University of California, Los Angeles
Department of Statistics

Statistics 13 **Instructor: Nicolas Christou**

Simple regression analysis

Part A:
Use `R` to access the Maas river data. These data contain the concentration of lead and zinc in `ppm` at 155 locations at the banks of the Maas river in the Netherlands. You can read the data in `R` as follows:

```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics13/soil.txt", header=TRUE)
```

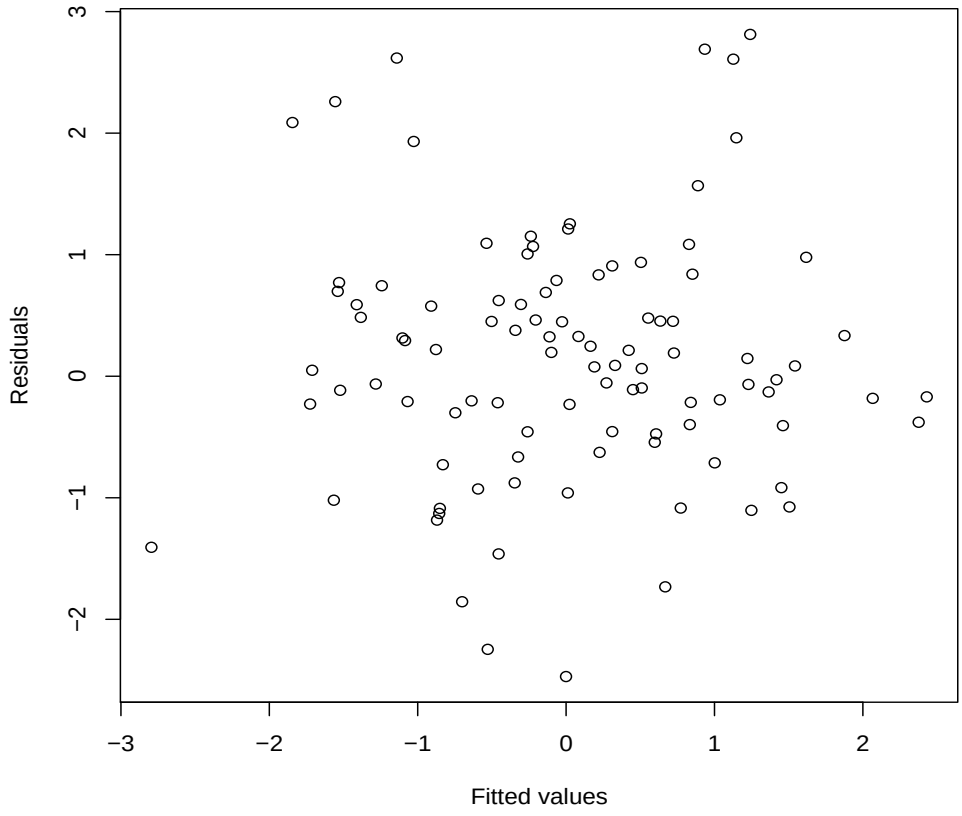
a. Run the regression of `lead` on `zinc`. Submit the `R` printout. What is the value of R^2.

b. Run the regression of `log(lead)` on `log(zinc)`. Submit the `R` printout. What is the value of R^2.

c. Construct a confidence interval for the slope of the model in question (b). What is your conclusion.

d. Test the hypothesis that the slope of the model in question (b) is equal to zero against the alternative that it is not equal to zero. Use the p-value from the `R` output.

e. Plot the residuals (`y`-axis) against the fitted values (`x`-axis) of the model of part (b). Are there any violations of the assumptions of regression? In general this plot must look like a patternless cloud of points (see an example below).



Part B:
The allocation of aid from New York State to individual school districts on Long Island is a very politically charged issue, since school taxes are the largest portion of local property taxes. For this reason, it is useful to understand the process that leads to differing levels of state aid. The following data give the proportion of students that were white, the state aid per student in 1991, and the percentage change in state aid from 1991 to 1994, for a set of school districts in Nassau County (source: various articles in *Newsday*). Is it possible to model the change in state aid from the proportion of white students? The data can be accessed in `R` as follows:

```
b <- read.table("http://www.stat.ucla.edu/~nchristo/statistics13/district.txt", header=TRUE)
```

District	Prop. of white	Aid p/s	Change in aid
Baldwin	0.7619150	2743.849	-20.87098
Bellmore	0.9600000	2714.702	-33.18658
Bellmore-Merick	0.9498513	3190.691	-20.94512
Bethpage	0.9349593	1984.253	-26.71606
Carle Place	0.8846421	1343.953	-25.3559
East Meadow	0.8879936	2274.382	-8.23419
East Rockaway	0.9398762	2832.949	-31.16606
East Williston	0.8946986	1288.161	-18.5612
Farmingdale	0.8609259	2676.262	-25.28701
Franklin Square	0.8845910	1520.793	-3.92579
Garden City	0.9348337	1208.642	-33.61306
Clen Cove	0.5908470	1728.227	-25.03693
Great Neck	0.8118150	1420.194	-34.08252
Hempstead	0.0038760	3584.258	34.54547
Herricks	0.7039627	1664.211	-17.22987
Hewlett-Woodmere	0.8997503	1482.684	-24.80647
Hicksville	0.7959093	2087.983	-31.84878
Island Park	0.8717949	1830.392	-18.33777
Island Trees	0.9215872	3902.785	-13.72291
Jericho	0.8789966	1631.453	-36.36158
Lawrence	0.7317935	2001.345	-27.12733
Levittown	0.9398691	3377.329	-11.03951
Locust Valley	0.9118414	1362.006	-22.56311
Long Beach	0.6329236	2109.607	-24.74878
Lynbrook	0.9149781	1407.610	-22.04312
Manhasset	0.7958278	1397.663	-27.38187
Massapequa	0.9758910	2751.454	-25.16039
Merrick	0.9674503	2148.133	-22.34513
Mineola	0.8239940	1602.114	-27.42883
New Hyde Park	0.7995320	931.8000	-23.83849
North Bellmore	0.9268664	2250.673	-21.11285
North Merric	0.9168854	2024.551	-2.31782
North Shore	0.9345088	1592.318	-23.70098
Oceanside	0.9279847	1863.777	-14.86319
Oyster Bay-East Norwich	0.8515929	1760.950	-35.53844
Plainedge	0.9549645	3060.421	-17.38123
Plainview-Old Bethpage	0.9478651	2527.303	-31.74576
Port Washington	0.7067465	1513.155	-22.61336
Rockville Centre	0.8309148	1537.153	-18.34864
Roosevelt	0.0026891	4448.562	37.80466
Roslyn	0.8107780	1535.186	-37.74107
Seaford	0.9609649	2991.140	-23.81355
Sewanhaka	0.6798685	2105.432	-12.24227
Syosset	0.8529753	1617.970	-40.20053
Valley Stream 13	0.9116326	1760.658	-20.43847
Valley Stream 24	0.8631347	1700.696	-25.59132
Valley Stream Central	0.8519185	2248.935	-14.45095
Wantagh	0.9626341	2490.192	-19.21851
West Hempstead	0.8237179	1883.239	-16.28749

a. Run the regression of `change` on `white`. Submit the `R` printout. What is the value of R^2.

b. Test the hypothesis that the slope of the model is equal to zero against the alternative that it is not equal to zero.

c. Plot `white` (`y`-axis) against `change` (`x`-axis). What do you observe?

d. Remove the outliers and answer again questions (a) and (b). What do you conclude now?