

Plot of Y versus X
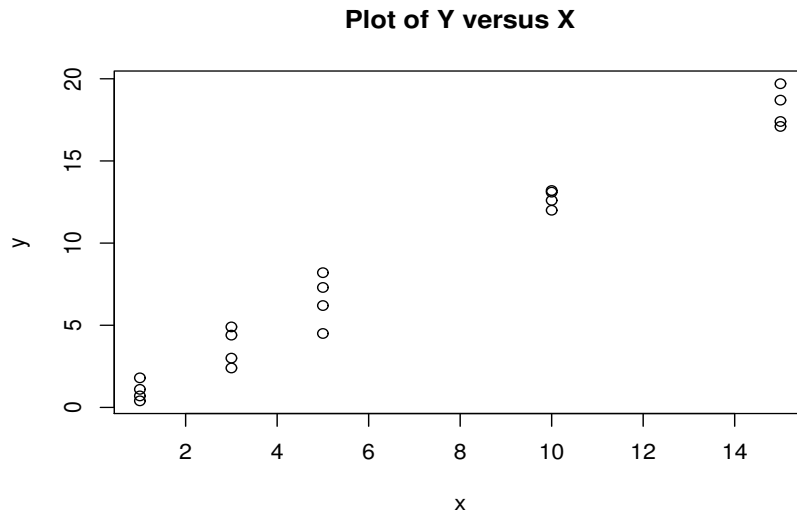

Figure 1.1: Plot of the data.

```
Residuals:
     Min        1Q   Median        3Q       Max
-1.71840  -0.78979   0.07438   0.74938   1.98160


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.06118    0.39157   0.156    0.878
x            1.23144    0.04615  26.686 6.31e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.047 on 18 degrees of freedom
Multiple R-squared:  0.9753, Adjusted R-squared:  0.974
F-statistic: 712.1 on 1 and 18 DF,  p-value: 6.31e-16
```

The estimated values of the intercept and of the slope parameter are 0.06 and 1.23 with a positive relation between the x and the y

3. Add the fitted line to your scatterplot by the following command
   ```
   plot(x,y, main="Fitted Line Plot")
   (abline(conc))
   ```

4. See the Analysis of Variance table by using the following command
   ```
   anova(conc)
   ```

   The result is

   ```
   > anova(conc)
   Analysis of Variance Table
   ```

2

Fitted Line Plot
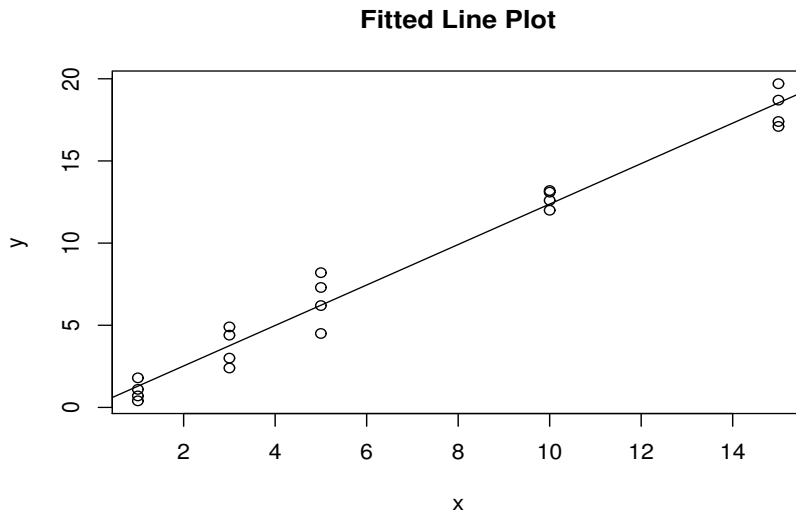

Figure 1.2: Plot of the data with the regression line.

```
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 781.28  781.28  712.12  6.31e-16 ***
Residuals 18  19.75    1.10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. We need to store the standardised residuals for plots
   ```
   stdres <- rstandard(conc)
   print(stdres)
   plot(x,stdres, main="Std residuals versus x")
   ```
   and save this plot.

   Is there any reason to doubt the linearity of the model?

   We run the previous command in R

   ```
   > stdres <- rstandard(conc)
   > print(stdres)
            1          2          3          4          5          6          7          8
   -0.19551730 -0.60152502  0.51499620 -0.90603081 -0.75120118 -1.34777736  1.13795670  0.64080989
            9         10         11         12         13         14         15         16
   -1.68880071  1.06296854  1.94746580 -0.01808367 -0.37183739  0.71708202  0.22211865  0.81607470
           17         18         19         20
    0.17629115  1.23092655 -1.19473487 -1.51112549
   > plot(x,stdres, main="Std residuals versus x")
   ```

   There is no reason to doubt the linearity of model has shown in the plot of the residuals

6. We also save the fitted values
   ```
   fits <- fitted(conc)
   print(fits)
   ```
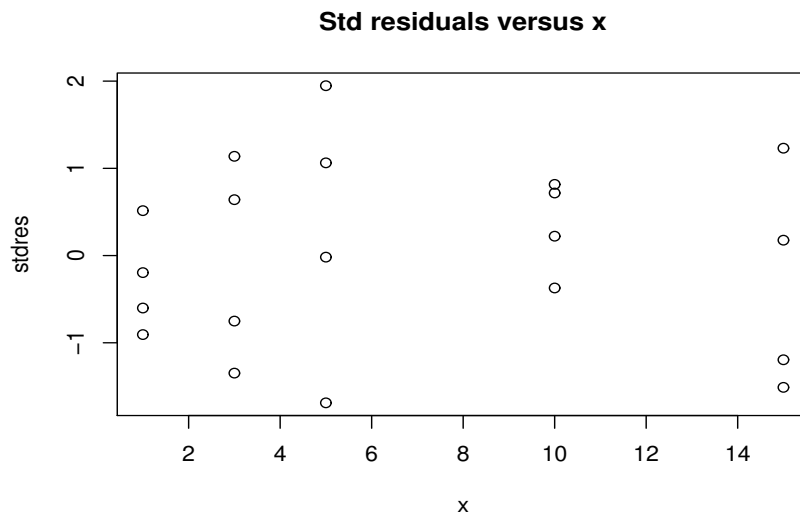
3

Std residuals versus x

Figure 1.3: Plot of the standardized residuals.

