Benford analysis of Covid-19 data

Andrew Jackson (ETH Zürich, Switzerland) and Malcoilm Sambridge (RSES, ANU, Australia)

1 Summary

In the effort to react appropriately to the spread of Covid-19 it is essential to have access to data concerning the effects of the virus; the two most commonly reported statistics are persons known to have tested positive for the virus (called "Cases") and persons who have succumber to the virus ("Deaths"). It is important that the data are timely and accurate; in the case of the latter, how can this be checked? Here we show that datasets such as the cumulative case and death data from different countries can be analysed by comparing the reported integers to the expected distribution for such rapidly-growing statistics, which are expected to have a Benford-law type distribution. Only the leading digit is of importance in such an analysis. We show that there is an expected good agreement between theory and reported values, although there are occasional strange deviations.

2 Analysis

Data for all countries analysed were obtained from `https://github.com/mhandley/COVID19/` on 9/4/20 (courtesy Prof. M. Handley, UCL, UK), who has distributed data from national authorities and occasionally from the WHO daily briefing. These data tend to be more up-to-date than the definitive Johns Hopkins data. We take as data for each country the cumulative cases or cumulative deaths as reported in the country files and neglect the date, so that the only numbers are the growing cumulative totals. We create two "All countries" datasets for cases and for deaths by concatenating the country data. These dataset consist of 2185 and 2070 integers respectively. The latter is slightly smaller than the former because on some days when there are cases, no death information is reported. The data we acquired contained no death information for China.

We begin by analysing the two "All counties" datasets for cases and deaths in Figures 1a and 1b. In all figures the dashed curve represents the expected Benford 1st digit distribution, in which there are around 30% leading digits that are "1"s and around 5% that are "9"s.

We see a clear adherence from all the data to the expected distribution. We can see this also at a country level (Figure 2). We take as examples the European countries France, Germany, Spain, UK, Switzerland, Italy. All these countries have experienced large growths in numbers. Since the datasets are smaller, we expect to see more variability around the expected distribution. Such variability would diminish as the datasets increase in size. We have examined the plots of 53 countries, and find general agreement in their behaviour, similar to in Figure 2.

There are, however, very rare outliers that signal a behaviour far from expected. An example is given in Figure 3. The likelihood of having so many data with "9"s is very low, and one can now trace this back to the original data. Subsequent days reported 9 deaths, and then two weeks later, subsequent days reported 90 and 99 deaths. Figure 4 shows the correleation coefficient for all countries' case data when compared to the theoretical curve.
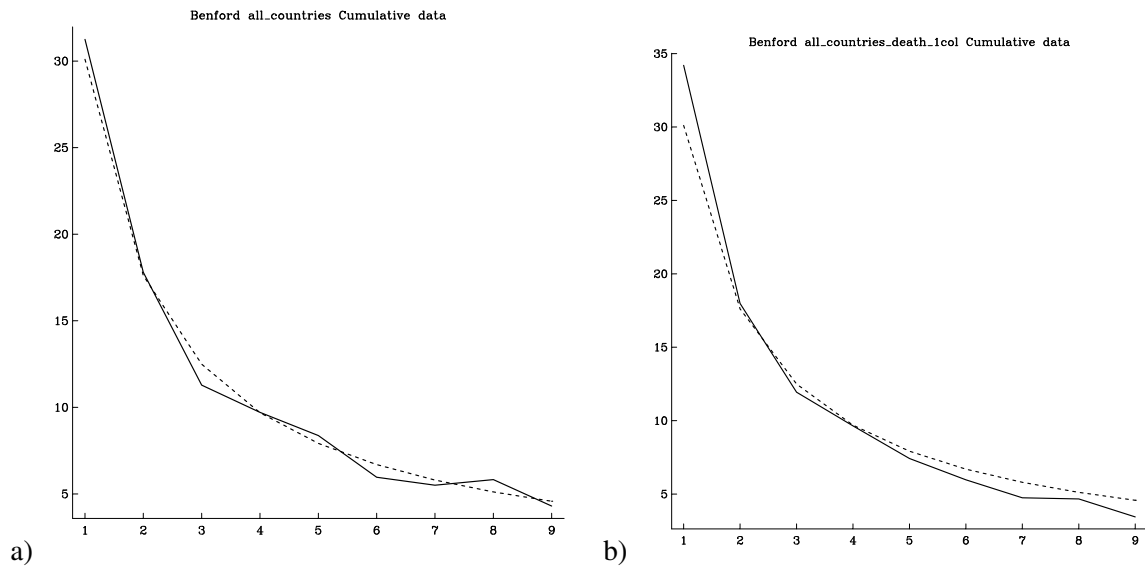
Figure 1: Comparison of data from all countries together with the expected distribution (dashed). a) Cases. b) Deaths.
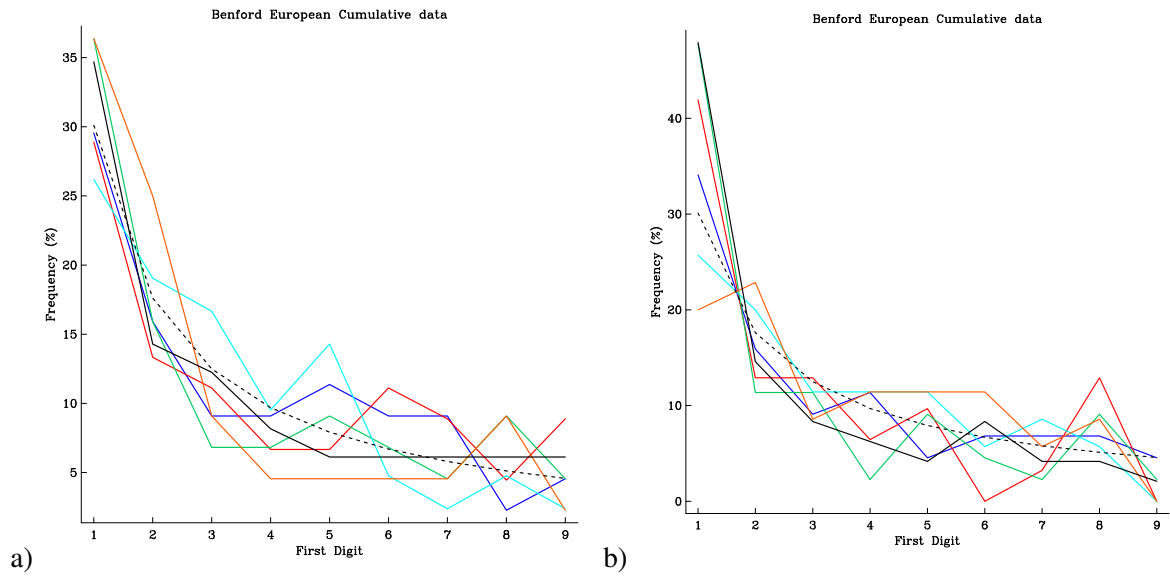

Figure 2: Comparison of data from selected European countries together with the expected distribution (dashed). Countries are France (blue), Germany (red), Spain (green), UK (cyan), Switzerland (Orange) and Italy (black). a) Cases. b) Deaths.
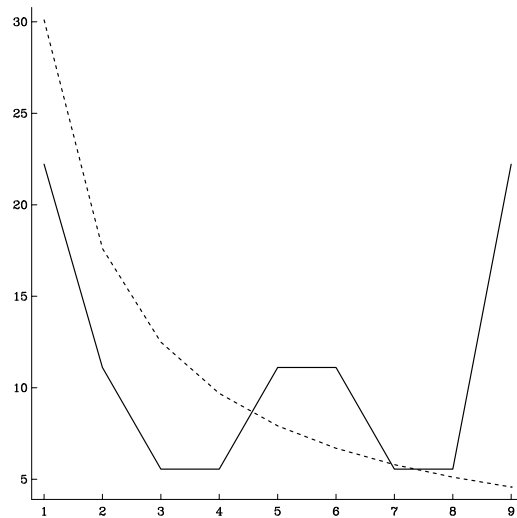

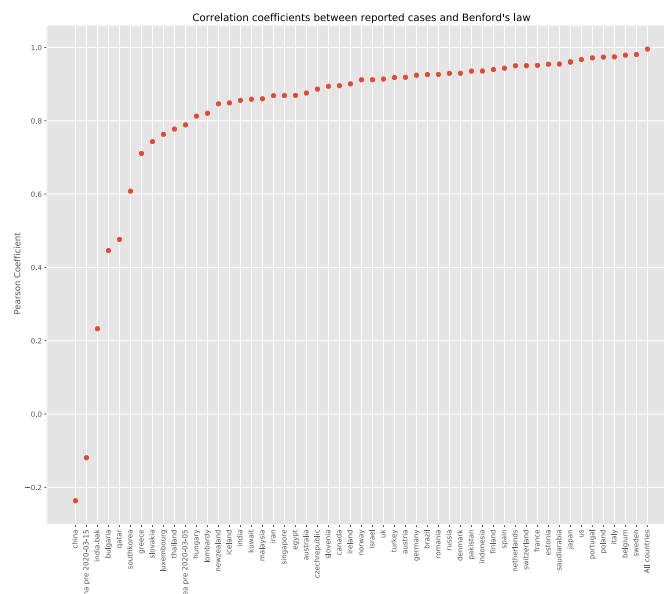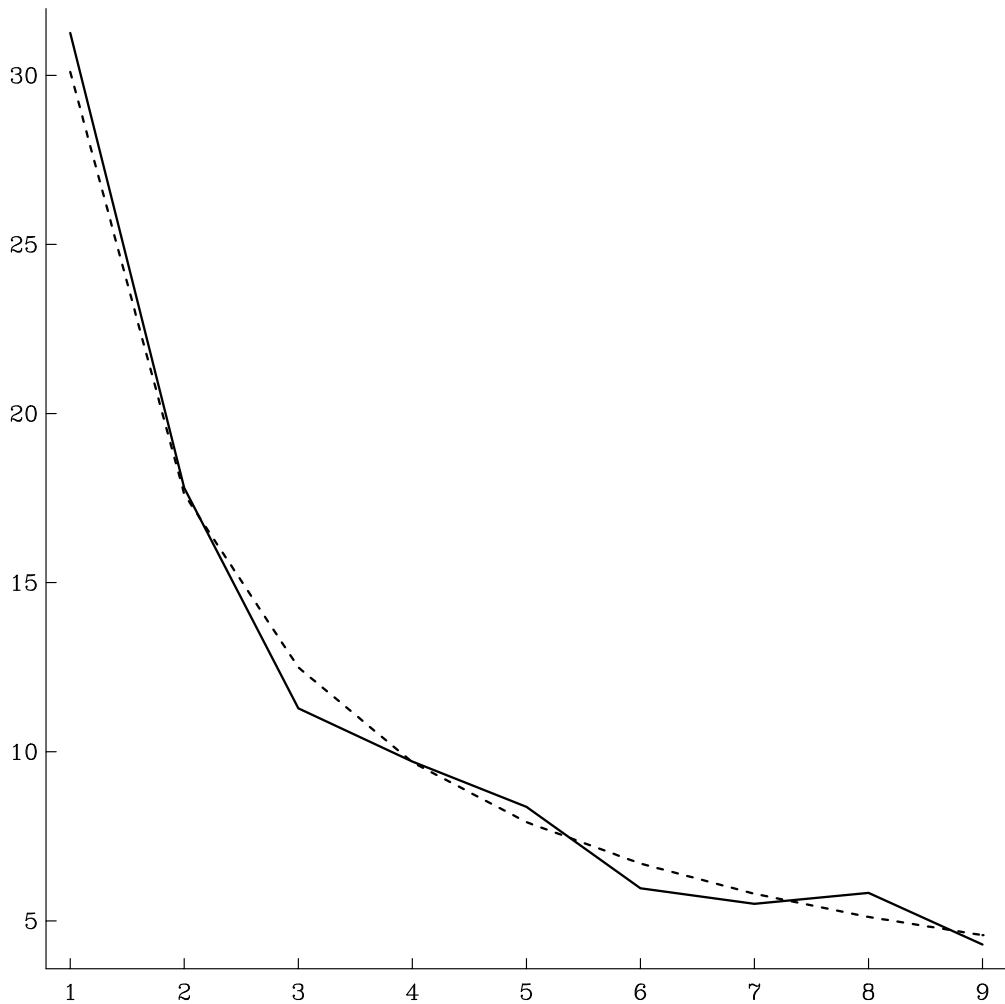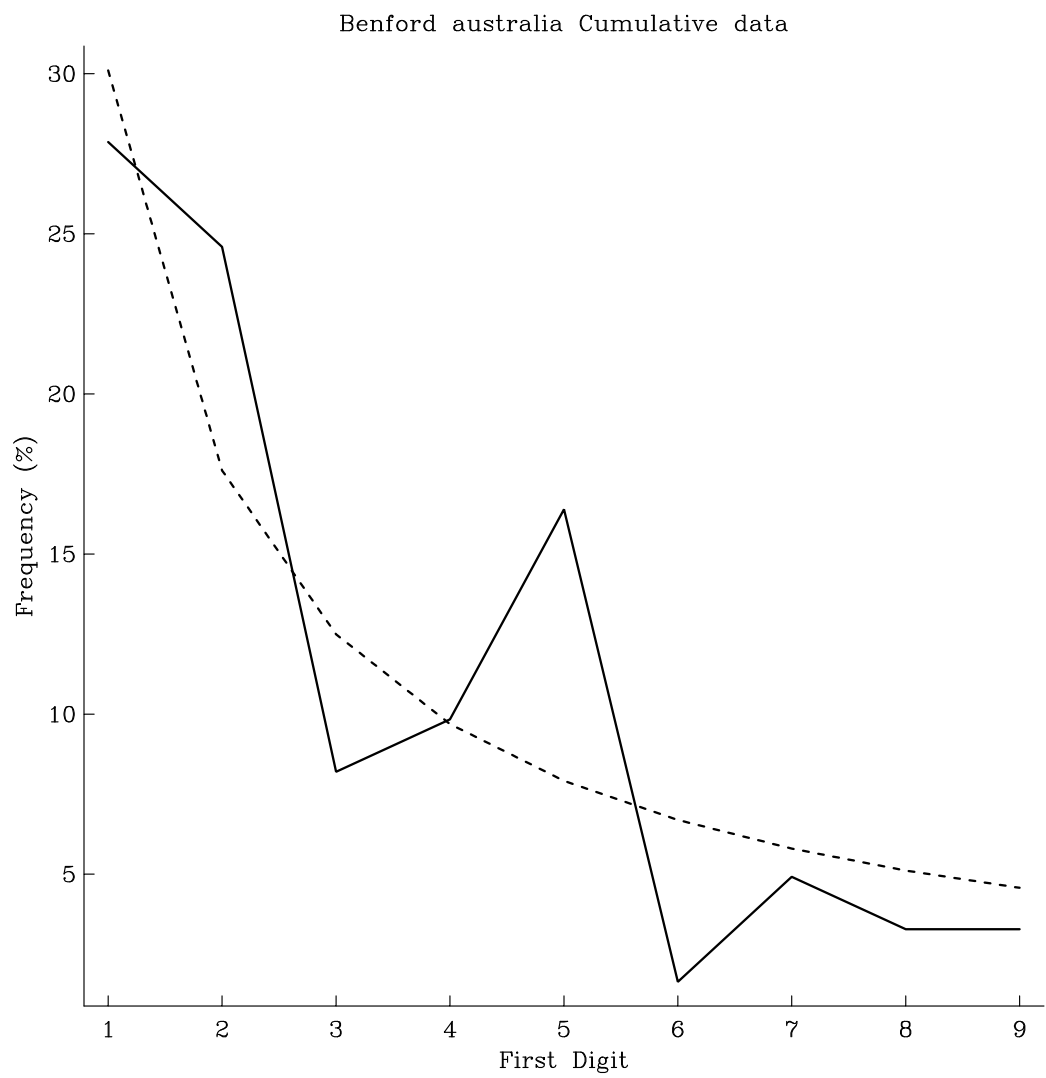Figure 3: Death data reported by a single country (name withheld).

Figure 4: Correleation coefficient for each of the countries' case data that was analysed.

Cases by country

Benford all_countries Cumulative data

Benford australia Cumulative data

Benford austria Cumulative data

Benford belgium Cumulative data
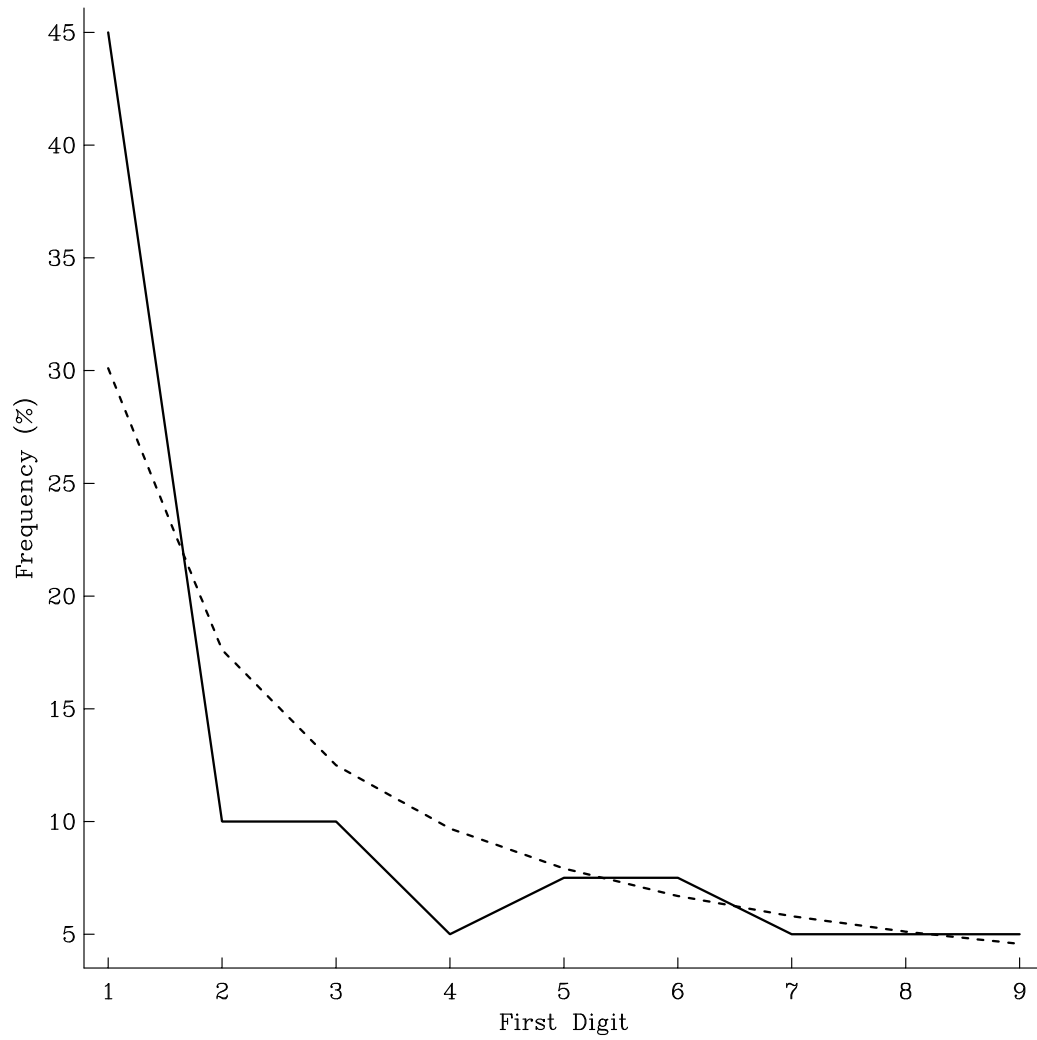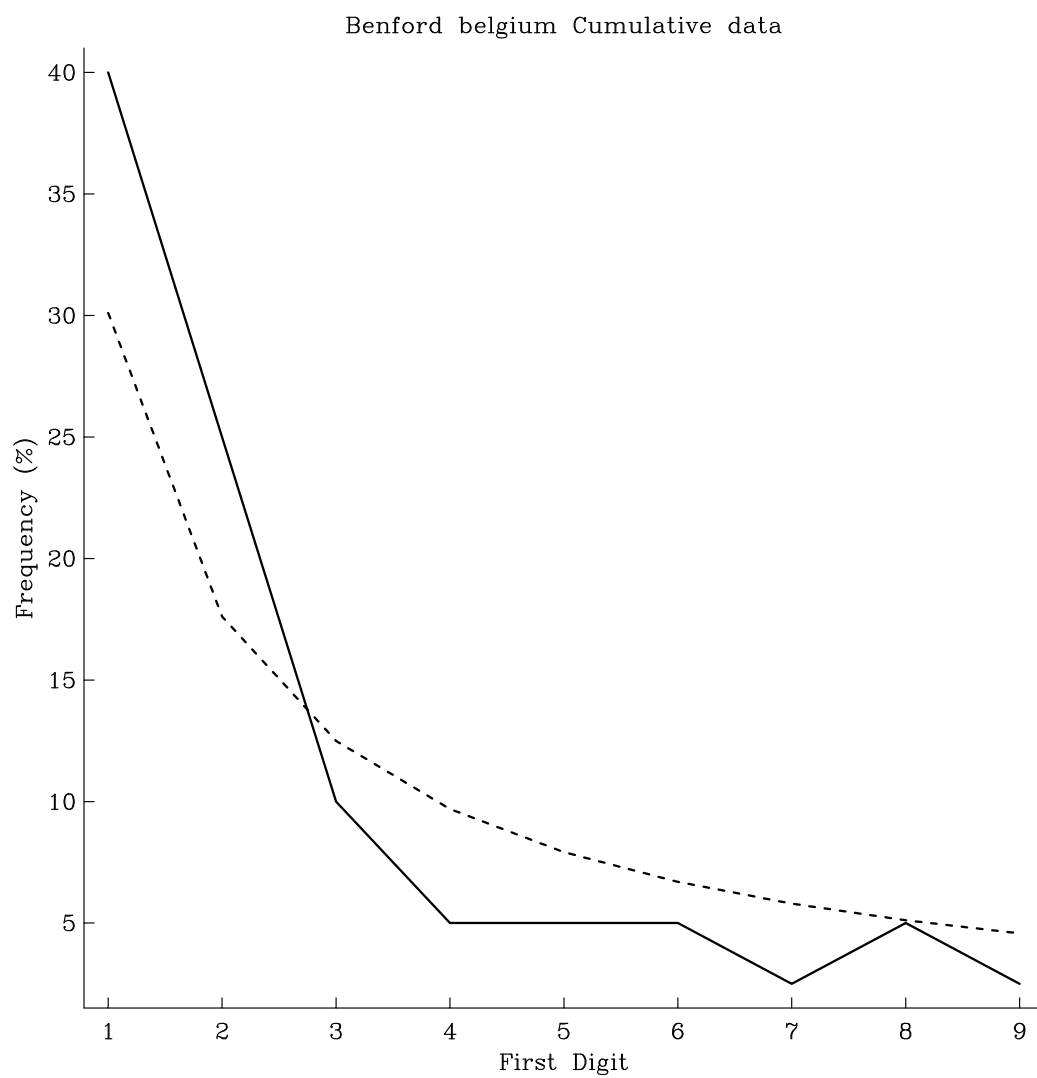
Benford brazil Cumulative data

