UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS

Generalised Regression Models

GRM: Coursework Assignment

Your answer to the following coursework assignment forms part of the assessment for this course (5% of the total mark for GRM).

In the petroleum refining industry, scheduling and planning play a very important role. One phase of the process is the production of petroleum spirit from crude oil, and its subsequent processing. To enable the plant to be scheduled in an optimum manner, an estimate is required of the percentage yield of petroleum spirit from crude oil, based upon certain rough laboratory determinations of the properties of the crude oil. The following table (see next page) shows values of actual per cent yields of petroleum spirit (y) and four properties of crude oil for samples from $n = 32$ different crudes. The properties recorded are as follows:

x_1: specific gravity of the crude, a function of the API measurement.

x_2: crude oil vapour pressure, measured in pounds per square inch absolute.

x_3: the ASTM 10 per cent distillation point, in °F.

x_4: the petroleum fraction end point, in °F.

It is required to use this data set to provide an equation for predicting y, from the measurements of these four explanatory variables.

Apply suitable models in R to analyse the relationship between y and the explanatory variables x_1, x_2, x_3, x_4.

Table: Observations on the variables x_1, x_2, x_3, x_4, y.

x_1	x_2	x_3	x_4	y
38.4	6.1	220	235	6.9
40.3	4.8	231	307	14.4
40.0	6.1	217	212	7.4
31.8	0.2	316	365	8.5
40.8	3.5	210	218	8.0
41.3	1.8	267	235	2.8
38.1	1.2	274	285	5.0
50.8	8.6	190	205	12.2
32.2	5.2	236	267	10.0
38.4	6.1	220	300	15.2
40.3	4.8	231	367	26.8
32.2	2.4	284	351	14.0
31.8	0.2	316	379	14.7
41.3	1.8	267	275	6.4
38.1	1.2	274	365	17.6
50.8	8.6	190	275	22.3
32.2	5.2	236	360	24.8
38.4	6.1	220	365	26.0
40.3	4.8	231	395	34.9
40.0	6.1	217	272	18.2
32.2	2.4	284	424	23.2
31.8	0.2	316	428	18.0
40.8	3.5	210	273	13.1
41.3	1.8	267	358	16.1
38.1	1.2	274	444	32.1
50.8	8.6	190	345	34.7
32.2	5.2	236	402	31.7
38.4	6.1	220	410	33.6
40.0	6.1	217	340	30.4
40.8	3.5	210	347	26.6
41.3	1.8	267	416	27.8
50.8	8.6	190	407	45.7

The data values are available on Learn.