```
plot(fits,stdres, main="Std residuals versus fits")
```
save this plot.

Is there any reason to doubt that the variance is constant?

We run the following code in R

```
> fits<- fitted(conc)
> print(fits)
        1        2        3        4        5        6        7        8        9       10
 1.292624 1.292624 1.292624 1.292624 3.755512 3.755512 3.755512 3.755512 6.218401 6.218401
       11       12       13       14       15       16       17       18       19       20
 6.218401 6.218401 12.375621 12.375621 12.375621 12.375621 18.532842 18.532842 18.532842 18.532842
> plot(fits,stdres, main="Std residuals versus fits")
```

There is no reason to doubt the constant variance since, although the spread at different values of x is slightly different there is no pattern to it.

7. Finally we look at a Q-Q plot
   ```
   qqnorm(stdres, main="Q-Q Plot")
   qqline(stdres)
   ```
   save this plot.

   Is there any reason to doubt the assumption that the errors are normally distributed?

   ```
   > qqnorm(stdres, main="Q-Q Plot")
   > qqline(stdres)
   ```

   The Q-Q plot does cause some doubt on the normality but there are not really enough points off the line to for this to be serious
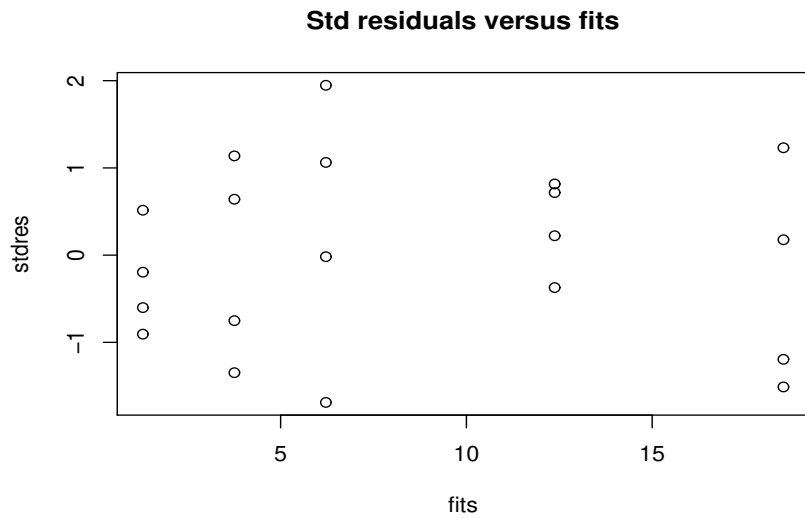
4

Std residuals versus fits
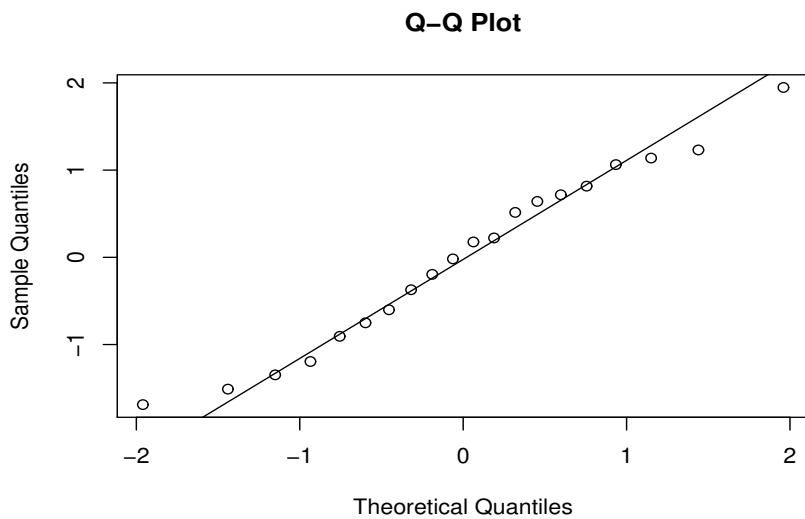

Figure 1.4: Plot of the standardized residuals versus the fitted values.

Q−Q Plot

Figure 1.5: Plot of the Q-Q Norm.