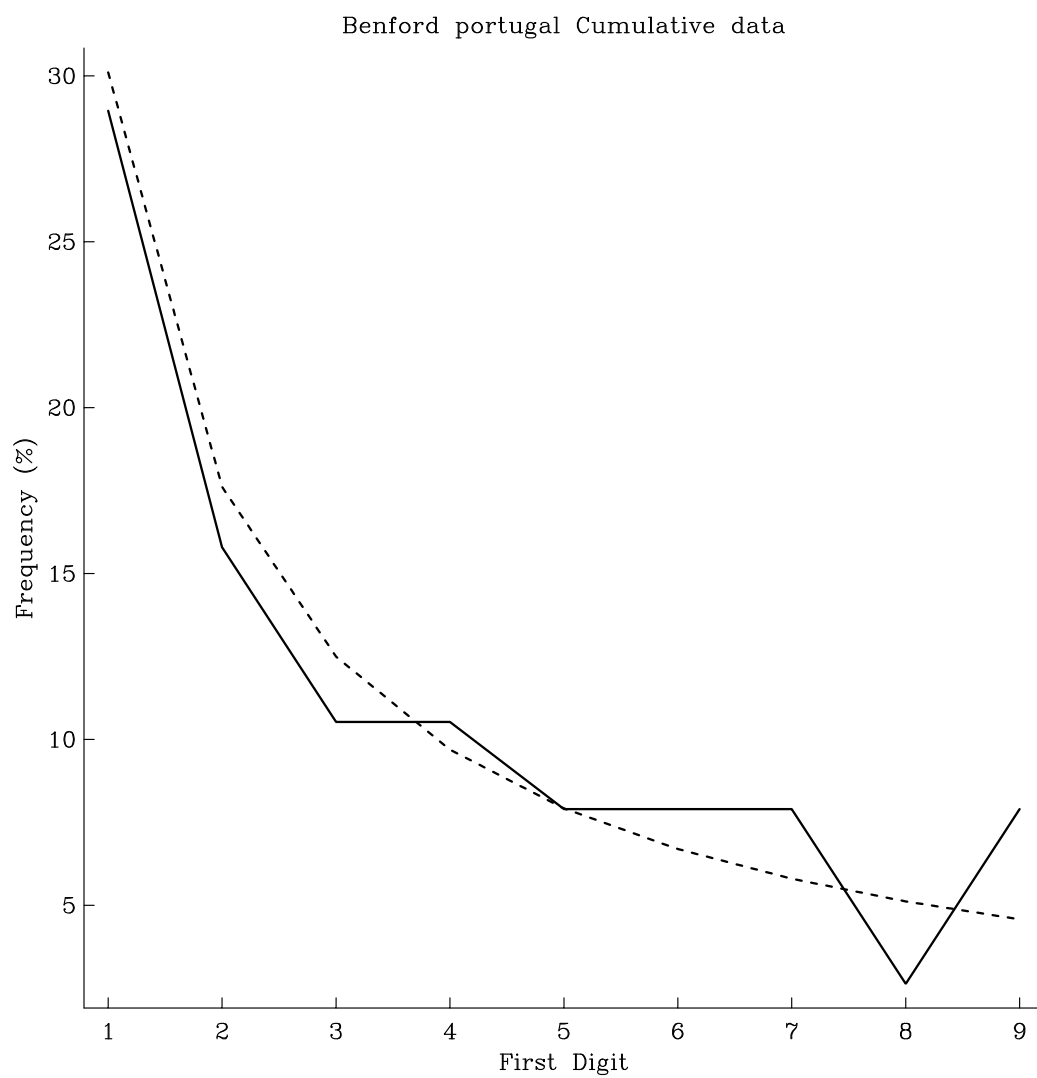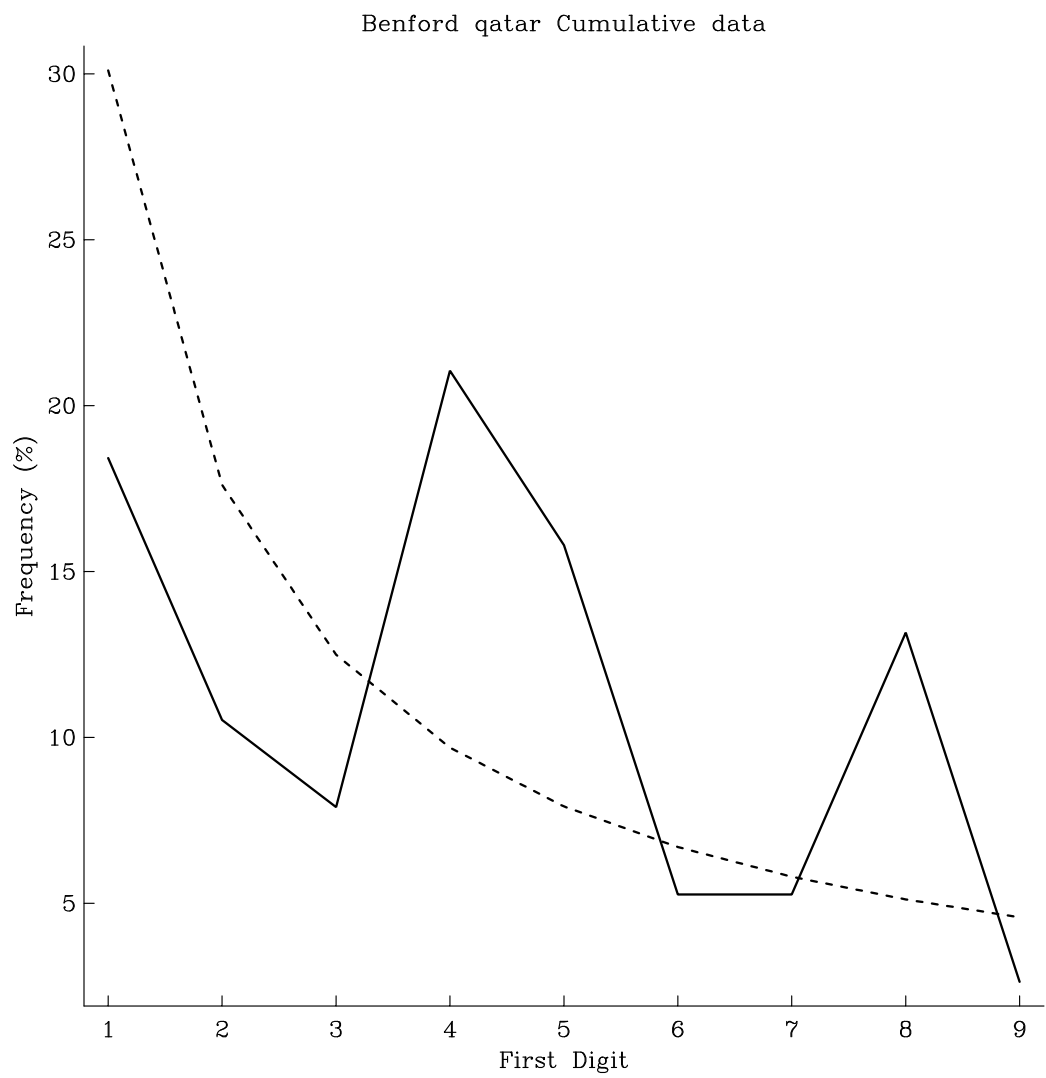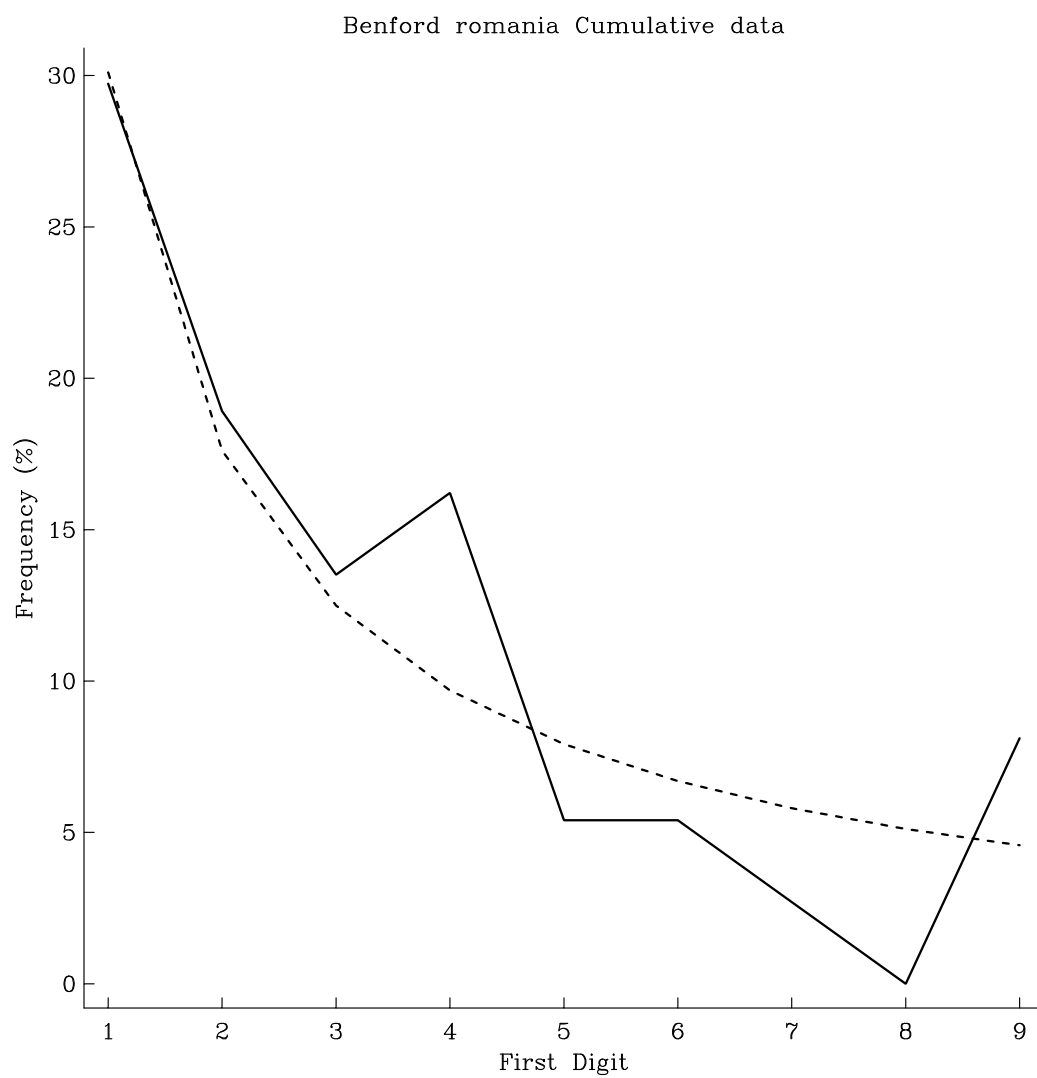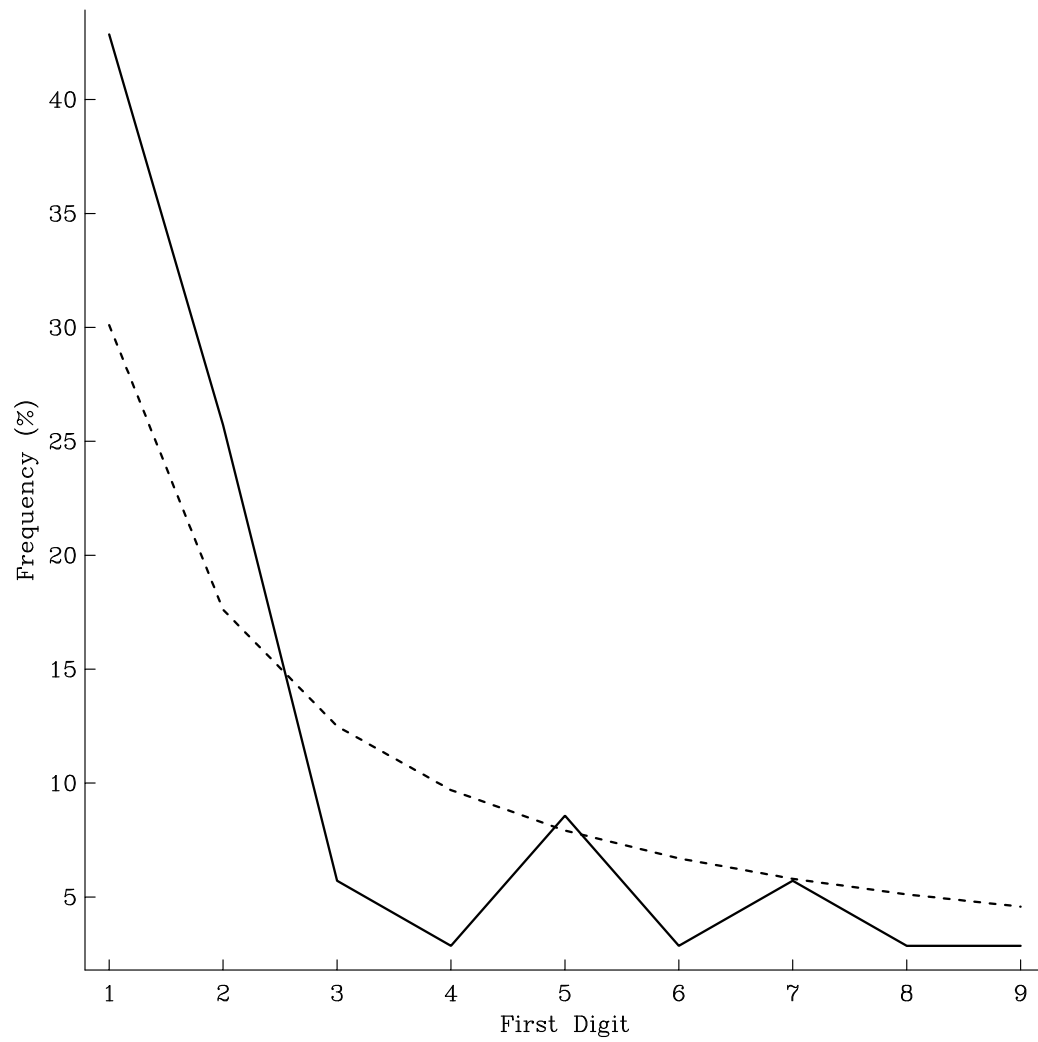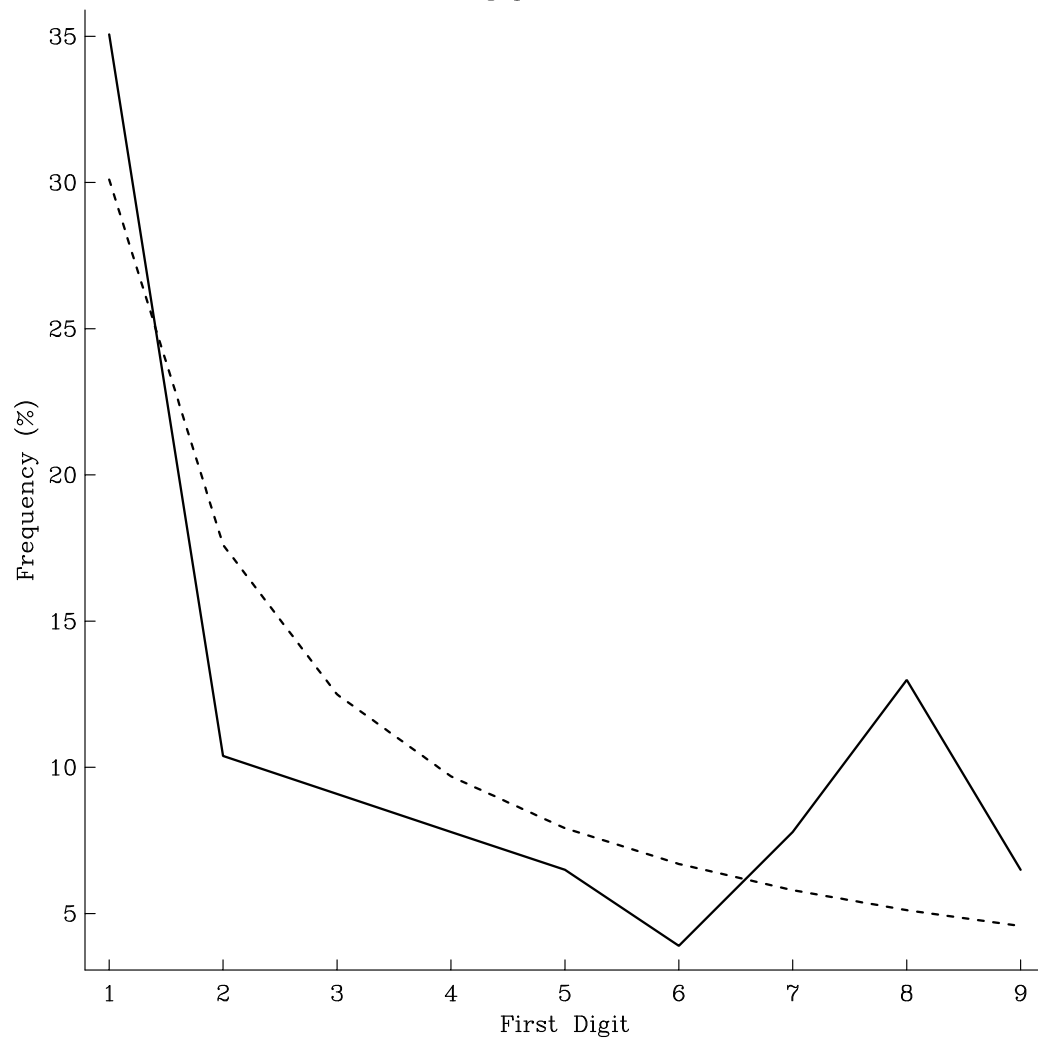Benford bulgaria Cumulative data
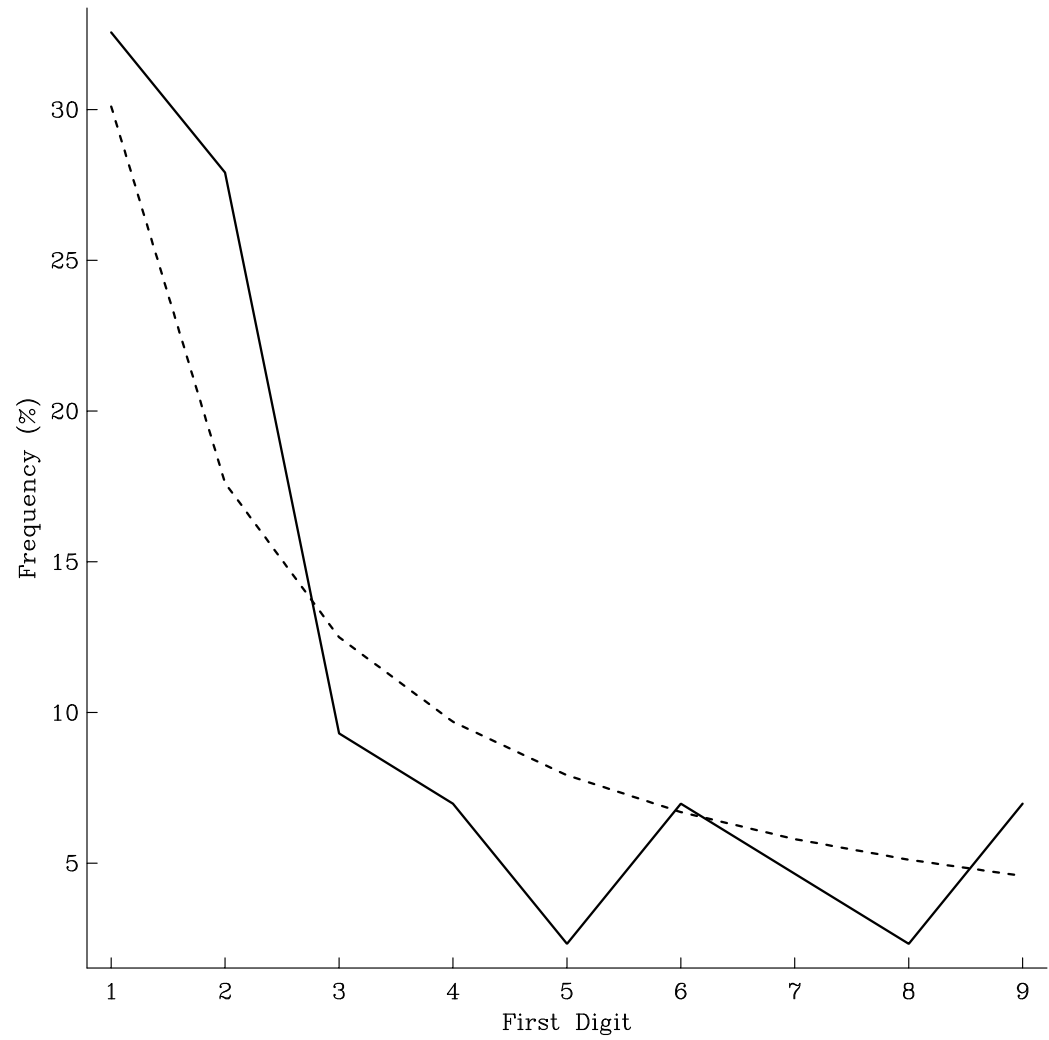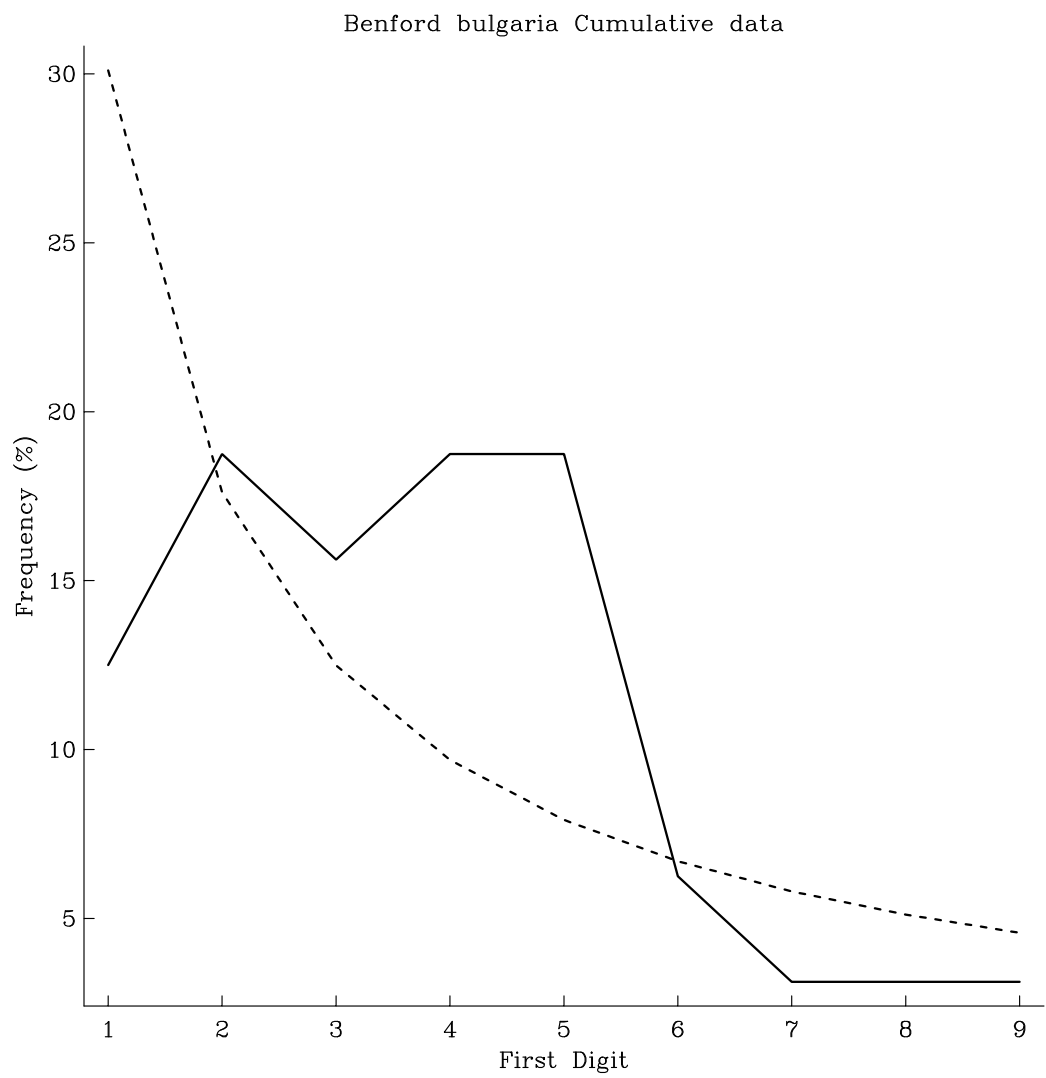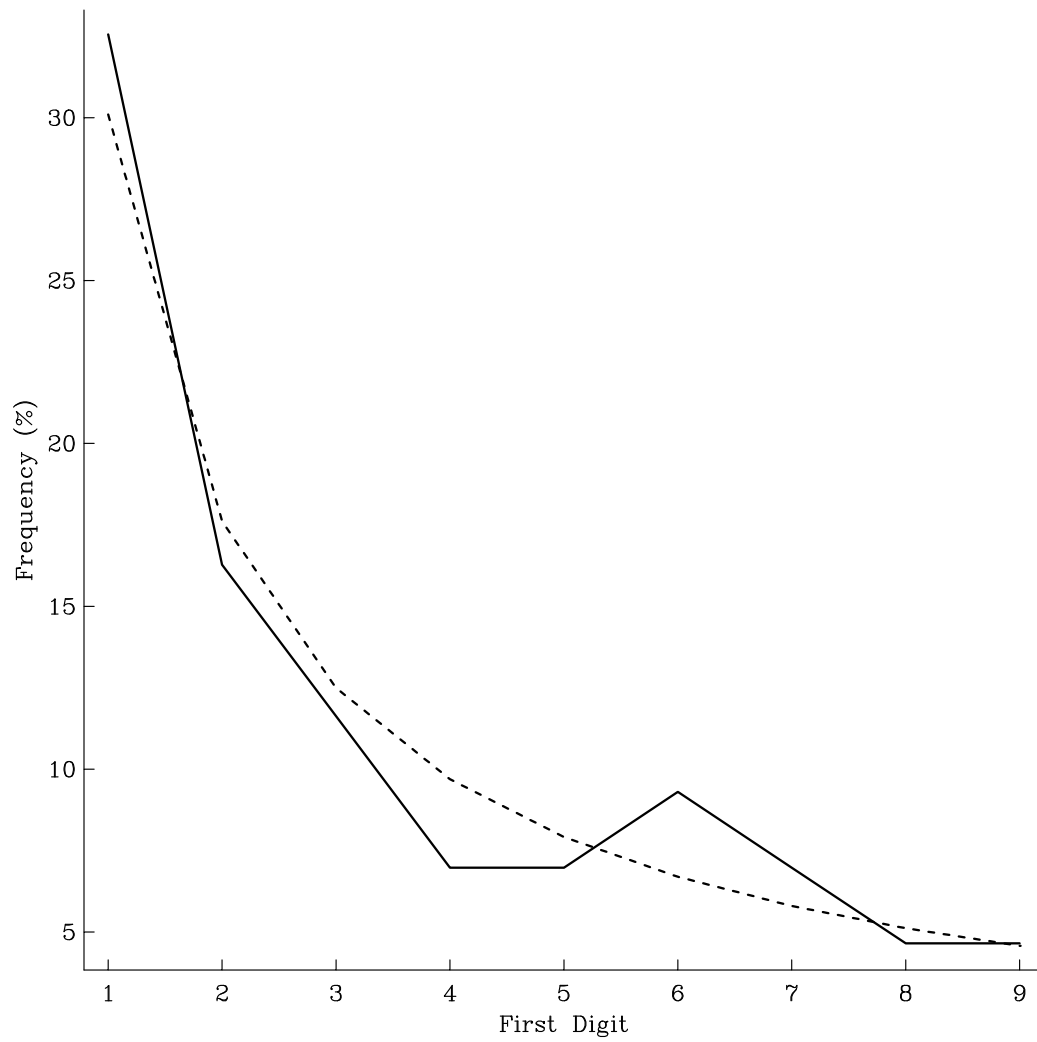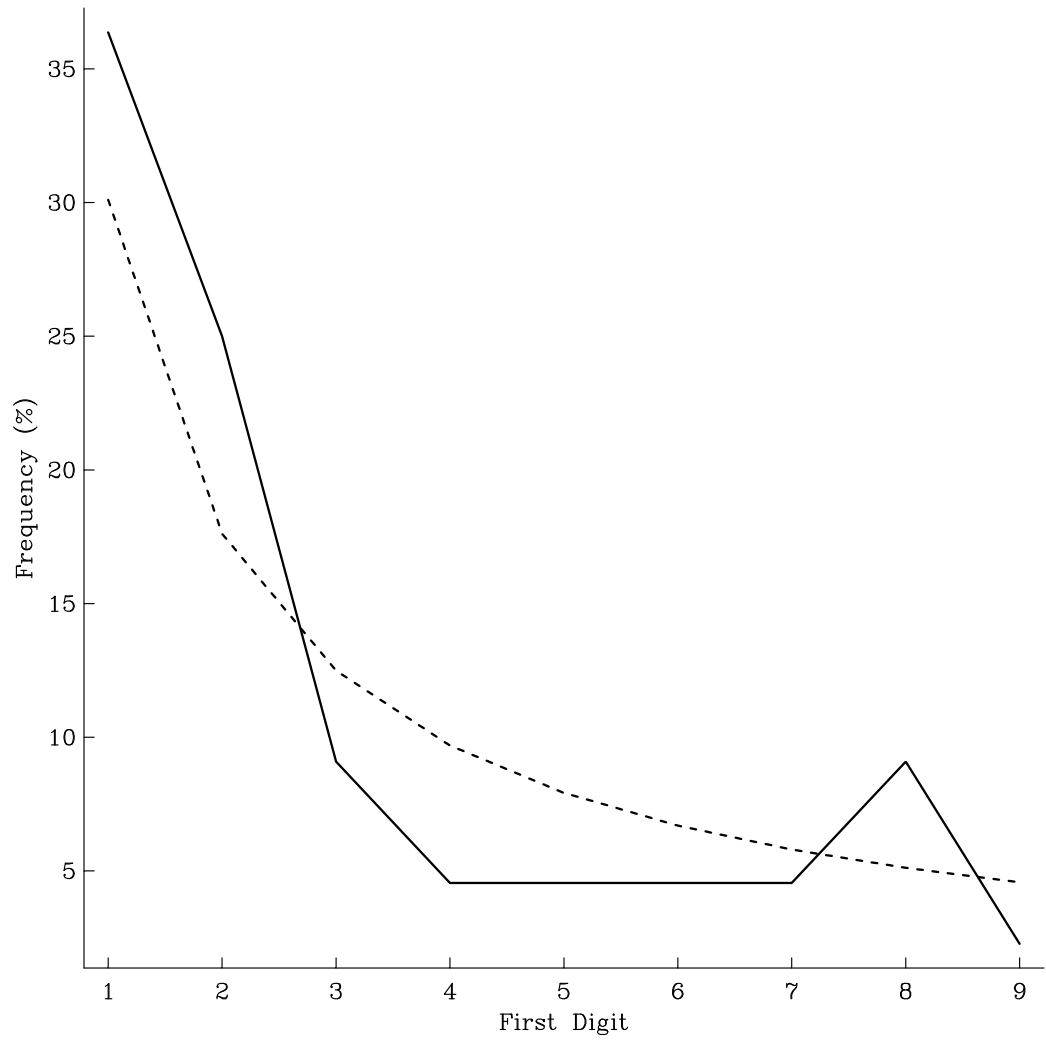
Benford canada Cumulative data

