Experimental Design and Data Analysis:
Assignment 4

Andrew Bedard(2566978) & Simone van Gompel(2567525)
Group 19

May 15, 2015

Exercise 1

1

2
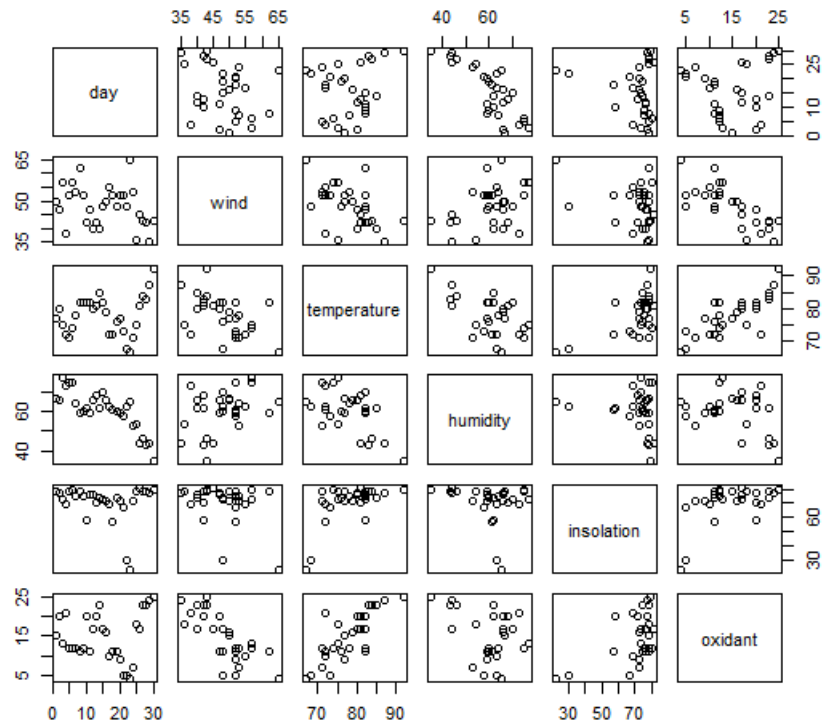
3

4

Exercise 2

1

Figure 1: Pairplot of the airpollution data
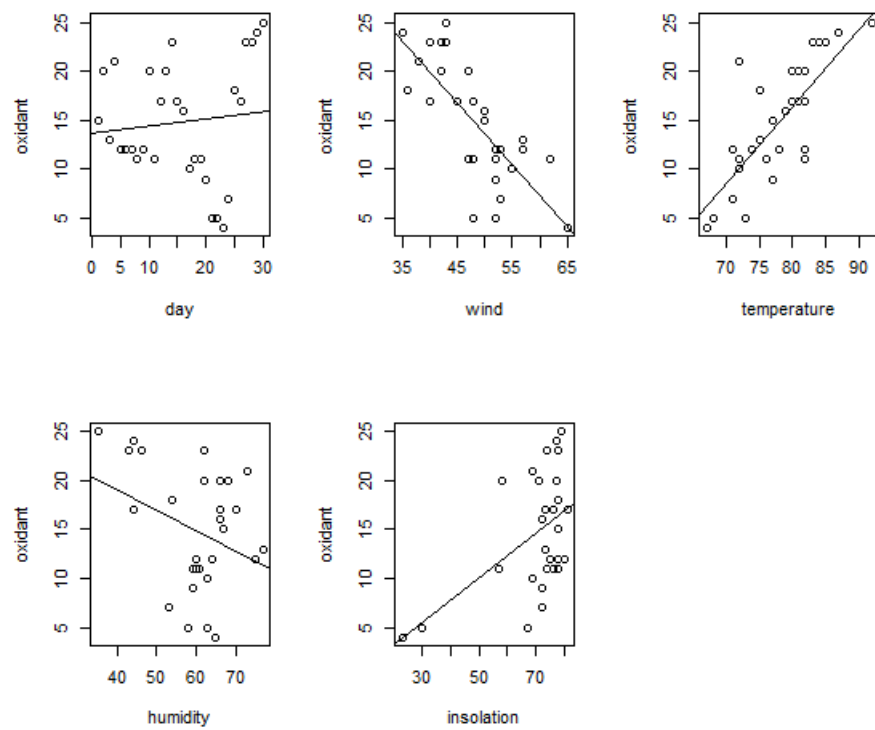
Figure 2: The linear regression of the explanatory variables

3

4

5

Exercise 3

1

2

3

4

5

Exercise 4

Following the step-up method we get for the first variable:

Table 1: Results of 1-way Anova on square root of *genal.txt* data

Variable	R^2
ExpendPop	0.9073
ExpendEmploy	0.954
ExpendLawyers	0.9373
ExpendCrime	0.1119
ExpendBad	0.6964

ExpendEmploy had the highest score, so we take this for the second step:

Table 2: Results of 1-way Anova on square root of *genal.txt* data

Variable	R^2
ExpendEmploy+Pop	0.9543
ExpendEmploy+Lawyers	0.9632
ExpendEmploy+Crime	0.9551
ExpendEmploy+Bad	0.9551

ExpendEmploy+Lawyer had the highest score, so we take this for the third step:

Table 3: Results of 1-way Anova on square root of *genal.txt* data

Variable	R^2
ExpendEmploy+Lawyers+Pop	0.9637
ExpendEmploy+Lawyers+Crime	0.9632
ExpendEmploy+Lawyers+Bad	0.9639

Adding these variables yield no significant change and so we stop at the second step. The result of ExpendEmploy+Lawyer is:

```
Coefficients:
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  -1.107e+02   4.257e+01   -2.600   0.01236 *
employ        2.971e-02   5.114e-03    5.810  4.89e-07 ***
lawyers       2.686e-02   7.757e-03    3.463   0.00113 **
Multiple R-squared:   0.9632
```

1 R-Code

1.1 Exercise 1

1.2 Exercise 2

1.3 Exercise 3

1.4 Exercise 4