

## HW1 Problem A for Math 420

(A). A dataset consisting of the national total numbers of births in the US on each day in 1978 can be found at

<http://www.math.umd.edu/~evs/MATH420/Births1978.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). See whether you can show, graphically or formally, that after subtracting a quantity depending only on the day of the week either from the numbers of births or from their logarithms, what remains is a series of numbers that looks more or less like a curvilinear trend plus ‘noise’ (at least with respect to temporal pattern) except for relatively few anomalous days. “Noise” here means an apparently patternless sequence of numbers which, either visually or by some other criterion, looks like a sequence of independent and identically distributed variable values with some distribution, not necessarily normal but roughly the same across time.

(c). What is special about the anomalous days in (b) ? Can you account for them in any way ? Did anything special happen in 1978 that might help account for anomalies ?

(d). How can you represent mathematically, 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)-(c) ? The operations asked for here are to approximate the remaining observations (after subtracting the day-of-week effect) by a member of a family of curves, for example by basis expansion (such as Fourier series or polynomial).

(e). Can you relate the function you fitted in (d) to any astronomical function of the days of the year ? (for example, length of daylight hours ? or maybe, average temperature in population centers ?)

(f). Now consider the original data points (numbers of births) minus the day-of-week effect and trend function you fitted. These differences, called *residuals* might be expected to show some dependence with negative correlation between consecutive days. Why might it be true ? Is it clearly true in the data ?

(g). Can you think of a way to characterize the overall size (or more ambitiously still, the probability distribution) of the residuals ?