Benford czechrepublic Cumulative data

Benford denmark Cumulative data
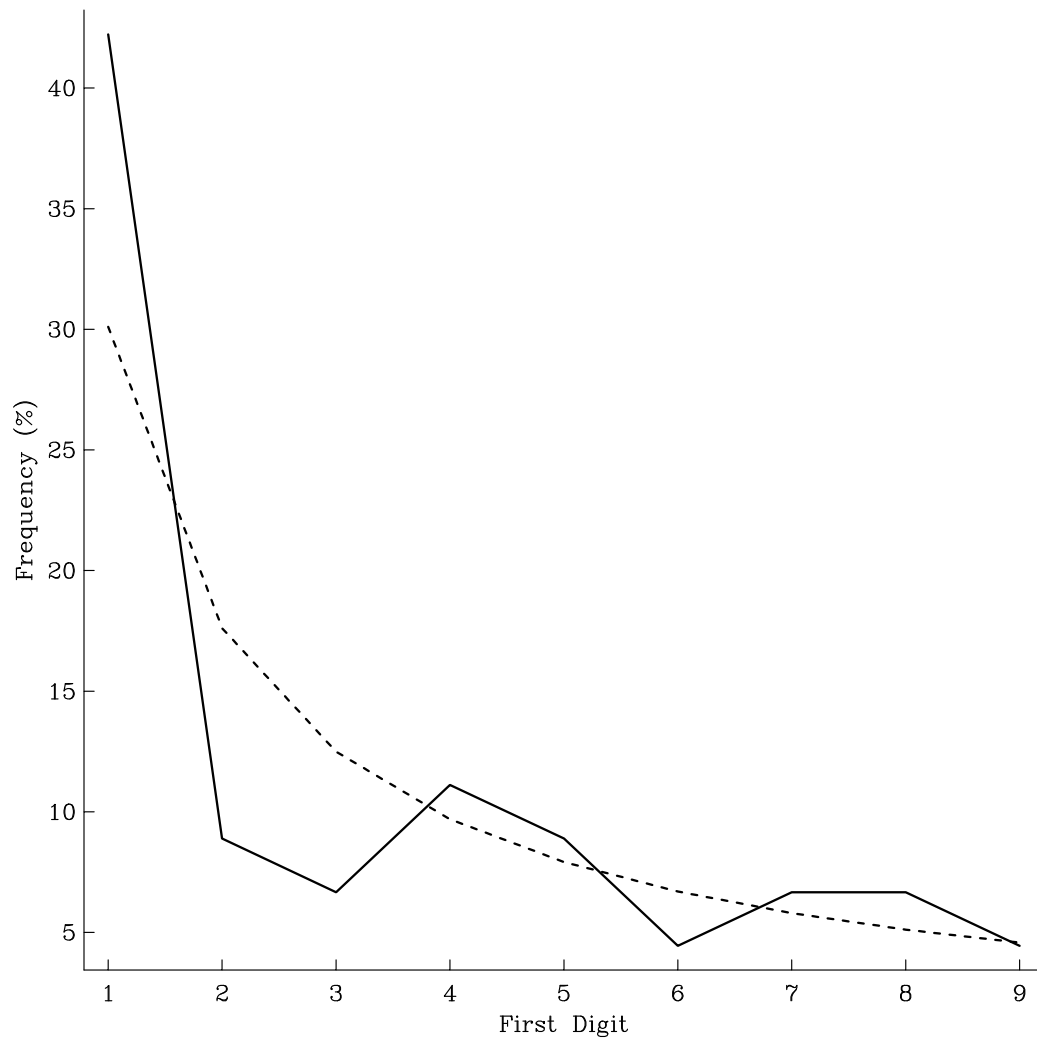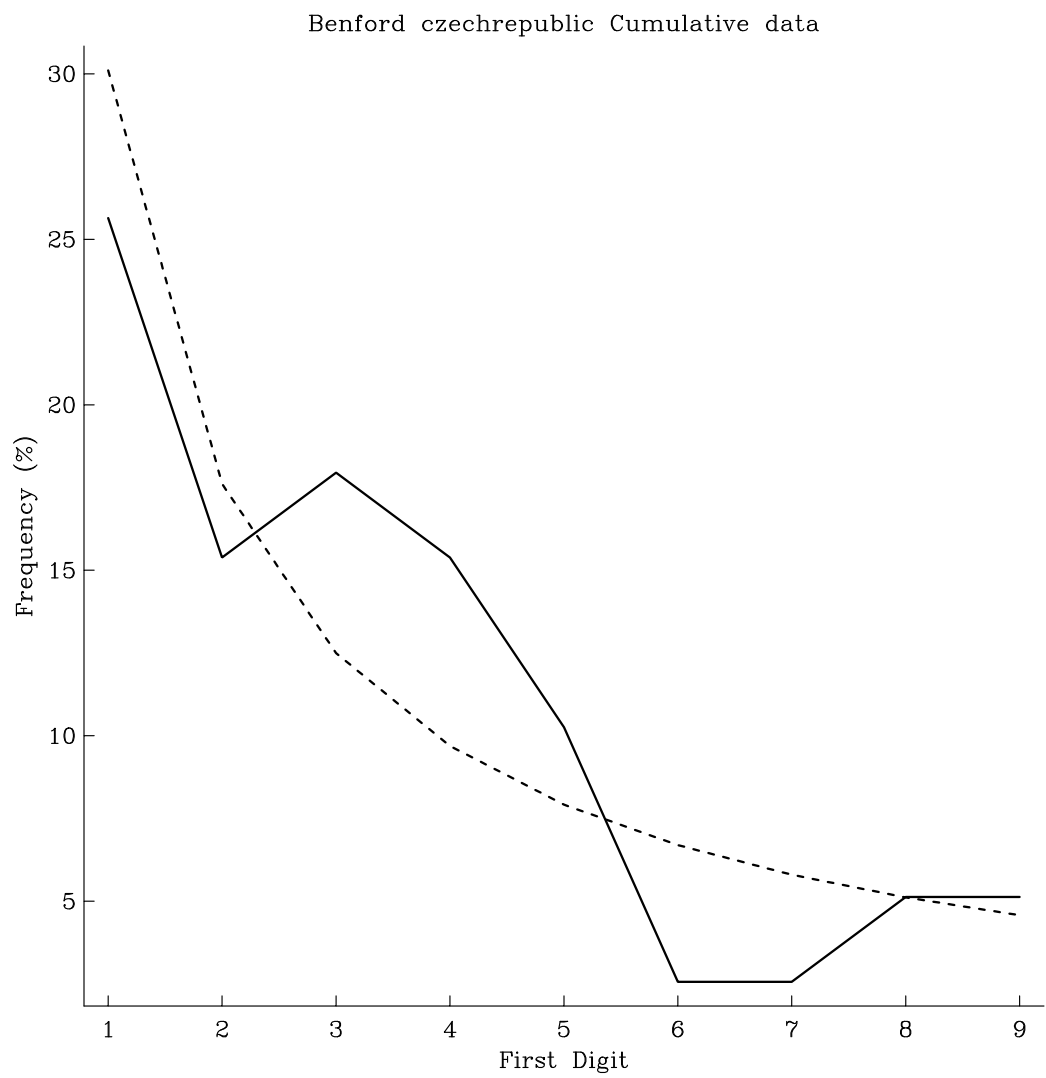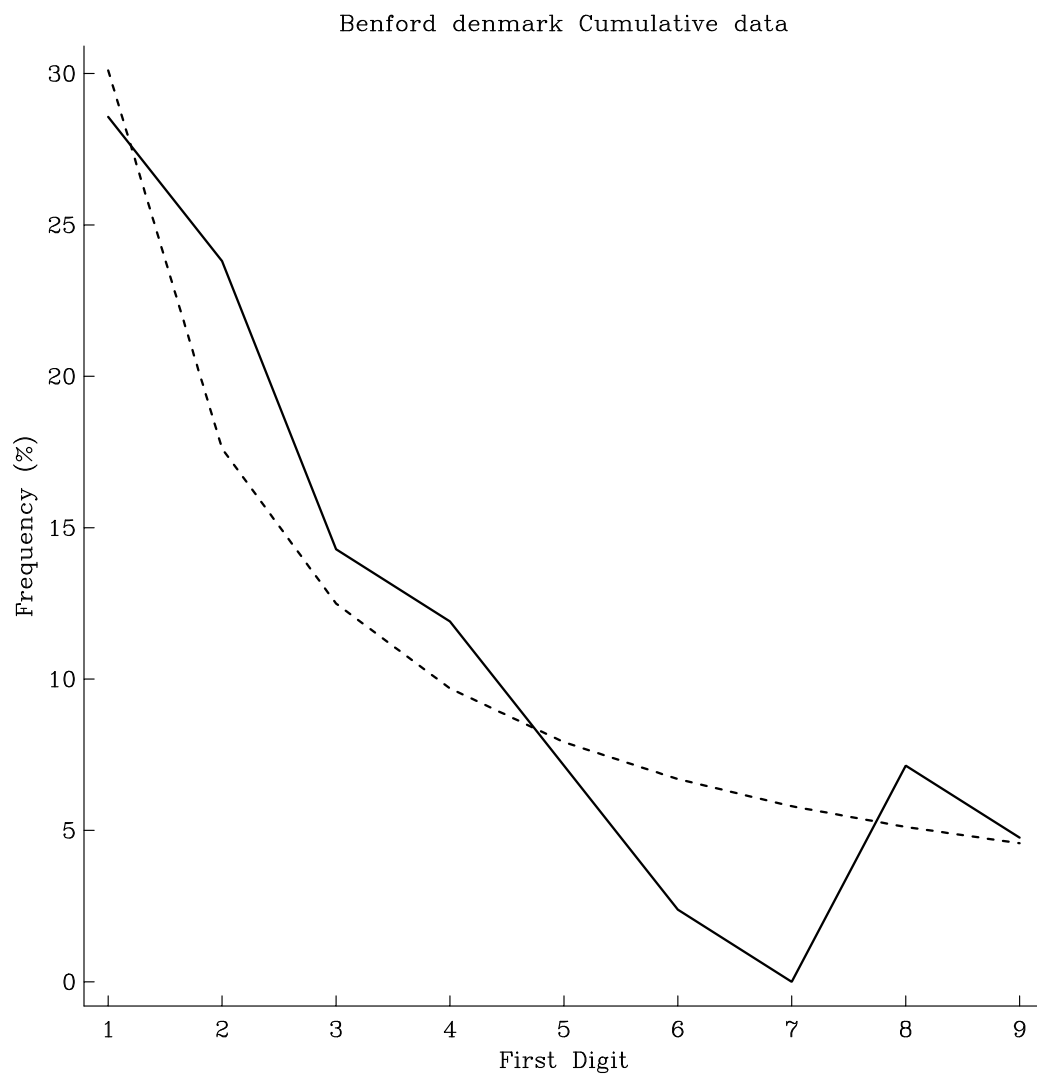
Benford egypt Cumulative data

Benford estonia Cumulative data
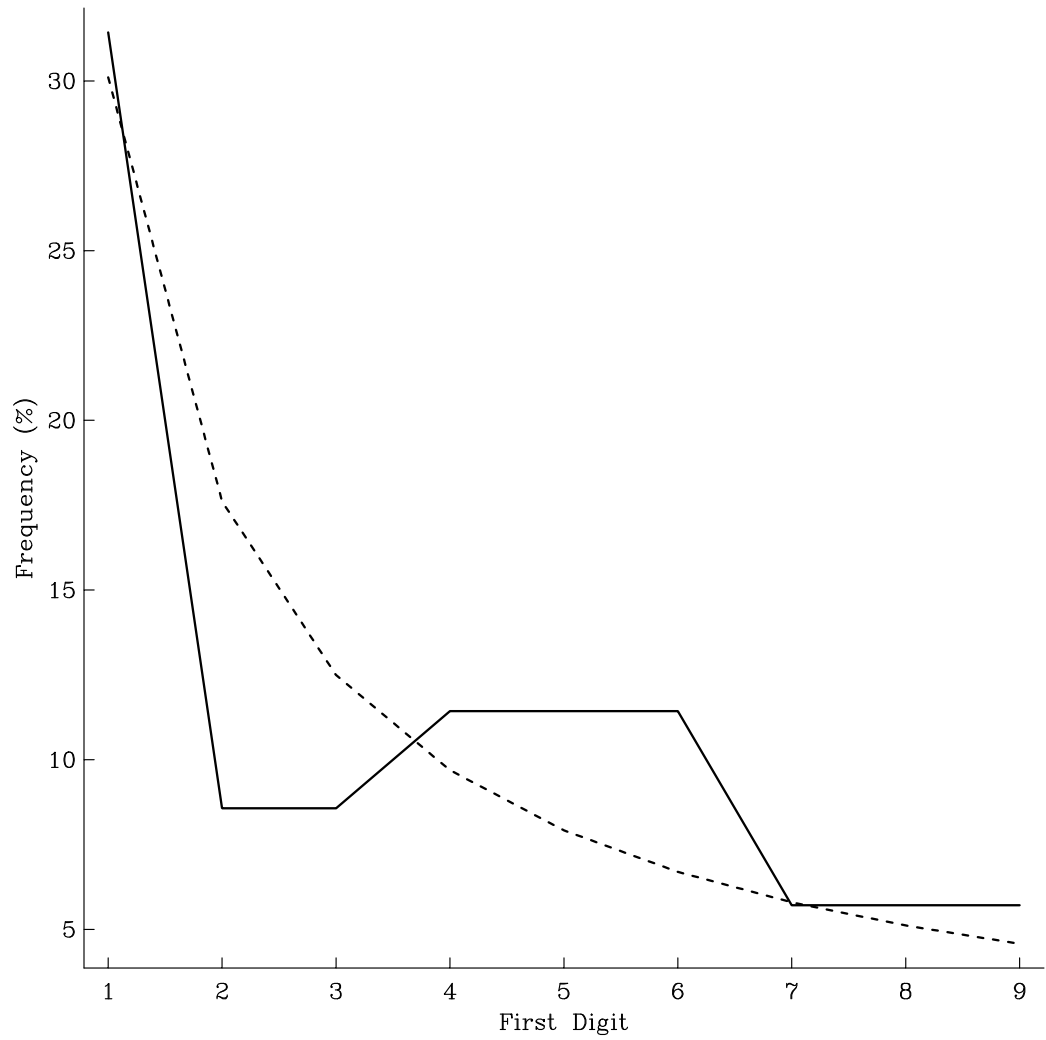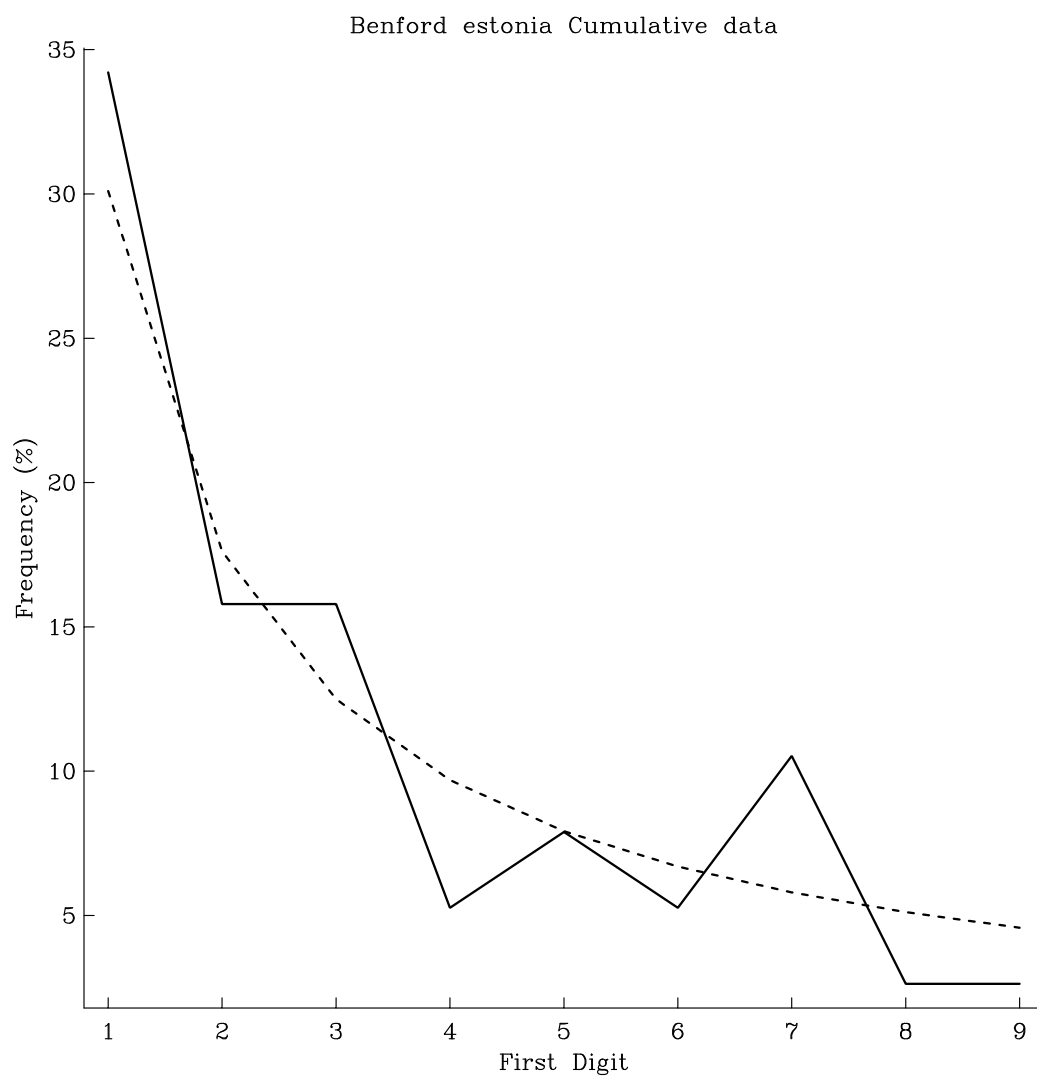
Benford finland Cumulative data

