

Regression: Exercises for Learners 25. Apr. 2015/MLR, R, Regression, Tutorial/No Comments A collection of Exercises in Regression Analysis that a Learner must know How to

In the table below, we have results from an environmental engineering study of a certain chemical reaction.

There are 21 concentrations of separately prepared solutions recorded at different times (three measurements at each of seven times).

In the last column we have natural logarithms of the concentrations for your reference.

The data can also be downloaded in CSV from [This Link](#)

You will use the data from this table to complete certain exercises mentioned in the last section after data table.

MAKE TABLE Solution Number (i) Time (Xi) (hrs) Concentration (Yi) (mg/ml) Ln of Concentration (lnYi) 1

6
0.029
-3.540
2
6
0.032
-3.442
3
6
0.027
-3.612
4
8
0.079
-2.538
5
8
0.072
-2.631
6
8
0.088
-2.430
7
10
0.181
-1.709
8
10
0.165
-1.802
9
10
0.201

-1.604
10
12
0.425
-0.856
11
12
0.384
-0.957
12
12
0.472
-0.751
13
14
1.130
0.122
14
14
1.020
0.020
15
14
1.249
0.222
16
16
2.812
1.034
17
16
2.465
0.902
18
16
3.099
1.131
19
18
3.614
1.285
20
18
3.402
1.224
21

18

3.913

1.364

Use R to complete the following exercises:

Exercise One: Graphs

Generate separate graphs of:

Concentration (Y) vs. Time (X) Natural Logarithm of Concentration ($\ln Y$) vs. Time (X)

Exercise Two : Equations and Plotting

Using the output from exercise one, obtain the following:

The estimated equation of the straight-line (degree 1) regression of Y on X The estimated equation of the quadratic (degree 2) regression of Y on X The estimated equation of the straight-line (degree 1) regression of $\ln Y$ on X Plots of each of these fitted equations on their respective scatter diagrams.

Exercise Three : Determine and Compare

Determine and compare the proportions of the total variation in Y explained by the straight-line regression on X and by the quadratic regression on X

Exercise Four: F-Tests

Carry out F-tests for the significance of the straight-line regression of Y on X Carry out an overall F-test for the significance of the quadratic regression of Y on X and a test for the significance of the addition of x^2 to the model For the straight-line regression of $\ln Y$ on X, carry out F-tests for the significance of the overall regression

Exercise Five:Determine and Compare

What proportion of the variation in $\ln Y$ is explained by the straight-line regression of $\ln Y$ on X? Compare this result with that obtained in Exercise Three for the quadratic regression of $\ln Y$ on X.

Exercise Six: Examine and Discuss

Use comment box on this page (below) and explain your thoughts on the following:

A fundamental assumption in regression analysis is variance homoscedasticity. By examining the scatter diagrams constructed in Exercises One and Two, state why taking natural logarithms of the concentrations helps with regard to the assumption of variance homogeneity. Do you think the straight-line regression of $\ln Y$ on X is better for describing this set of data than the quadratic regression of Y on X? Considering the overall table, what key assumption about the data would be in question if, instead of 21 different solutions, there were only 3 different solutions, each of which was analyzed at the seven different time points?