

Project: Regression Analysis

The project is on regression analysis. The first part of your paper will concisely summarize regression techniques. In the second part, it is your choice to do a regression analysis yourself, or to look at recent research papers, and to comment on the use of regression therein.

Your project will (and should) involve material beyond that discussed in class, and you may, but need not, display formulas. You are encouraged to talk to each other about the projects, but *you must write your paper on your own. The paper is due Thursday, May 31, in quiz section, and late papers will not be accepted.* Keep in touch with Mathias during all stages of your project, and feel free to ask for help. *Make use of office hours and come to see us.*

First part

Present a concise summary of regression analysis. Questions that may guide you include the following: What is the aim of regression analysis? How is it done – what are the key formulas? Are there any graphical tools? What are the dangers and limitations?

First look at Part III of our text (Chapters 8–12). Some of the material there has not been discussed in lecture. Then check additional references on regression analysis. For example, you might look at the paper by Webster (1997), especially pages 557–559 and 562–566, or at other introductory statistics texts in the library. Most of them have chapters on regression.

Then summarize your findings. Use your own words. *Do not copy, and give proper reference to all material and sources used.*

Second part

Do one (and only one) of the following two alternatives:

Alternative 1. Analyze a suitable data set of your choice, using regression techniques. You might find the data set within your department or on the web, or

you might check the literature. Particularly useful sources are the StatLib Datasets Archive at <http://lib.stat.cmu.edu/datasets/> and the book by Hand et al. (1994). You may use a computer package to analyze your data set, or carry out the computations by hand – either option is acceptable. Describe the data, their source, and your analysis carefully. Provide illustrations, and give proper reference to all material and sources used.

Alternative 2. Look at the recent research literature in your field of study or interest. Go to the library and browse the journals. You will note very many examples for the use of regression analysis. Pick one or two of these papers, published in 1997 or thereafter. Describe the scientific problem under study and the data set used. Explain in which way the authors use regression as a tool in the investigation. Comment critically. Again, *give proper reference to all material and sources used*, and append a copy (copies) of the paper(s) you discuss.

Grading

The maximum score on the project is 20; and your paper is worth 20% of the total score (*cf.* the syllabus). *Present the paper in a neat and concise manner.* The written part (excluding illustrations) should be no longer than six to eight pages of standard size paper.

Final comments

We have deliberately chosen to give you several options for the project. You should, however, decide fairly soon which approach you are going to take. *Please see one of us during our office hours on or before Wednesday, May 9, to discuss your choice.* At all stages of your project, keep in touch with Mathias.

References

- Hand, D. J., F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski, *A Handbook of Small Data Sets*, Chapman & Hall, London, 1994.
- Webster, R., Regression and functional relations, *European Journal of Soil Science*, **48**, 557–566, 1997.