Benford france Cumulative data

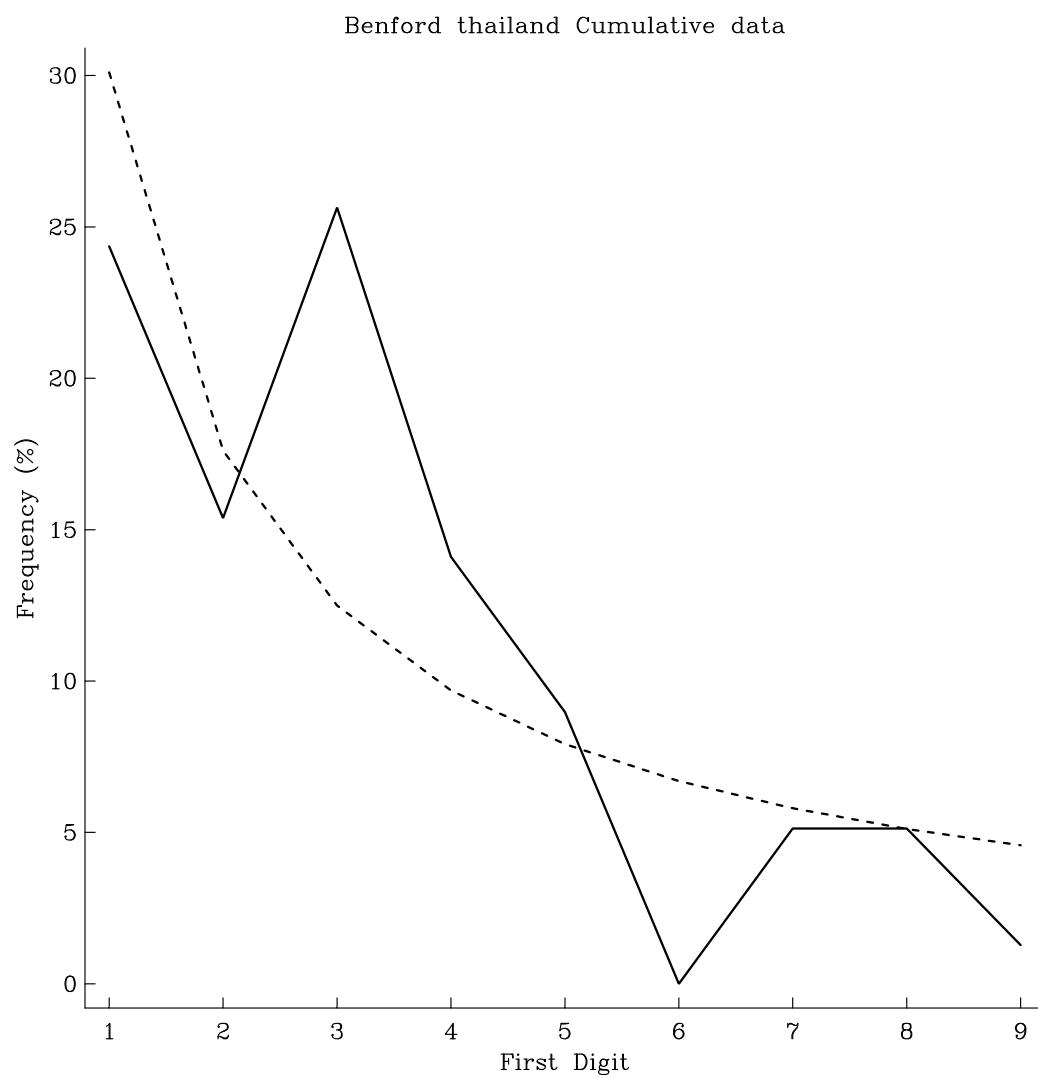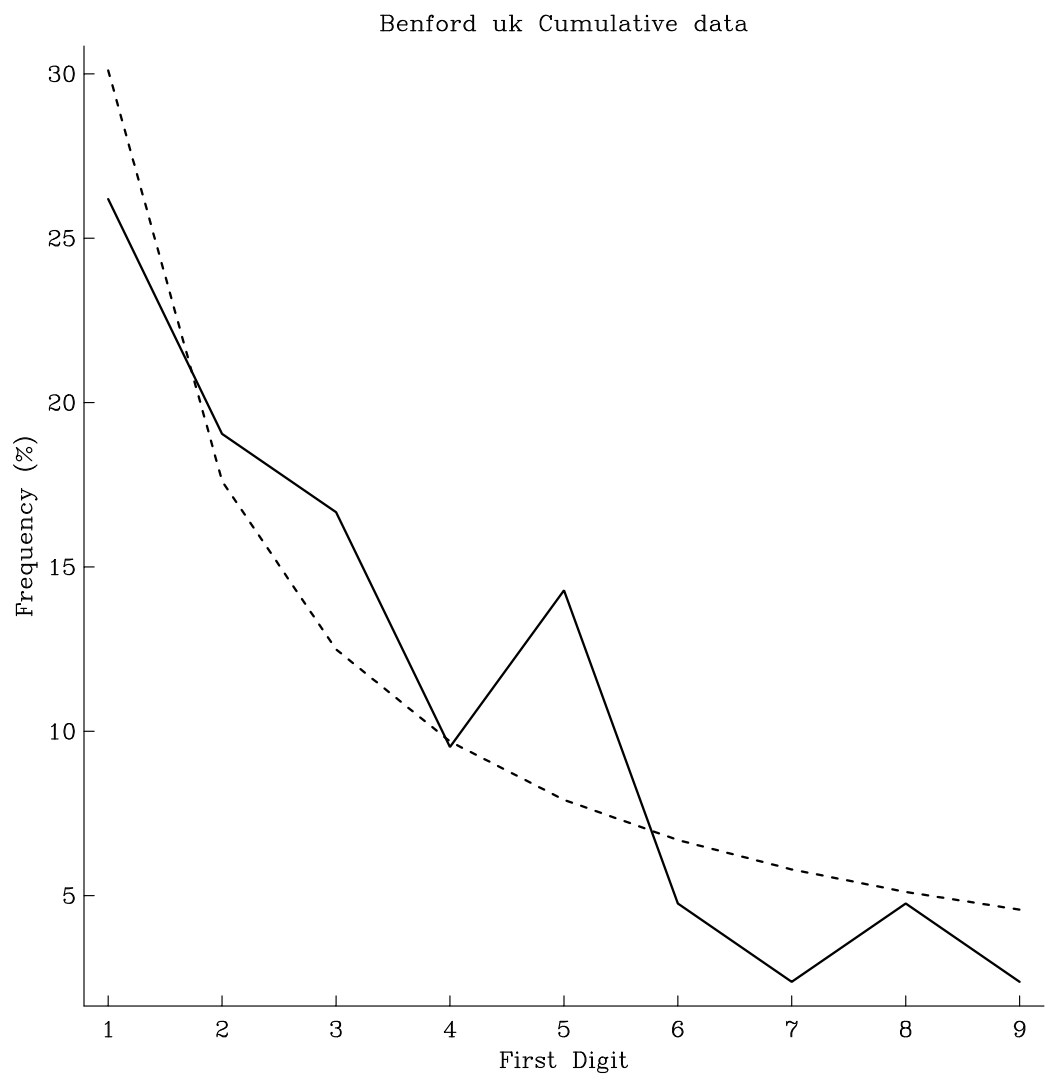Benford germany Cumulative data
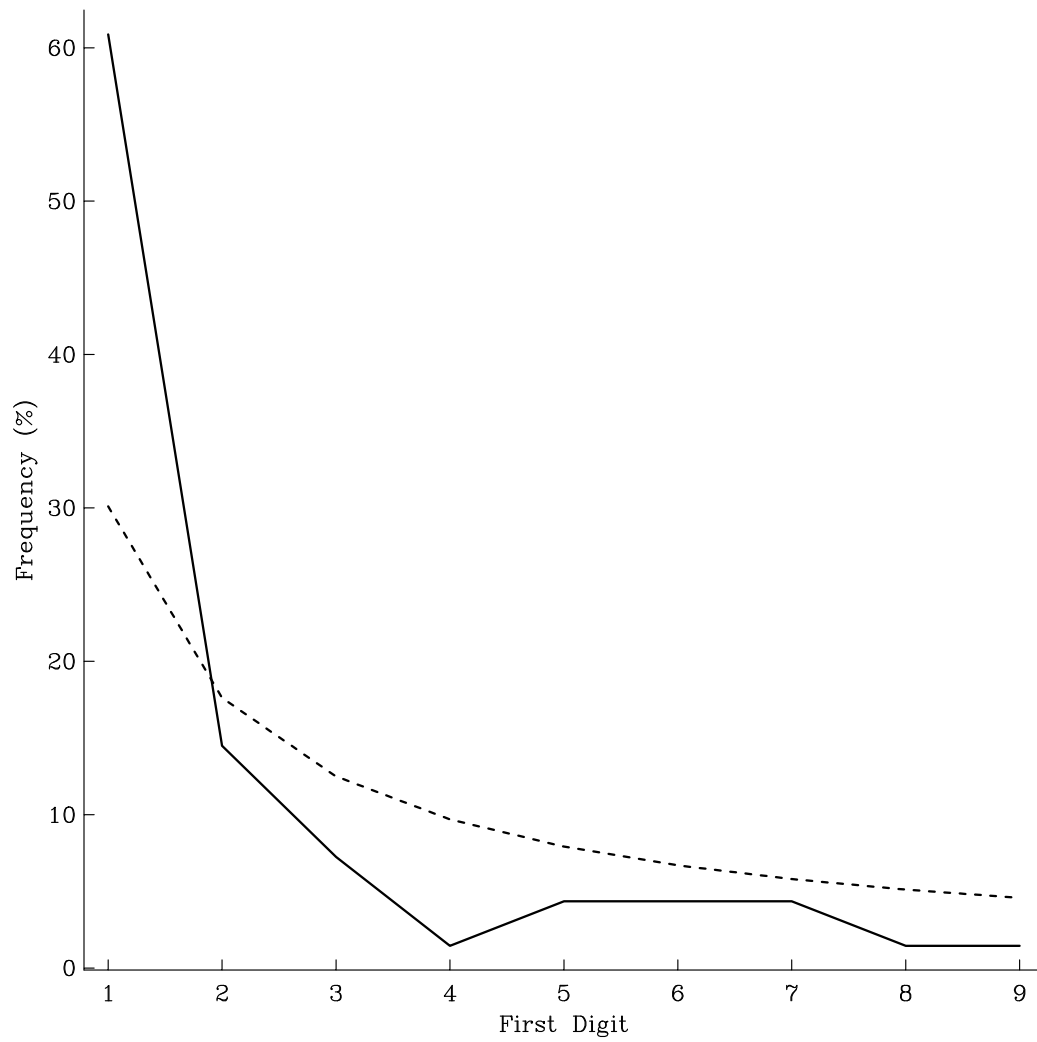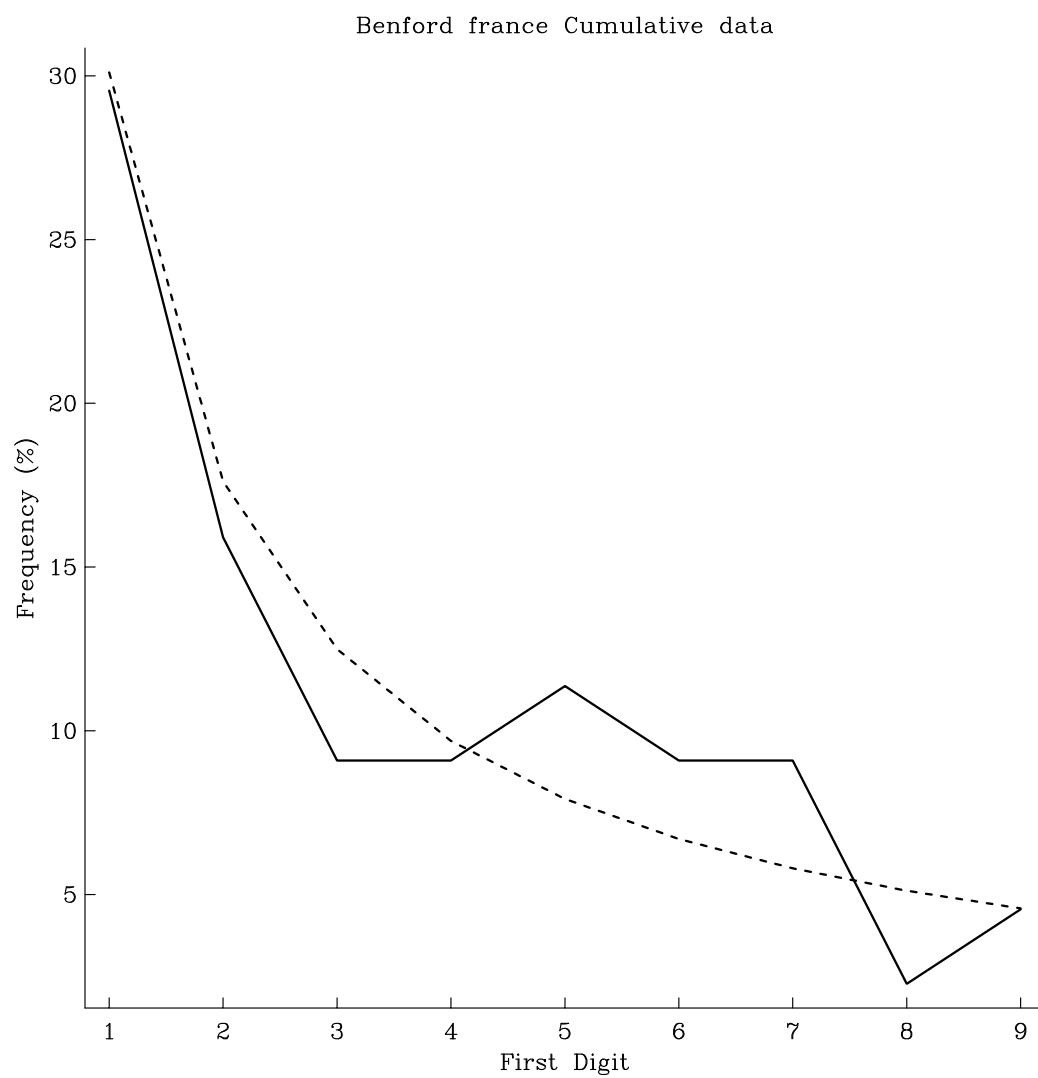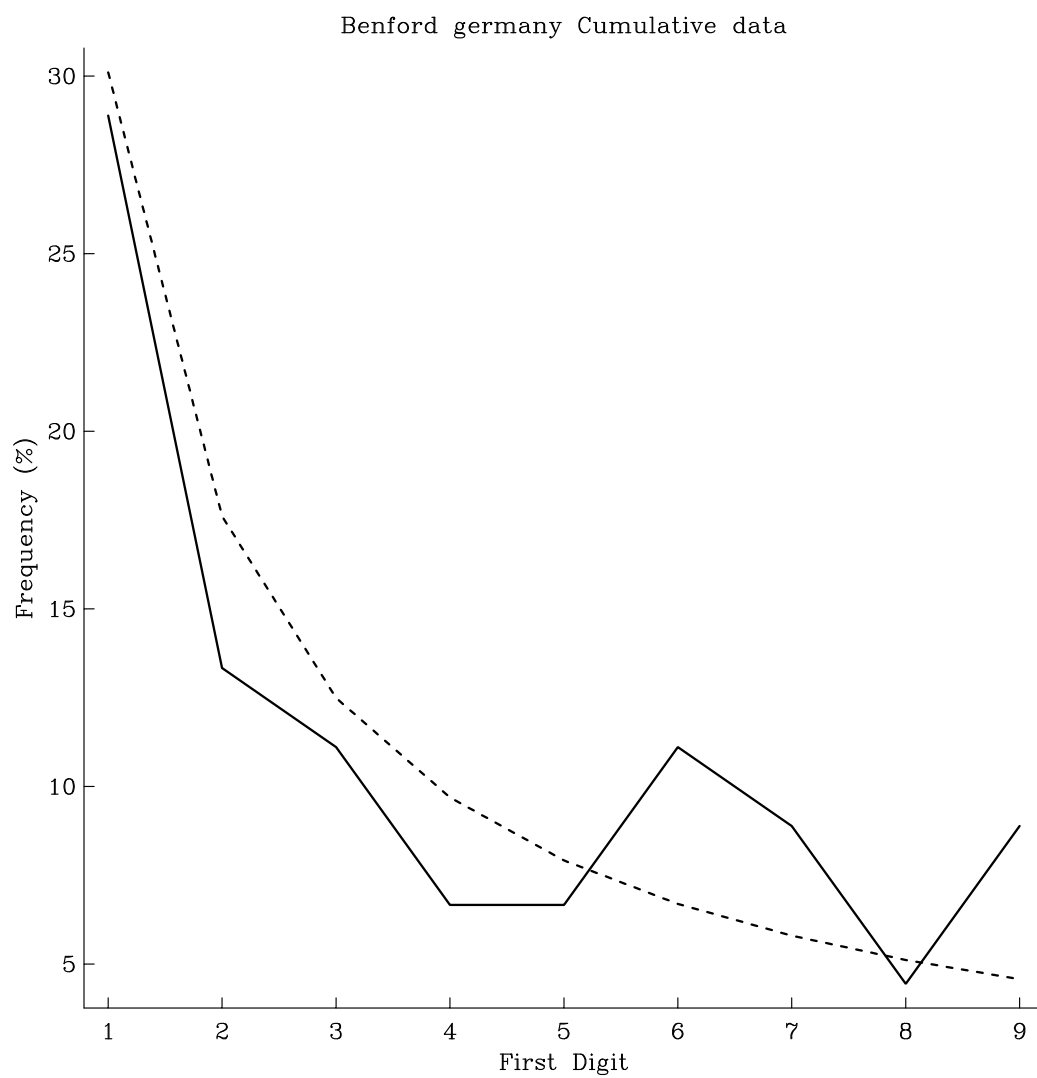
Benford greece Cumulative data

Benford hungary Cumulative data
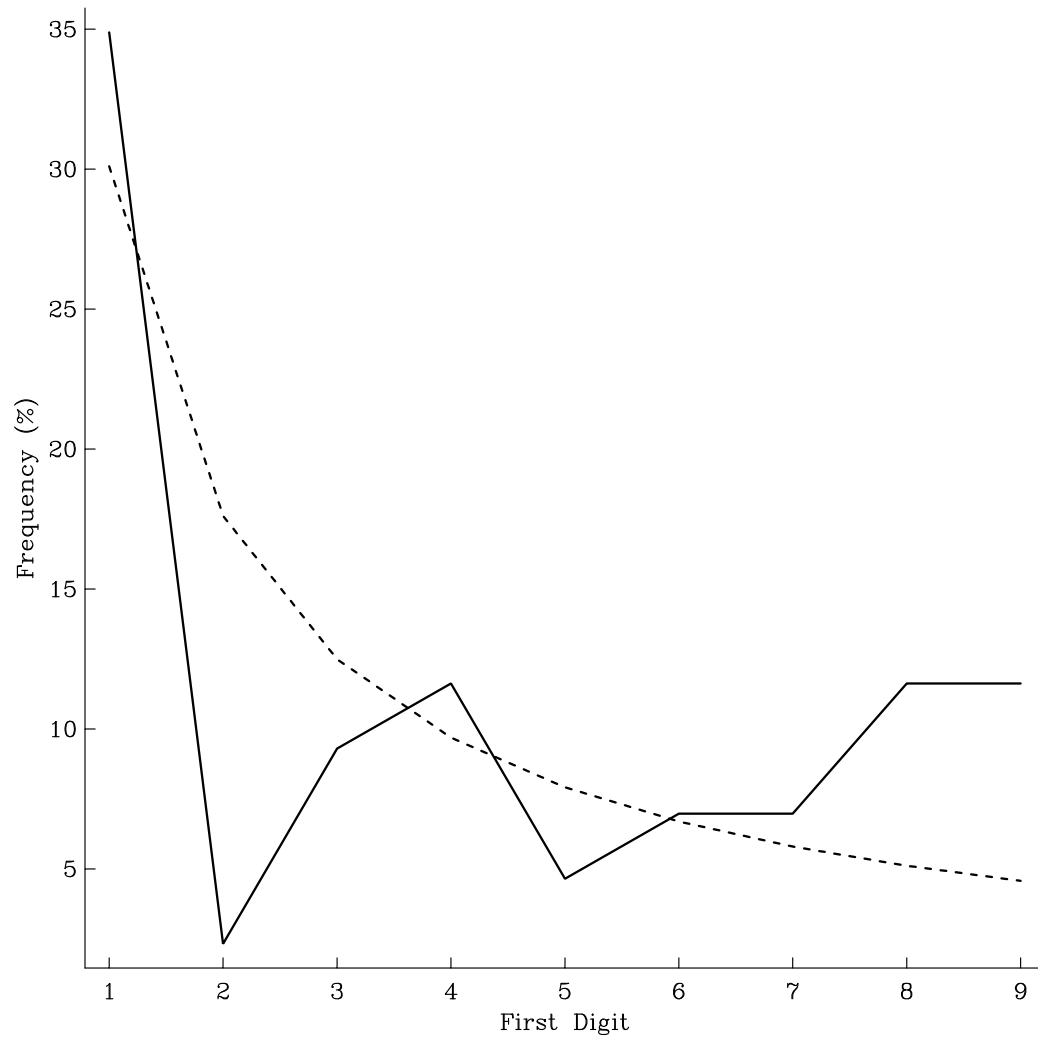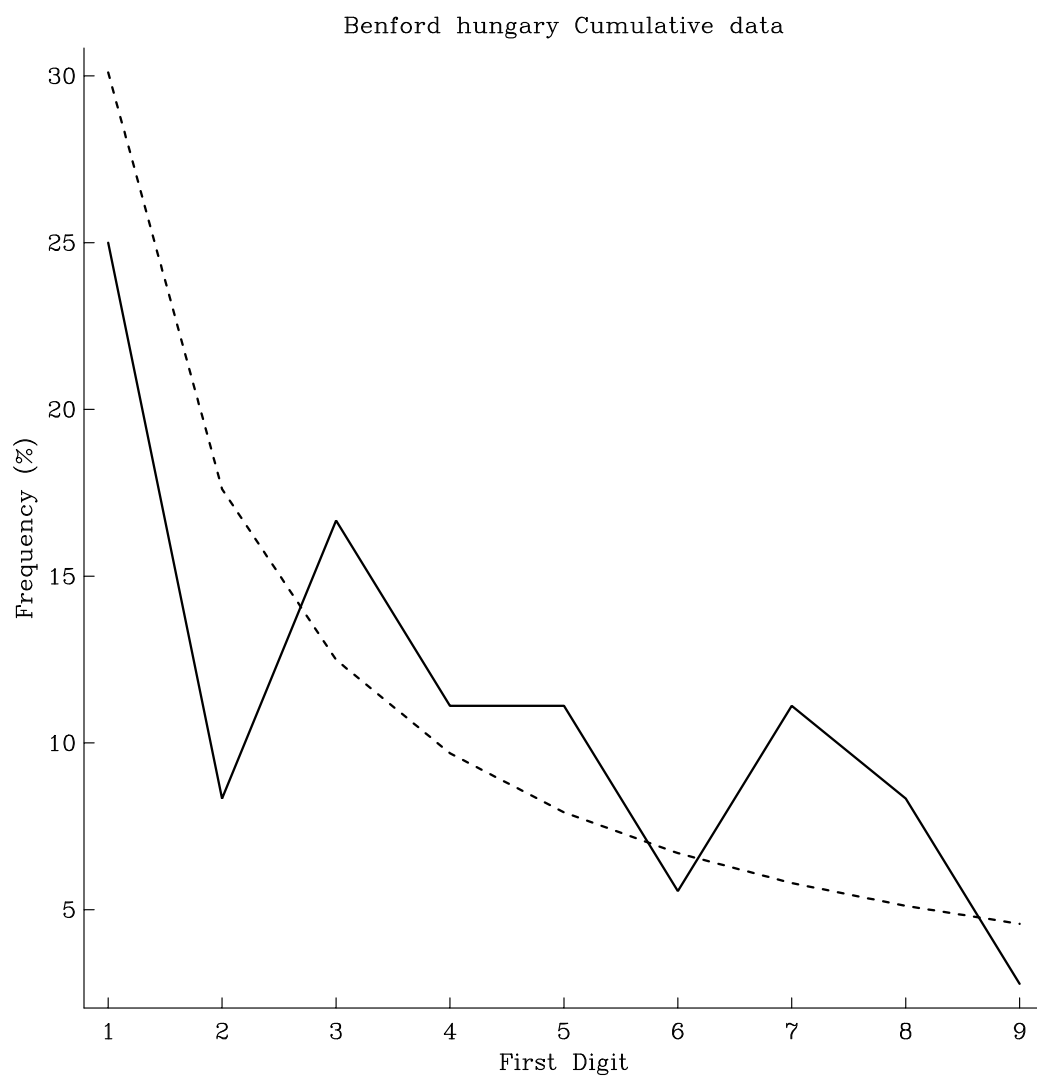
Benford iceland Cumulative data

