

AMSC/MATH 420, Spring 2014
First Solo Homework:
Fitting Linear Statistical Models to Data

Due Monday, February 3

A dataset consisting of the national total numbers of births in the US on each day in 1996 can be found on the course web page as a text file `births1996.txt`. Using these data:

(a) Show that there is an important day-of-the-week effect on the way these numbers of births turn out. Which days of the week regularly have the smallest numbers of births?

(b) Demonstrate graphically and/or numerically that after subtracting a quantity that depends only on the day of the week, either from the numbers of births or from their logarithms, what remains is a sequence of numbers that looks more or less like a curvilinear trend plus “noise” except for relatively few anomalous days. Here “noise” means an apparently patternless sequence of numbers which, either visually or by some other criterion, looks like a sequence of independent, identically distributed values across time.

(c) Identify and examine the anomalous days in (b). Was there anything special about these days in 1996 that might help account for anomalies?

(d) Using a linear least-squares fit, express as simply and smoothly as possible the common curvilinear trend remaining in (b) after adjusting for day-of-week effects and possibly for the “outliers” you found in (b).

(e) Discuss the function you fitted in (d) in relation to environmental factors. Is there significant seasonal variation, and why or why not?