Benford india Cumulative data

Benford indonesia Cumulative data
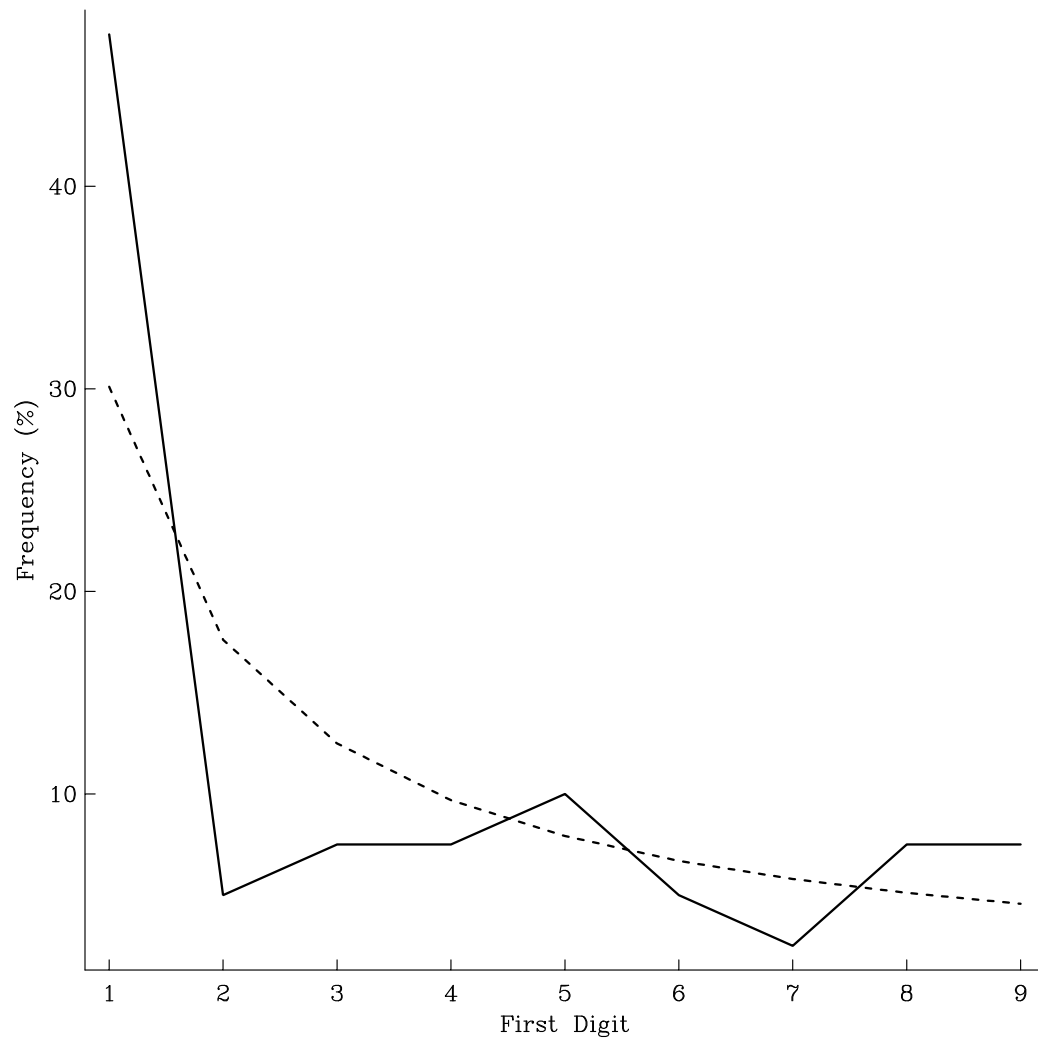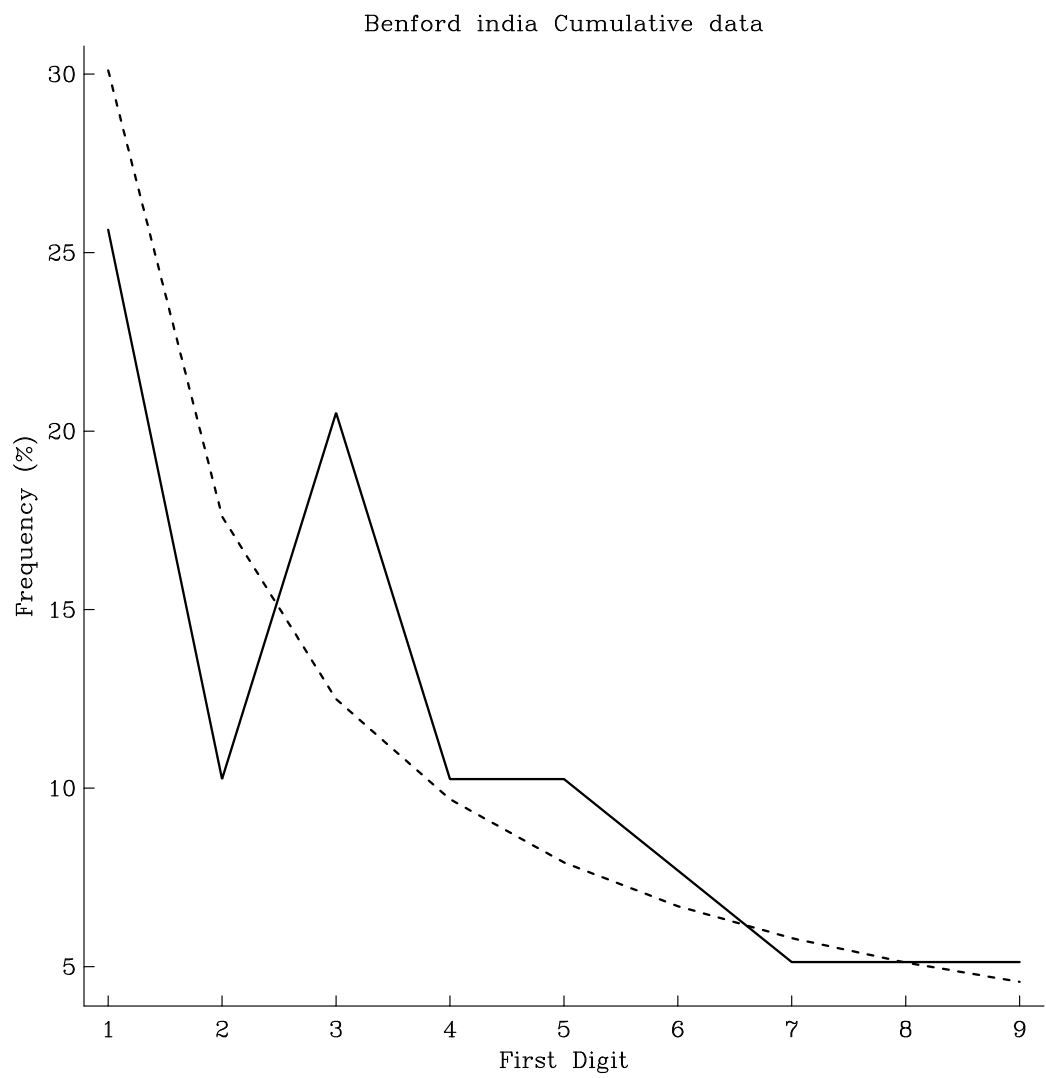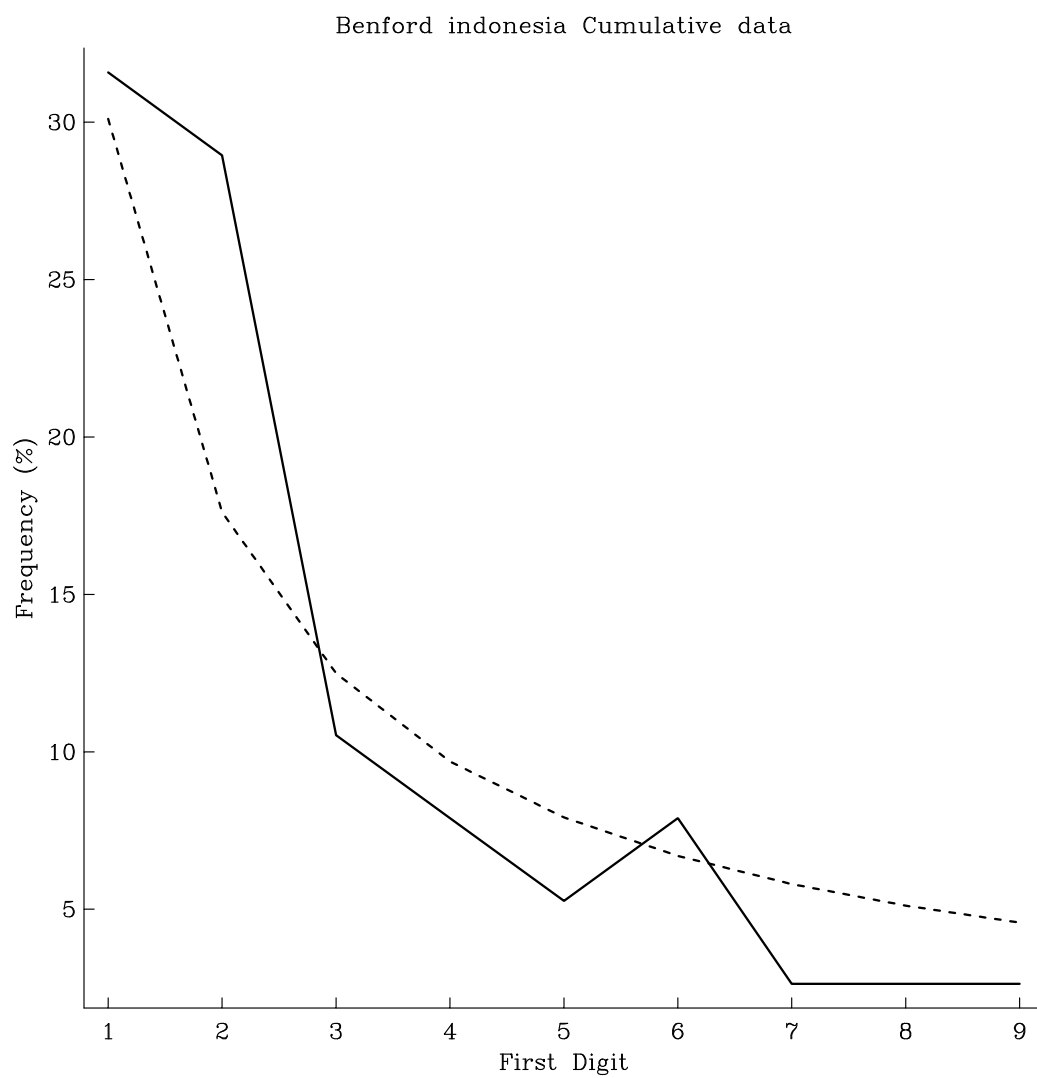
Benford iran Cumulative data
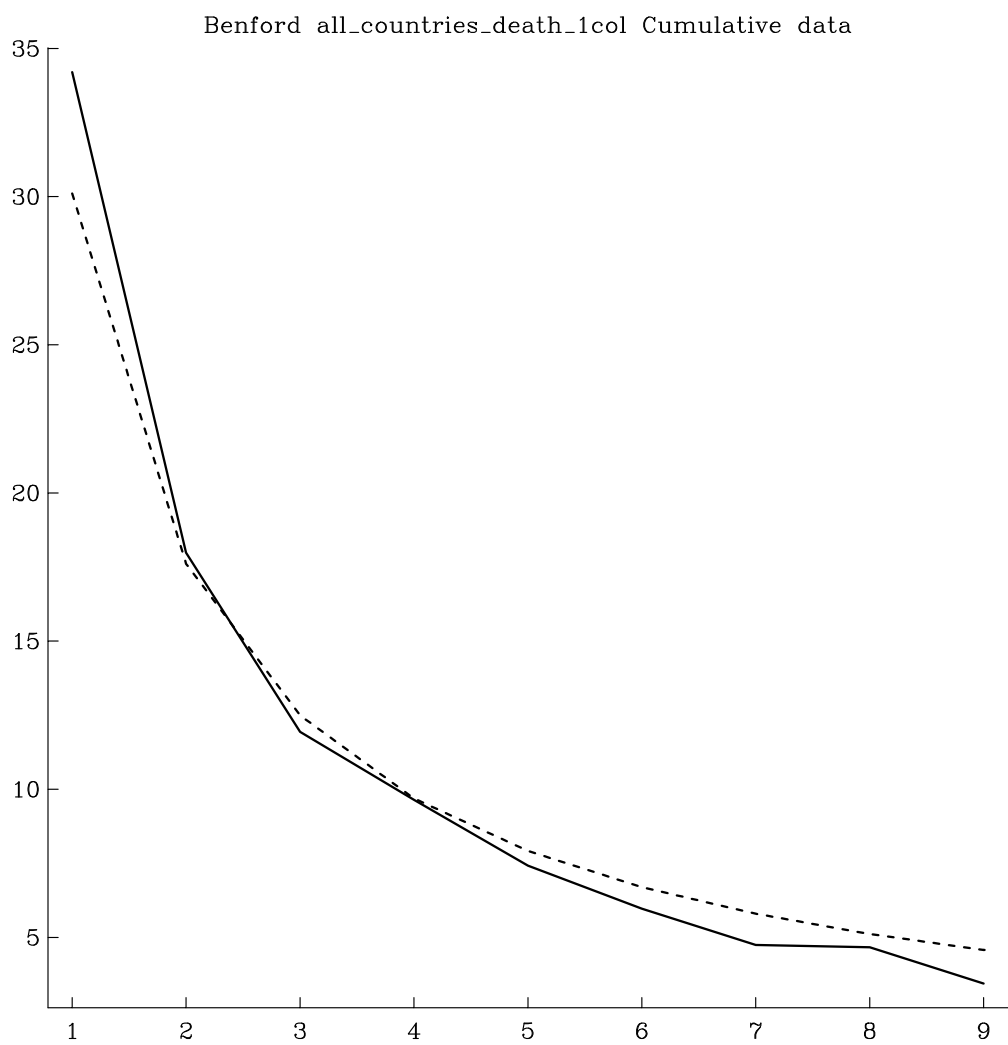
Benford ireland Cumulative data
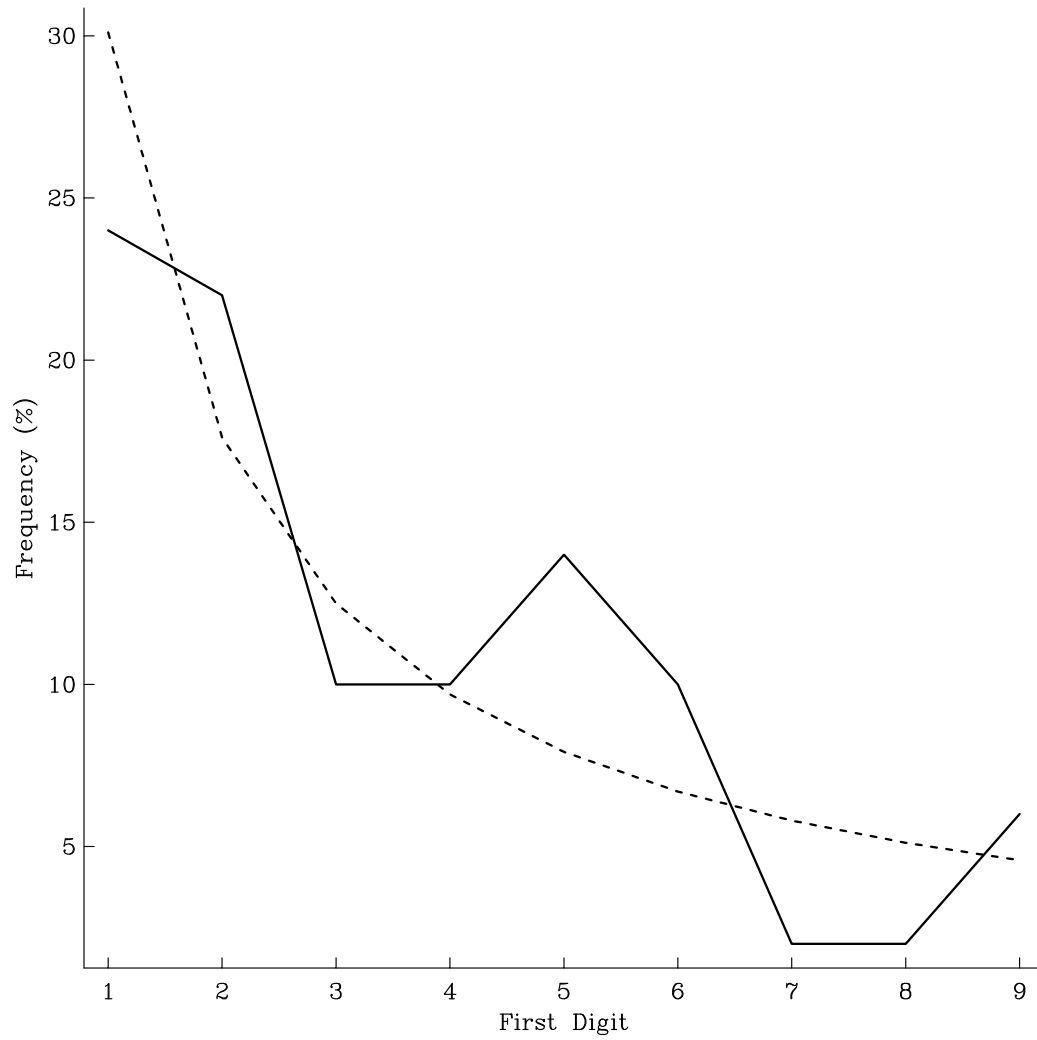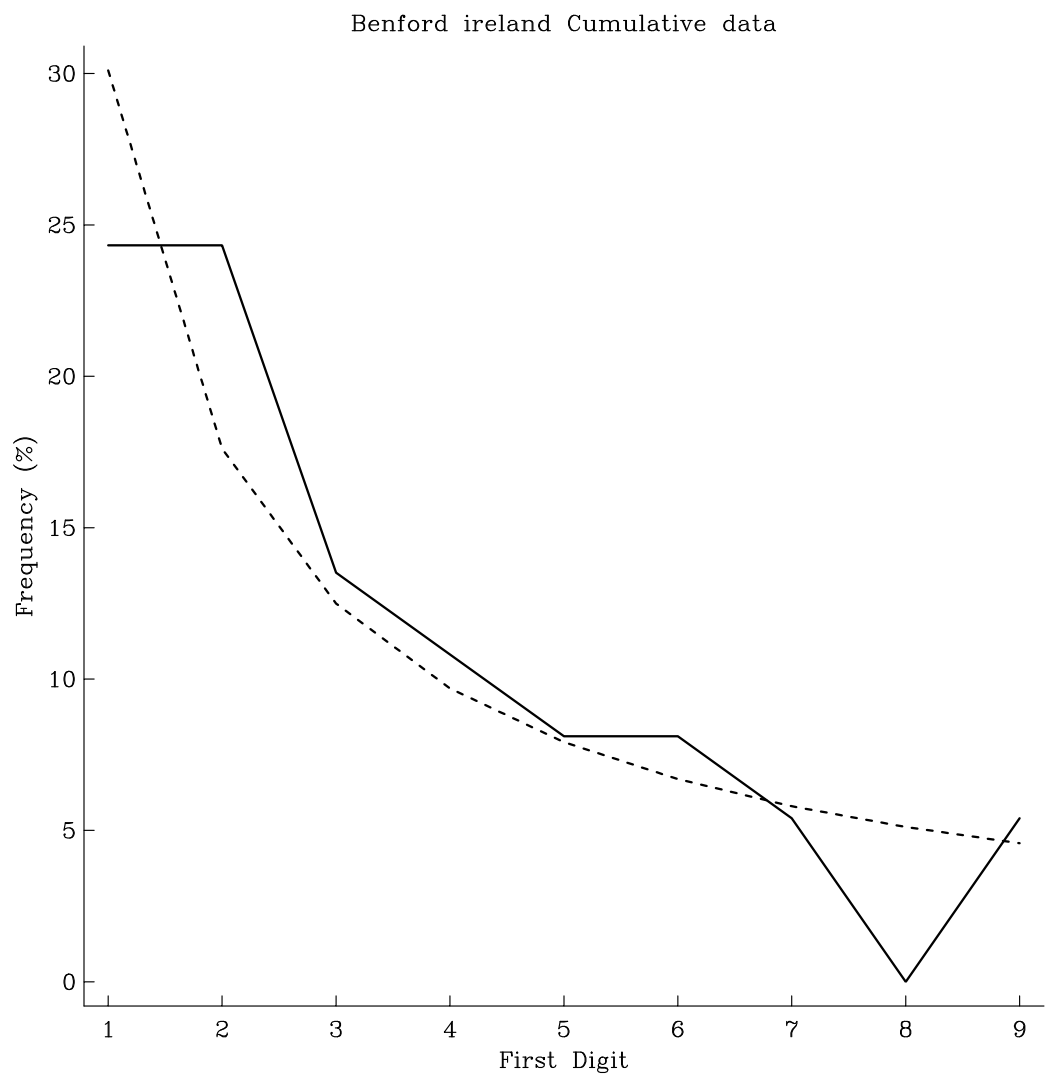
Benford israel Cumulative data
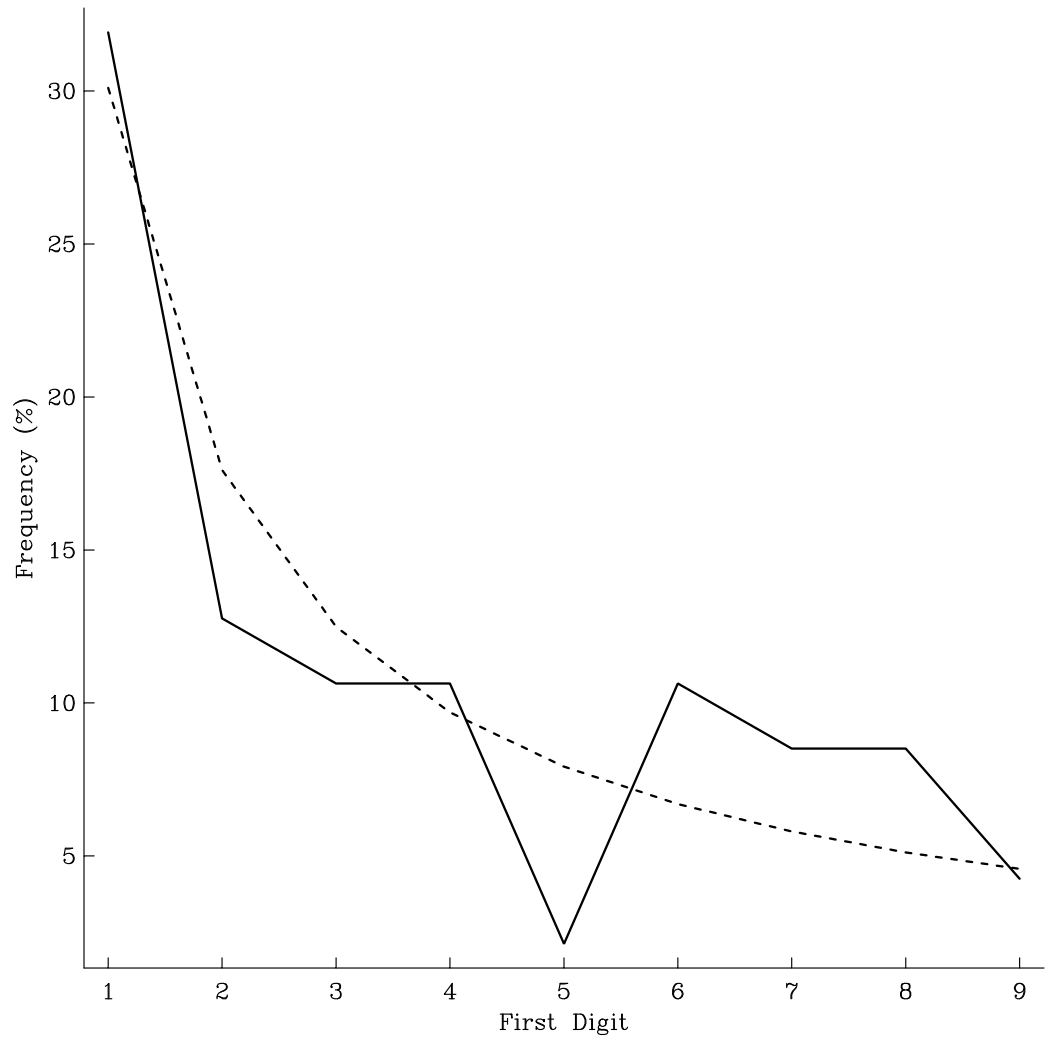
Benford italy Cumulative data

Frequency (%)

First Digit
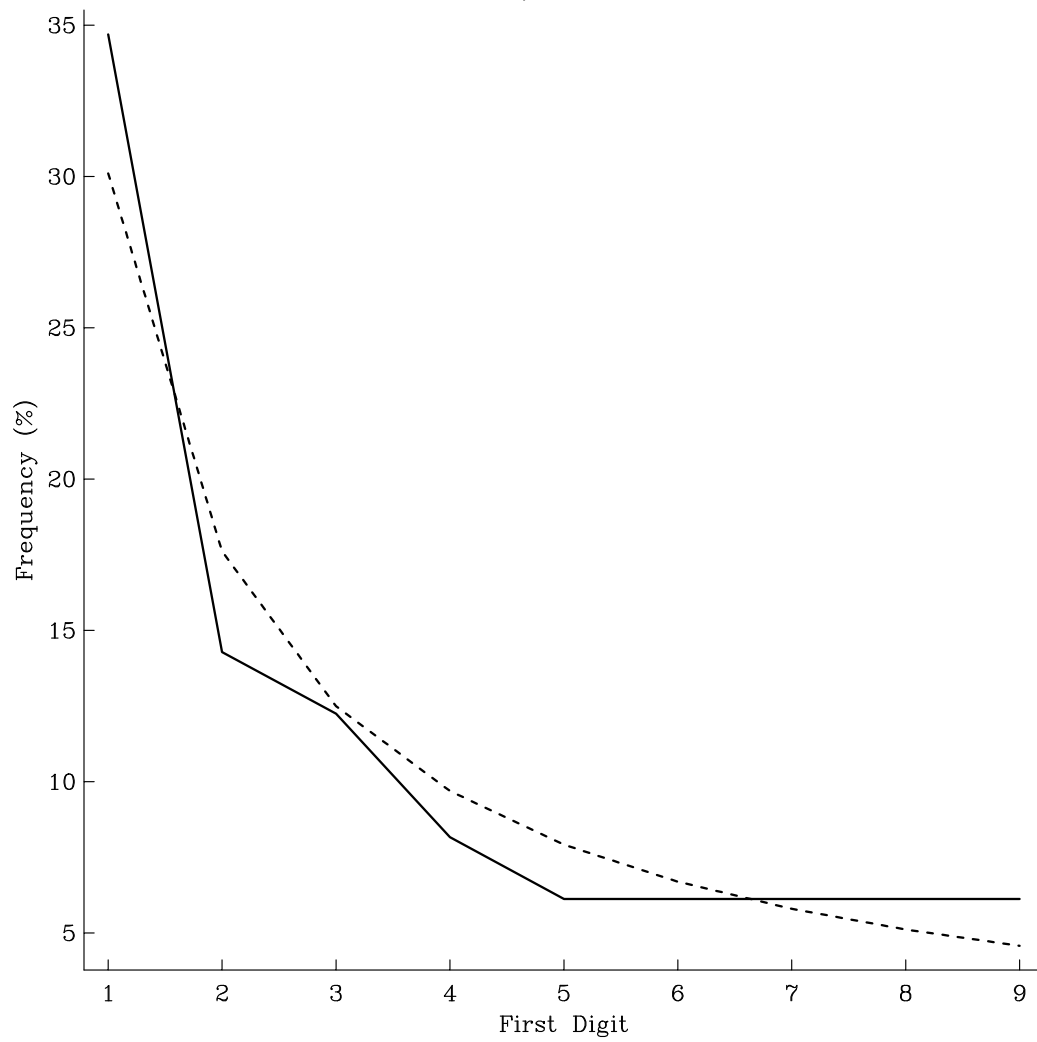
Benford japan Cumulative data
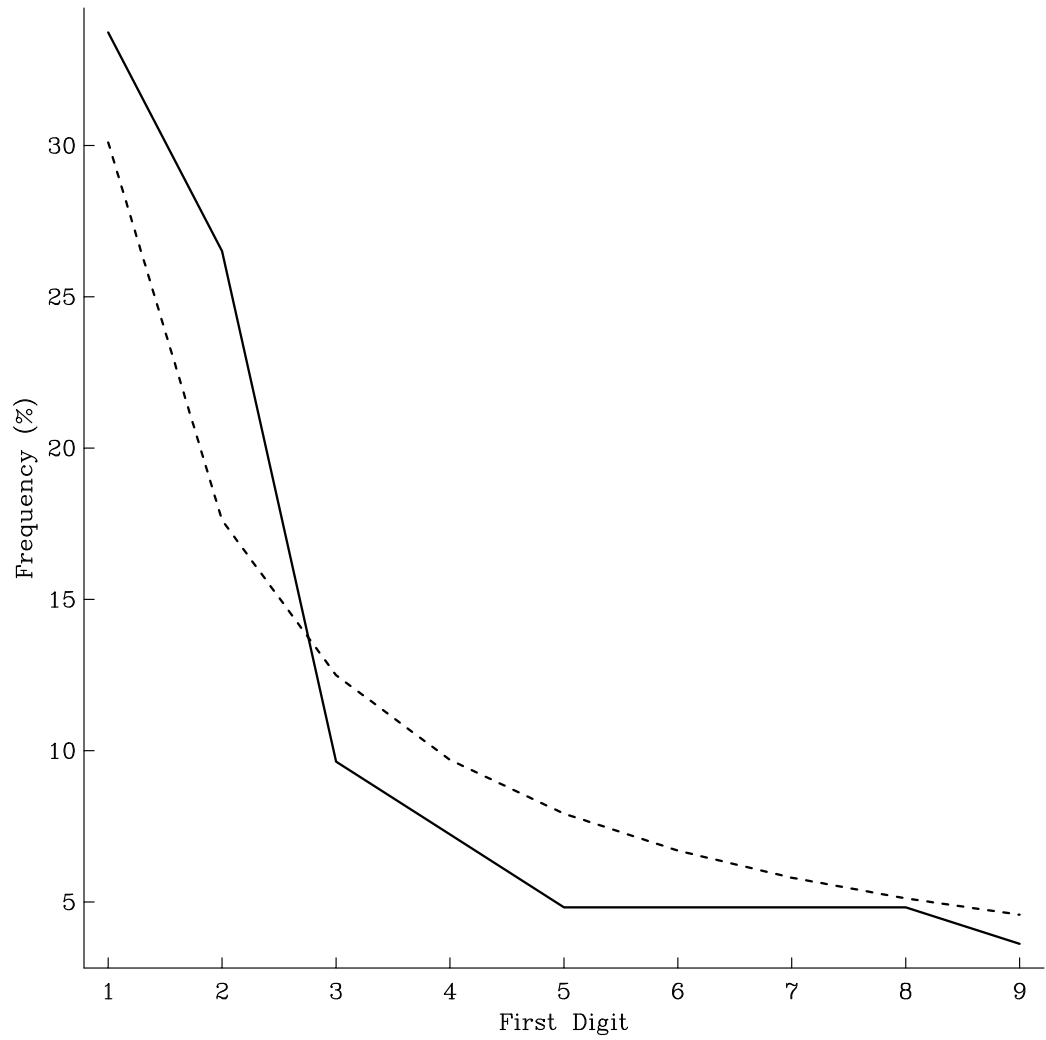
Benford kuwait Cumulative data

Benford lombardy Cumulative data
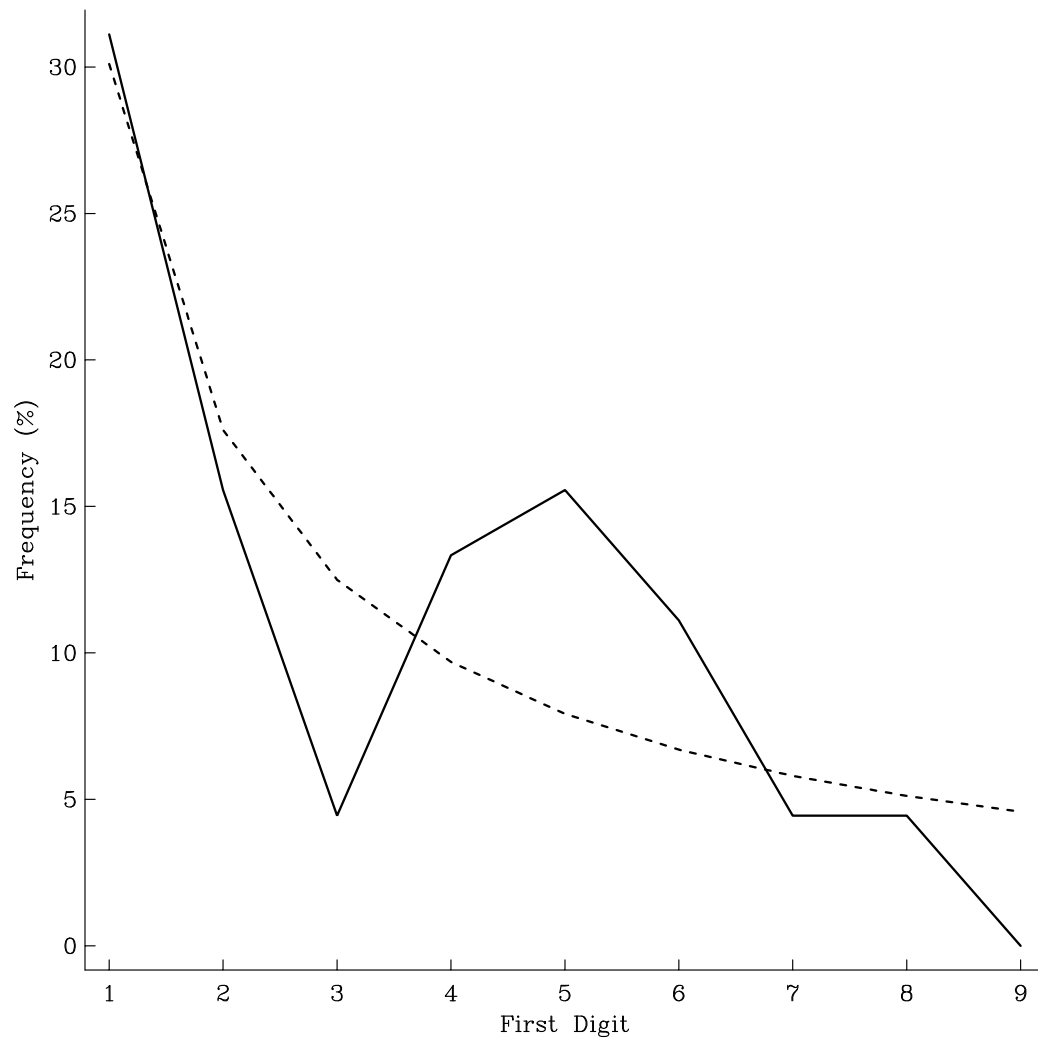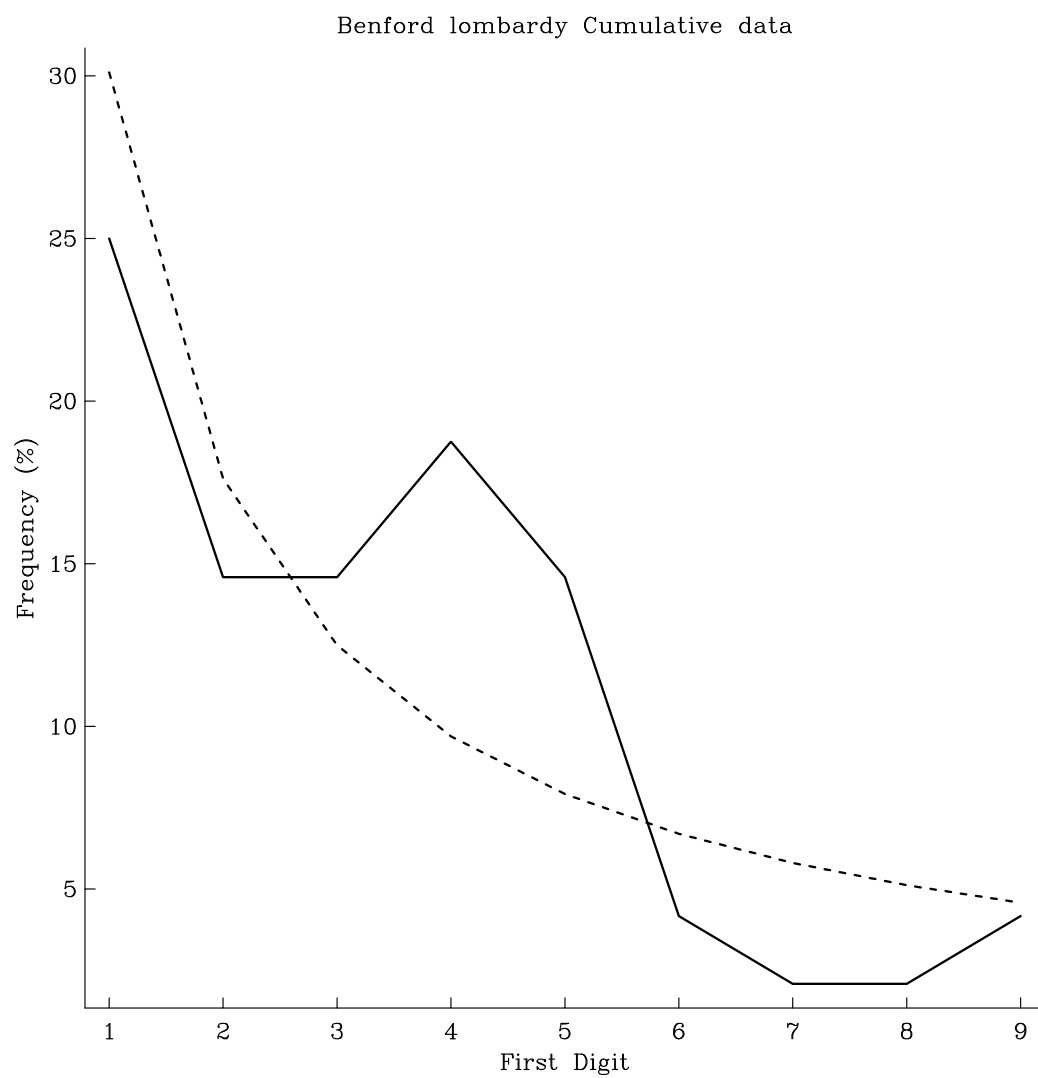
Benford luxembourg Cumulative data

Benford malaysia Cumulative data
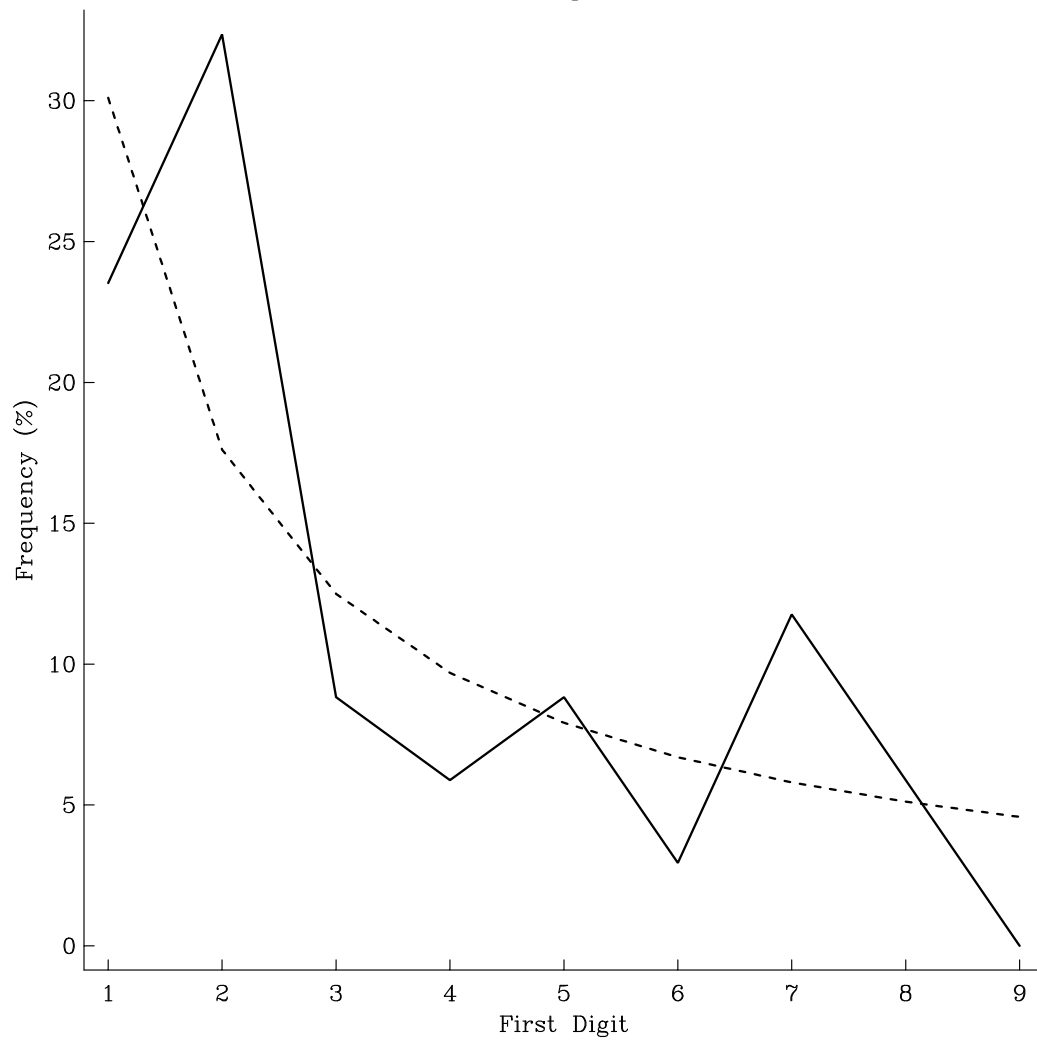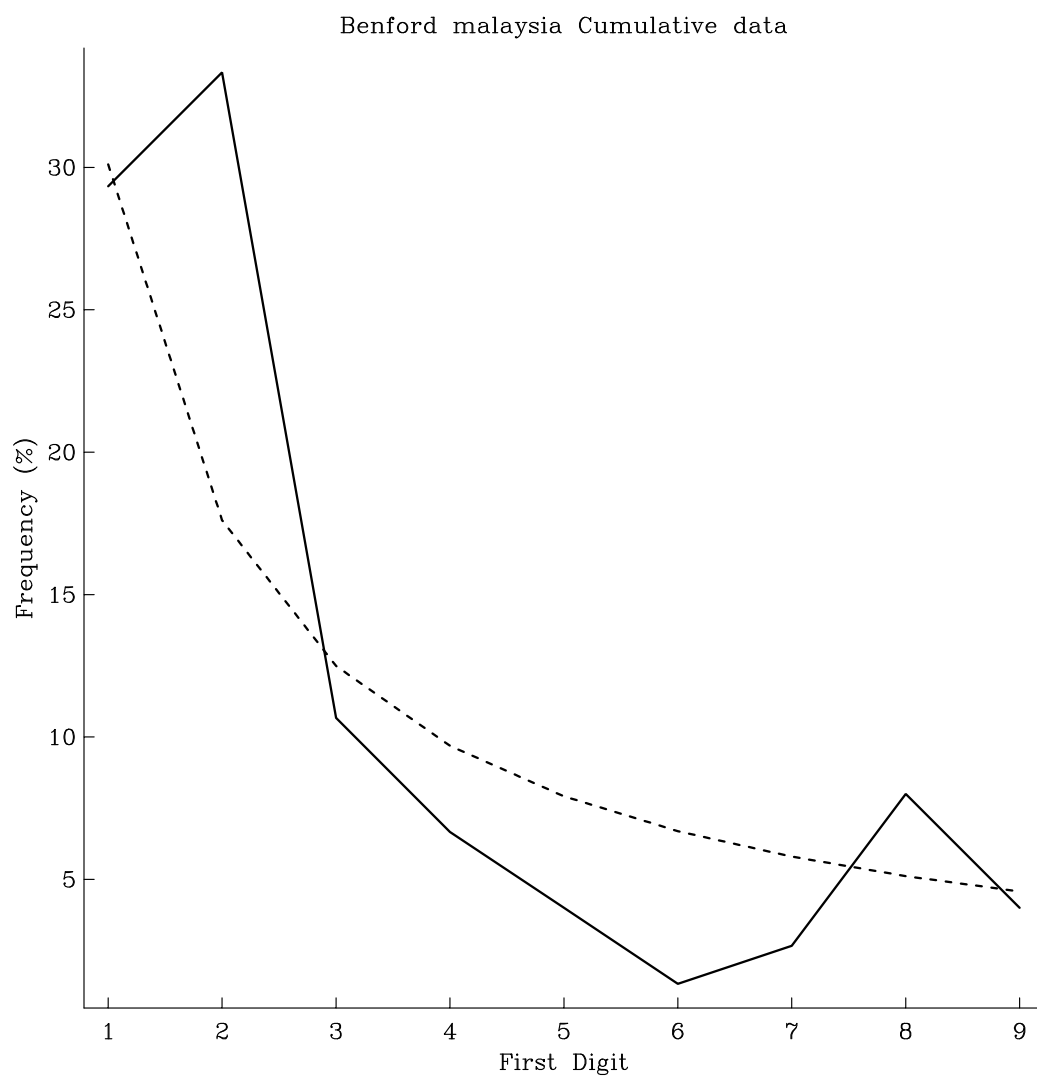
Benford netherlands Cumulative data
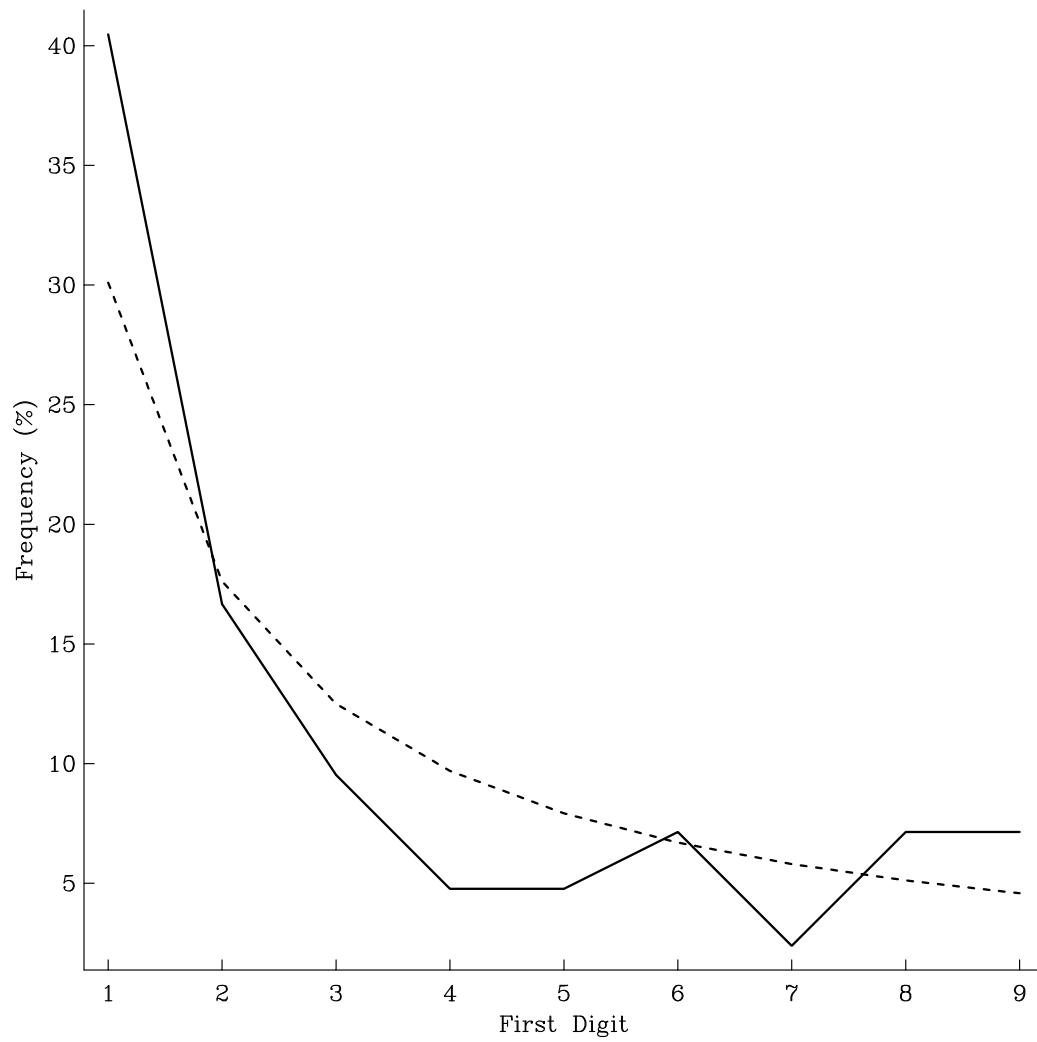
Benford newzealand Cumulative data

Benford norway Cumulative data

Benford pakistan Cumulative data
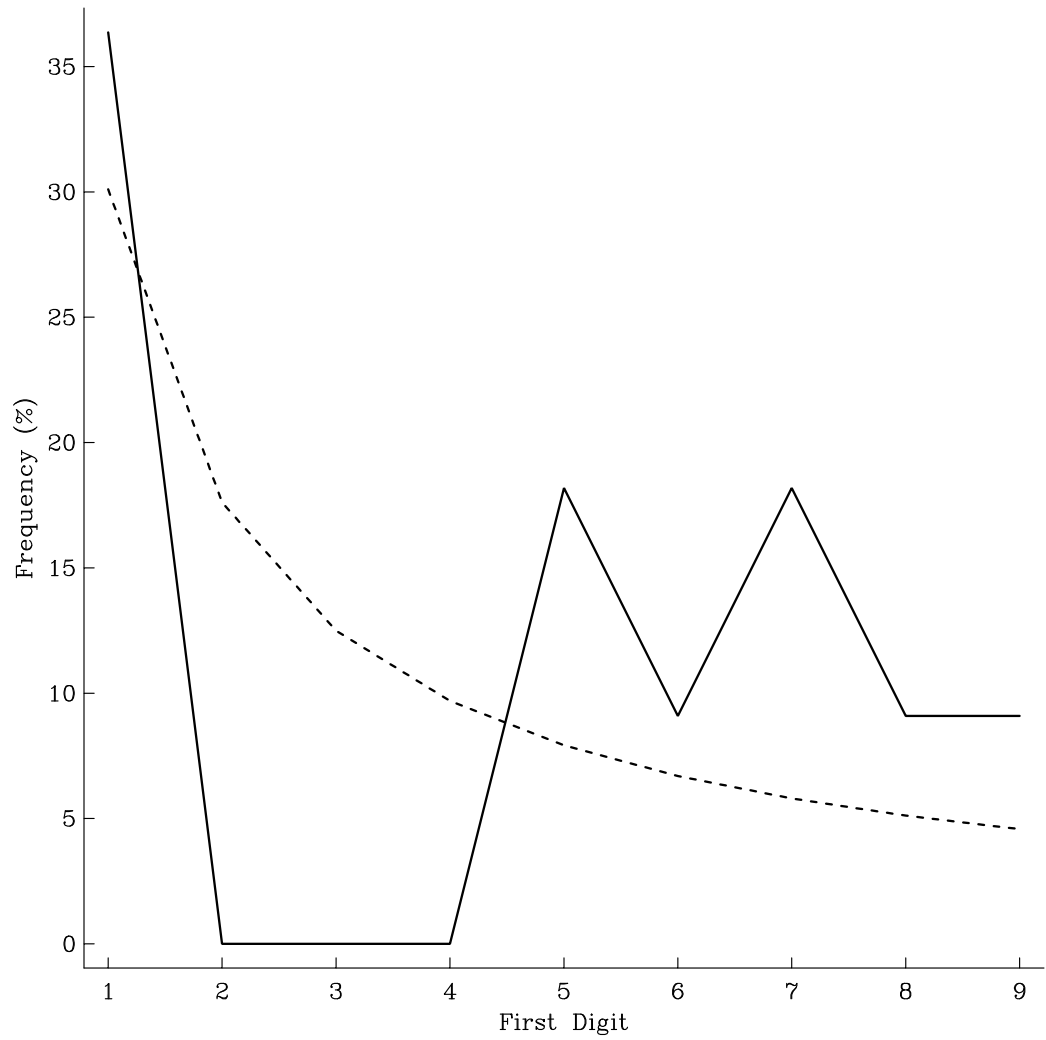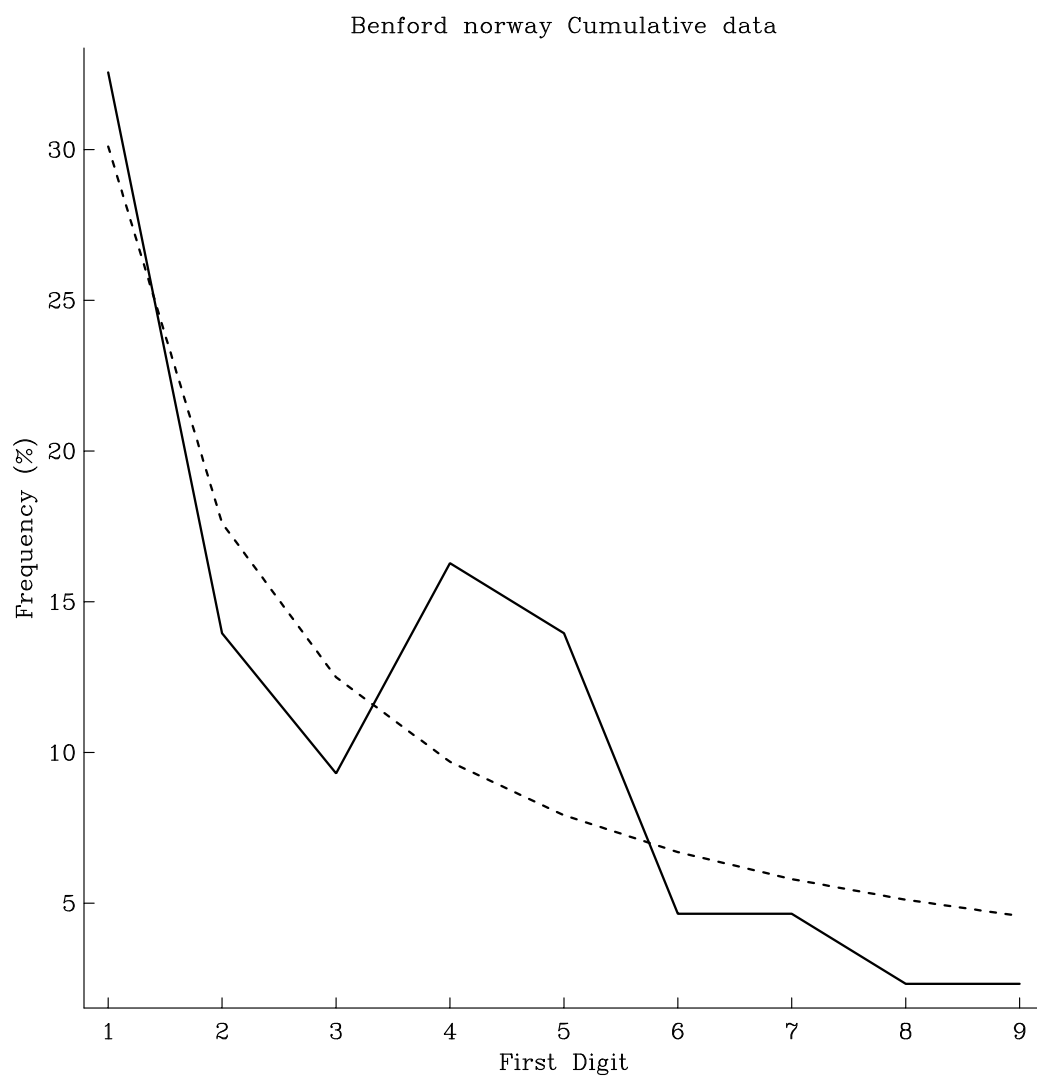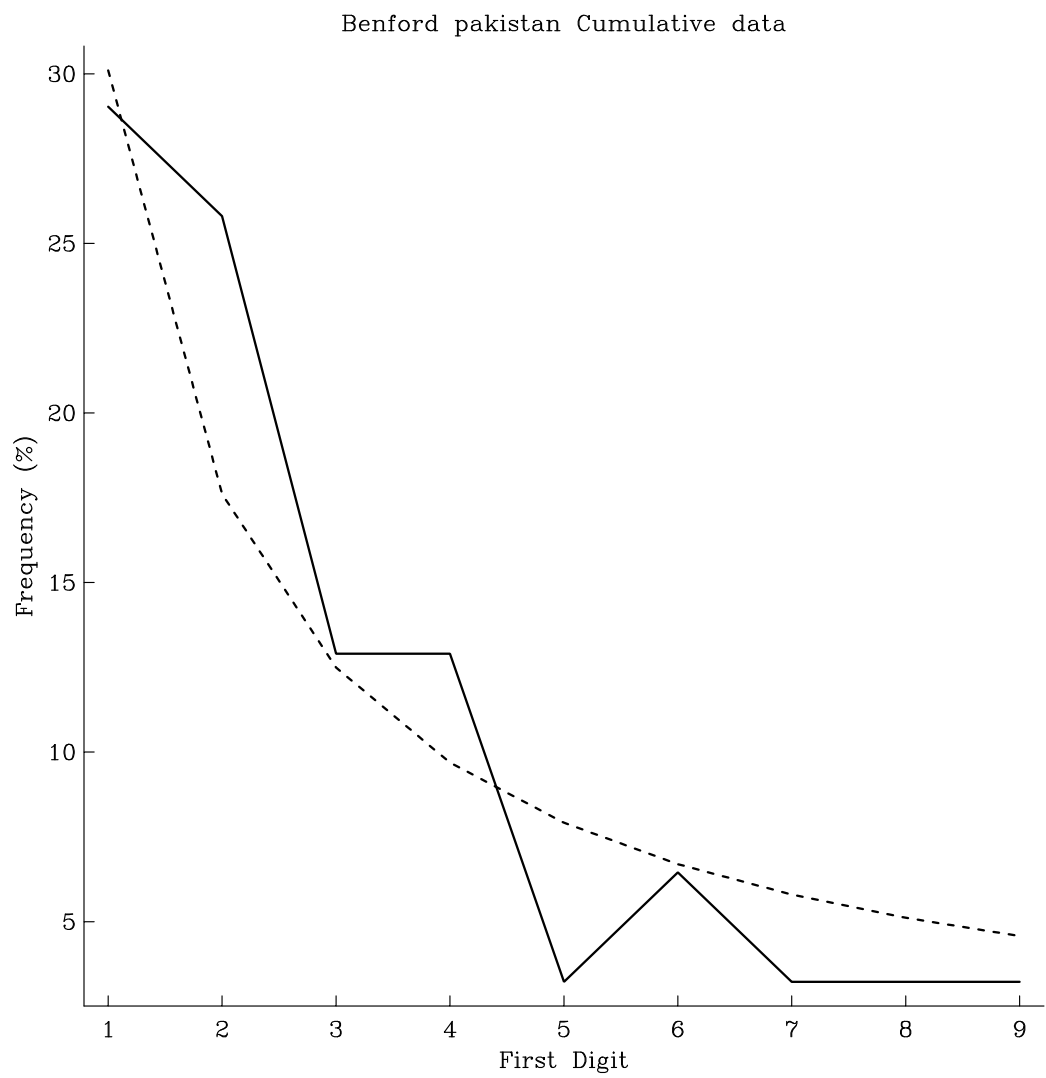
Benford poland Cumulative data

Benford portugal Cumulative data

Benford qatar Cumulative data

Benford romania Cumulative data

Benford russia Cumulative data
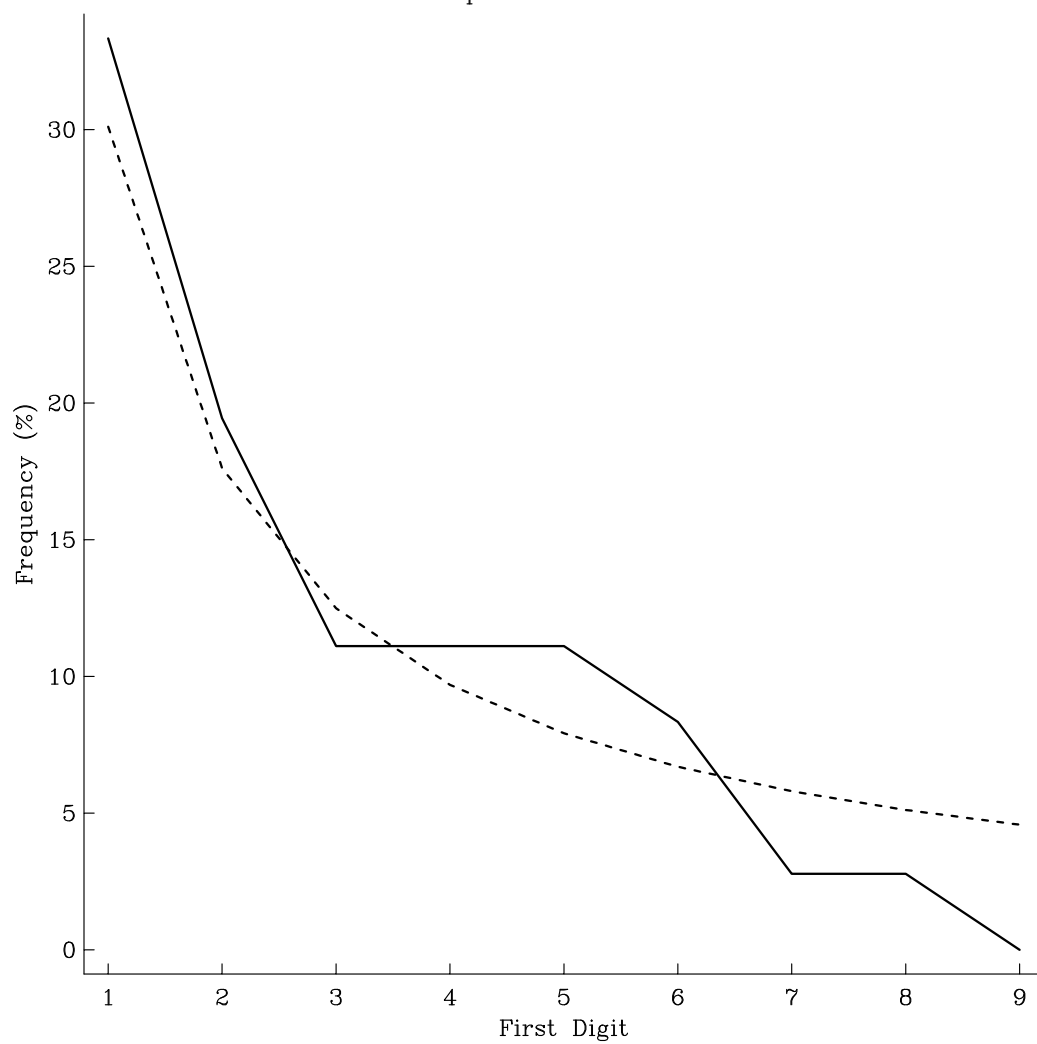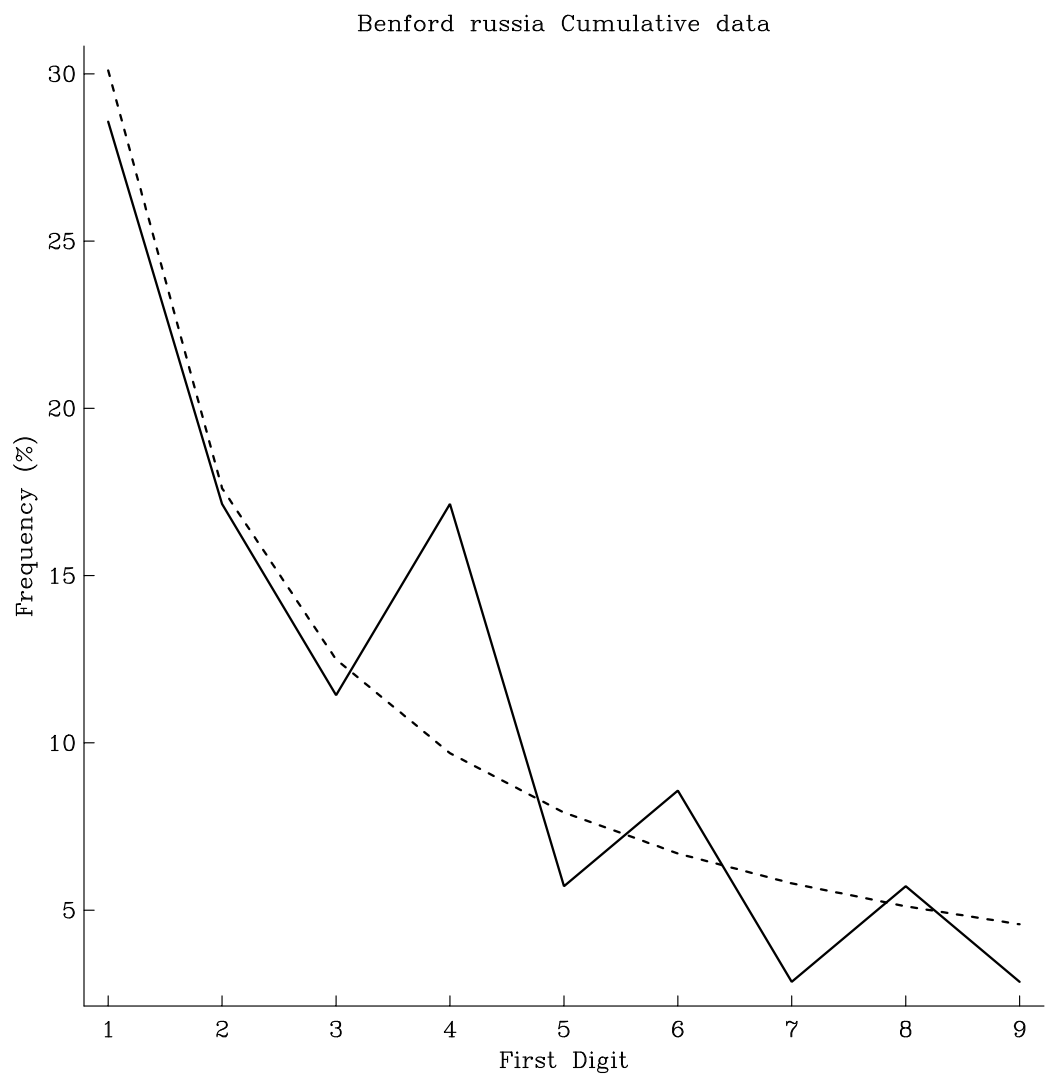
Benford saudiarabia Cumulative data

Benford singapore Cumulative data

Benford slovakia Cumulative data

Benford slovenia Cumulative data

Benford southkorea_pre_mar_5 Cumulative data

Benford spain Cumulative data
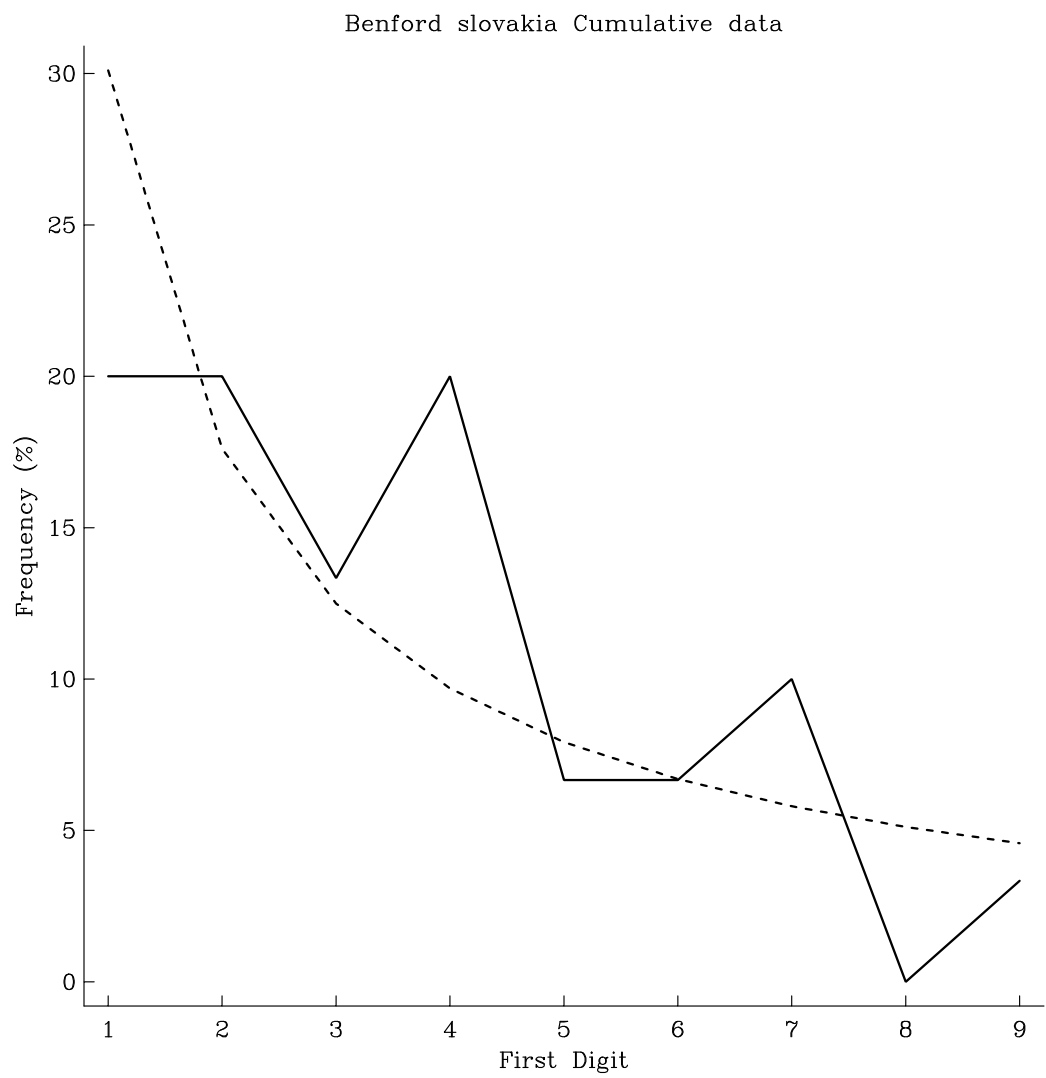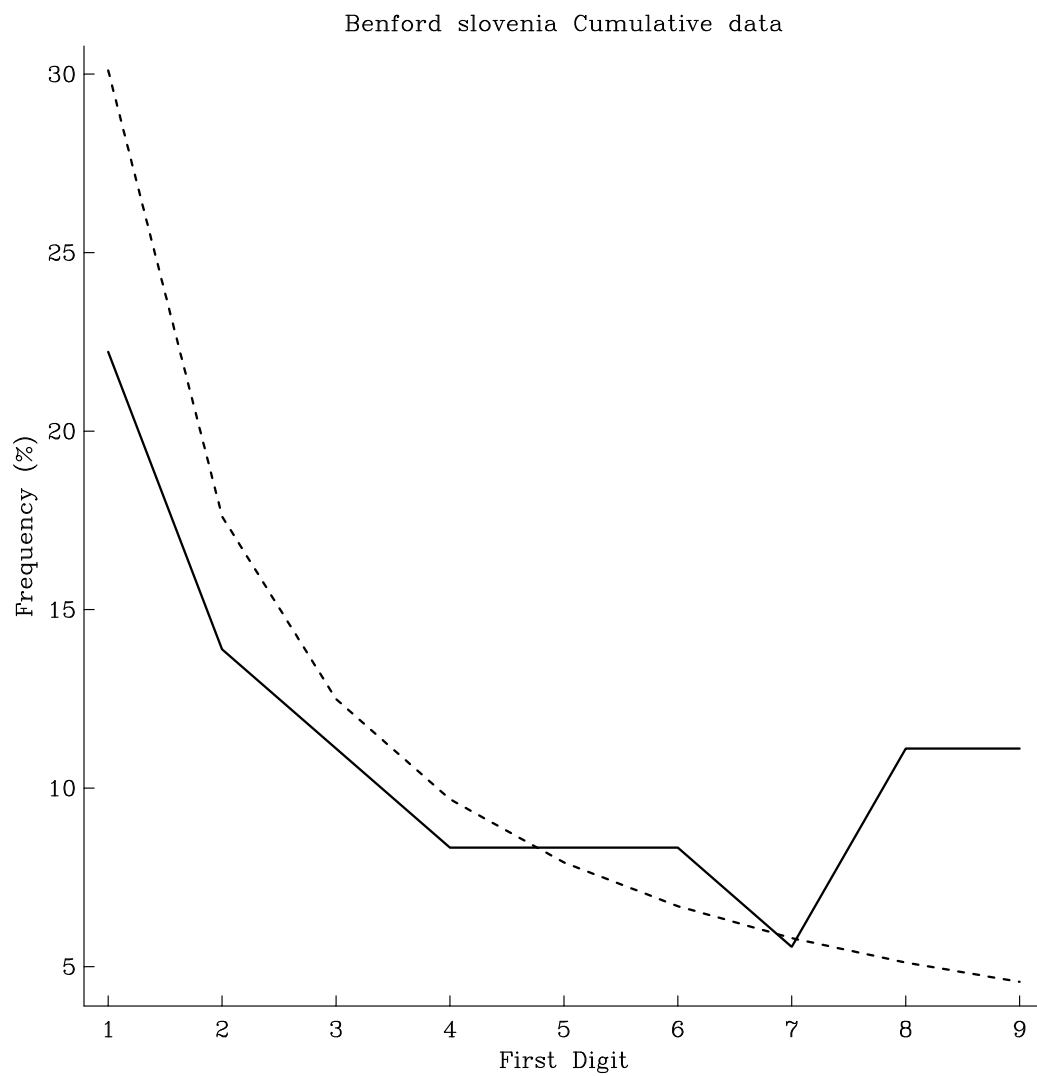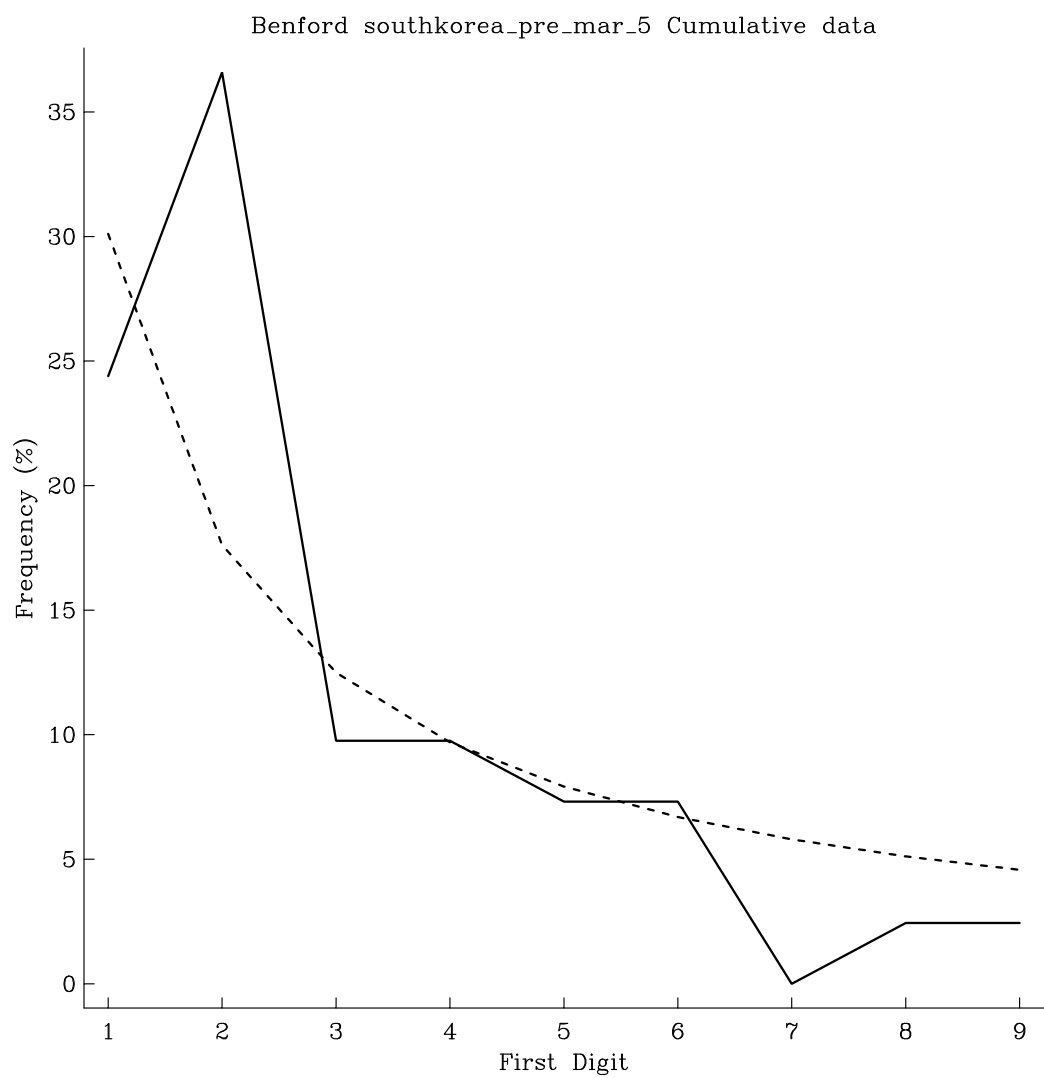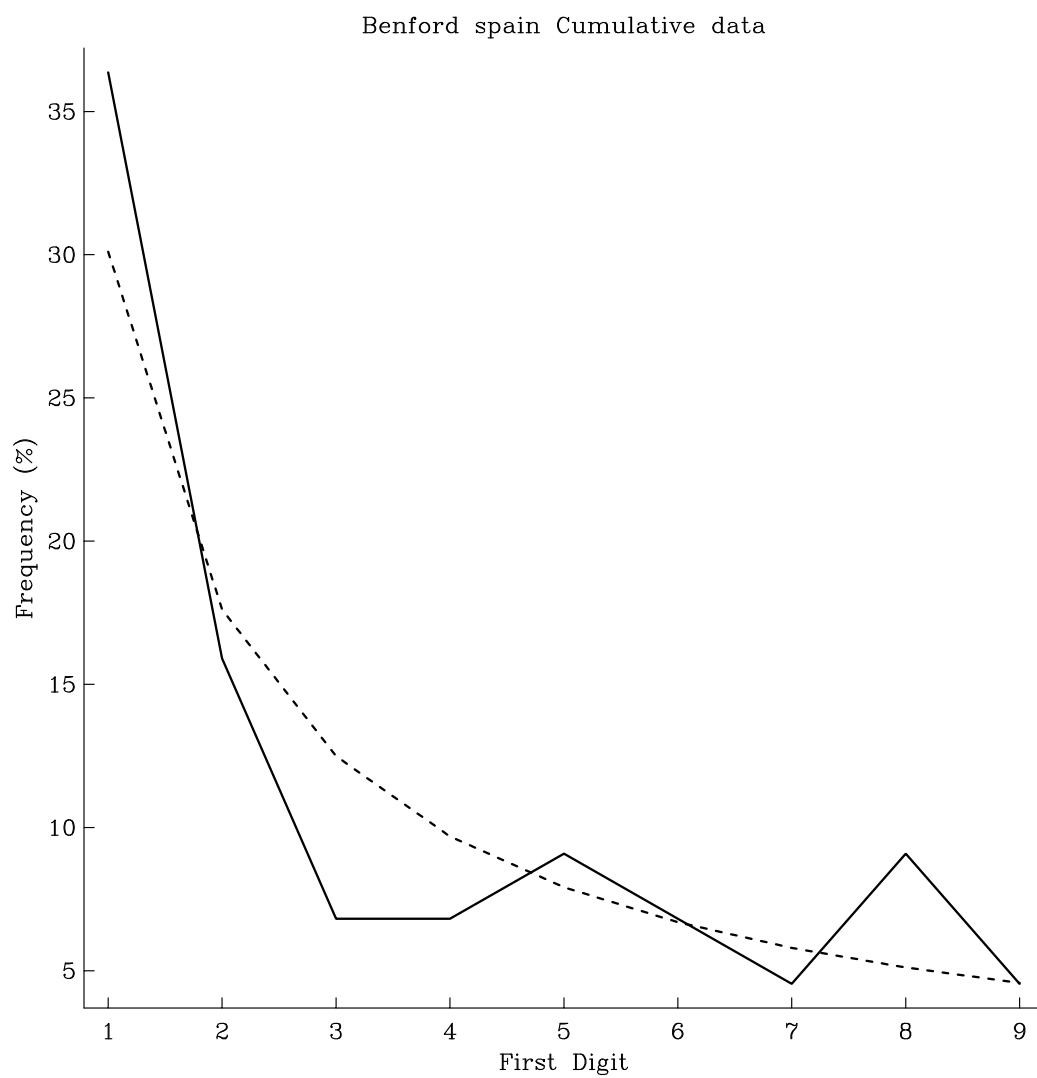
Benford sweden Cumulative data

Benford switzerland Cumulative data

Benford thailand Cumulative data

First Digit

Benford turkey Cumulative data

Benford uk Cumulative data
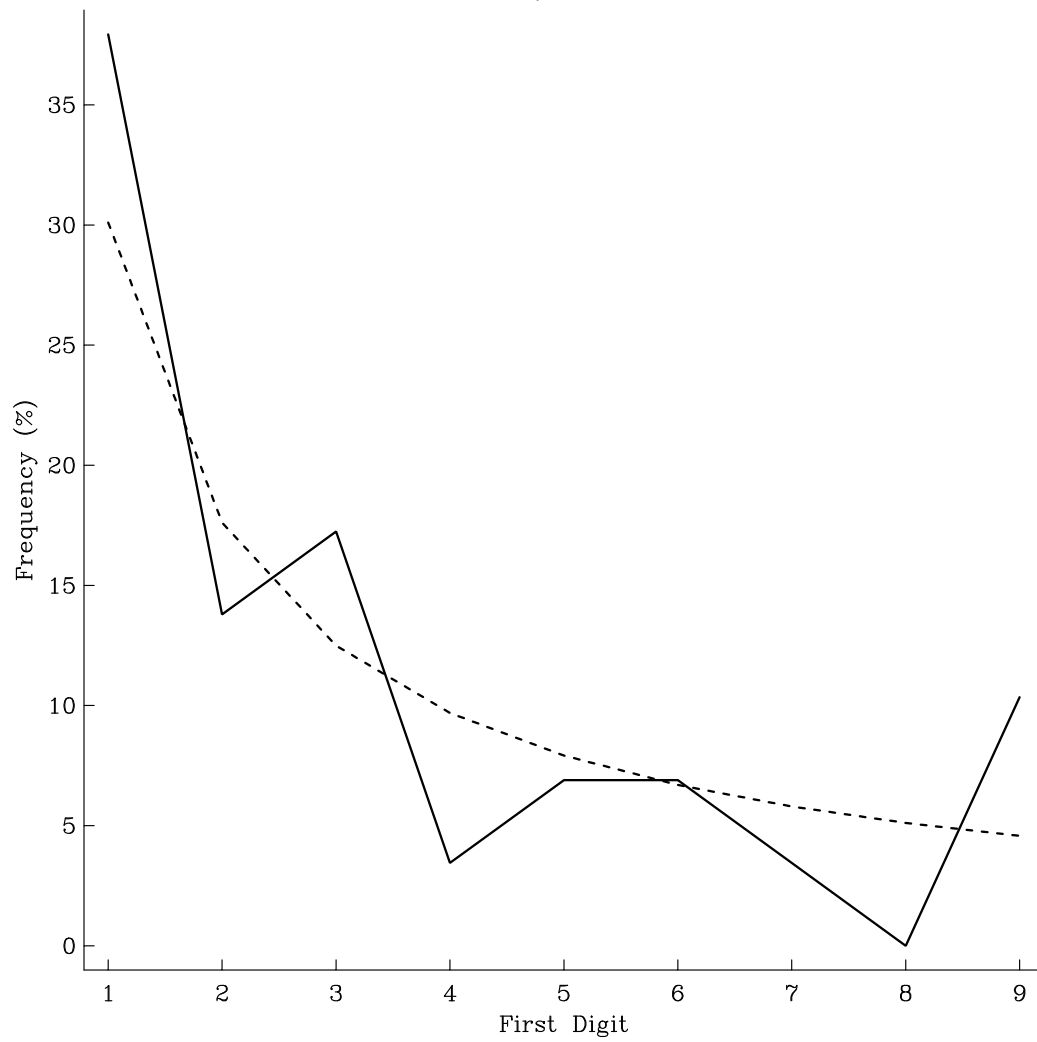
Benford us Cumulative data

Deaths by country

Benford all_countries_death_1col Cumulative data
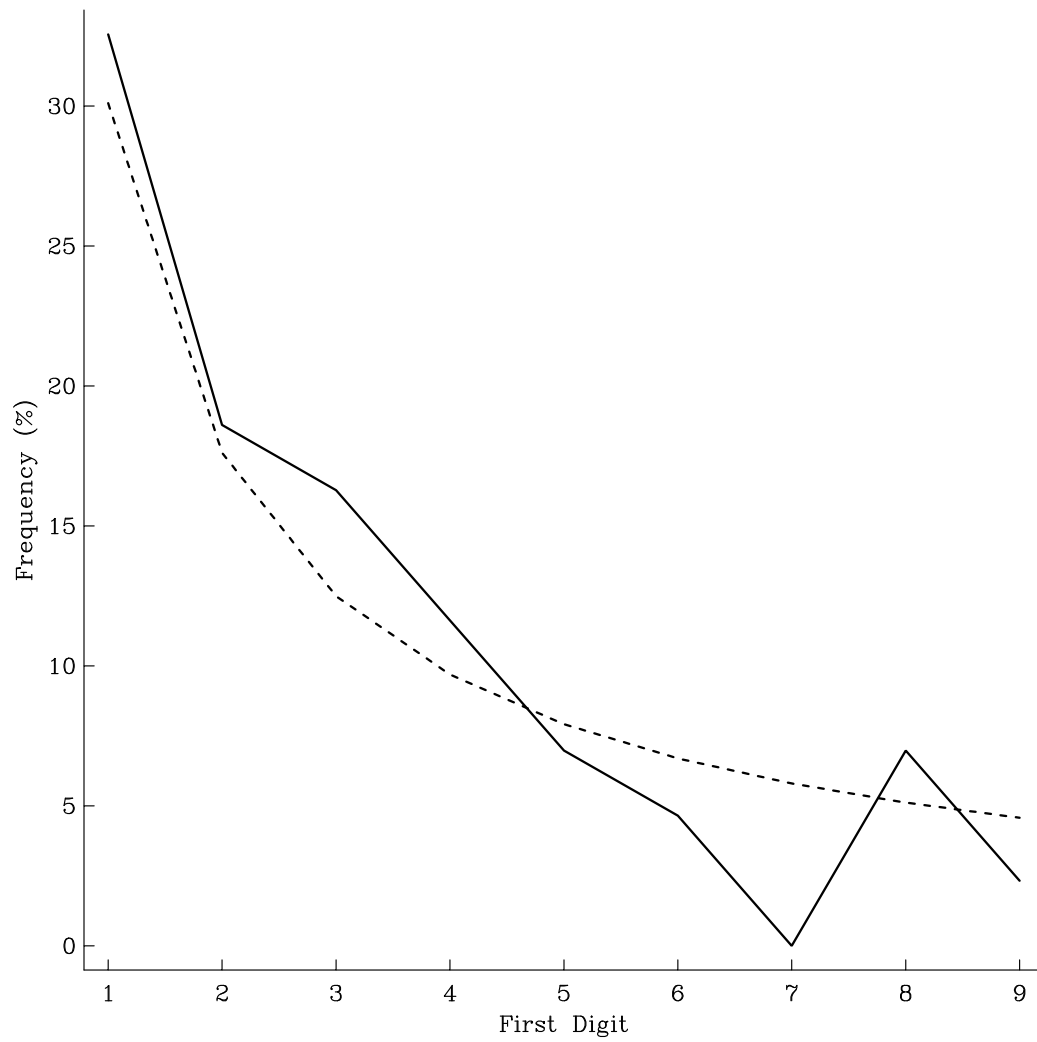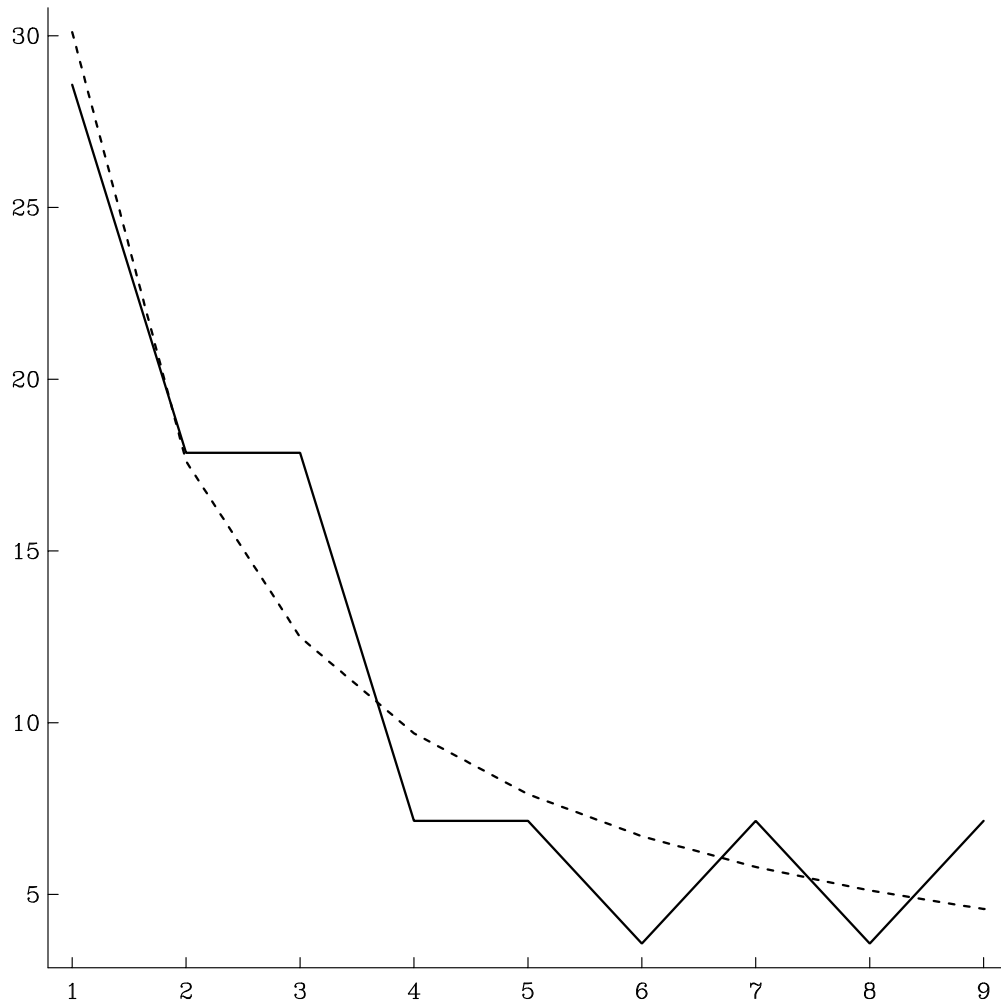
Benford australia.dat Cumulative data
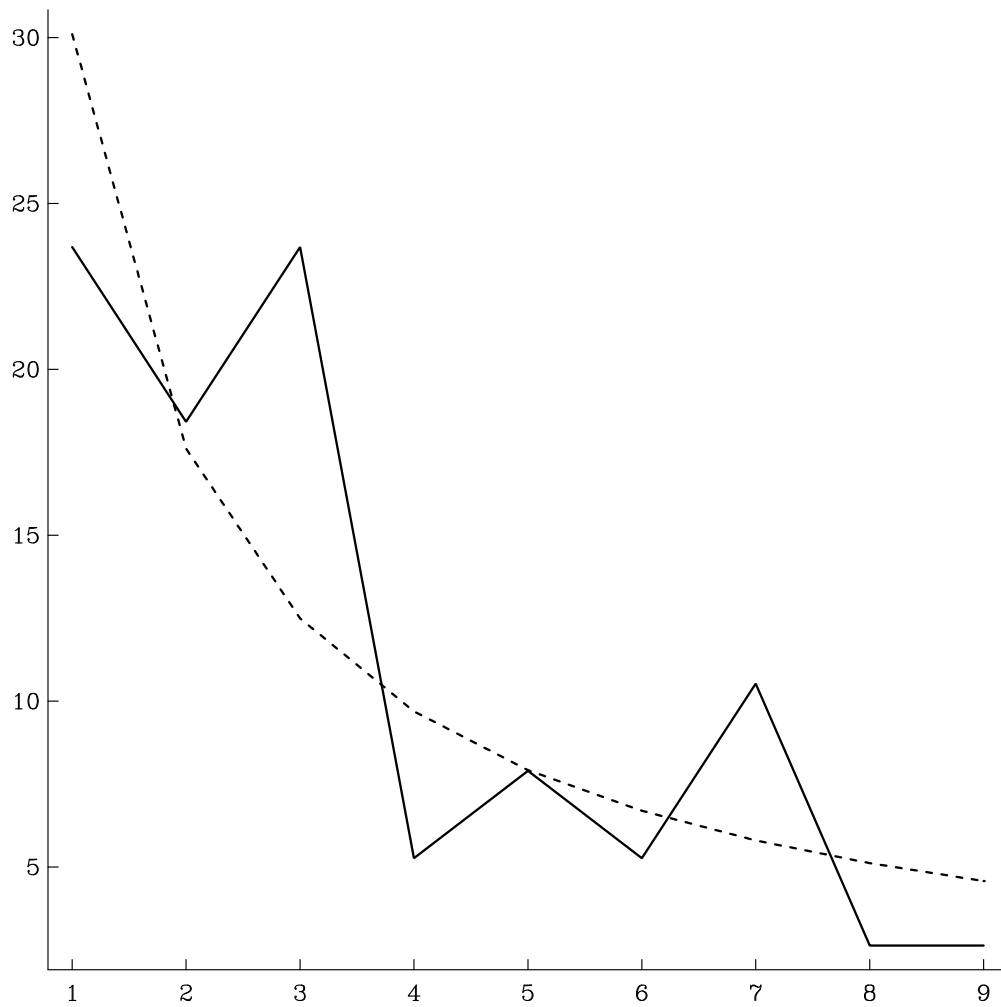
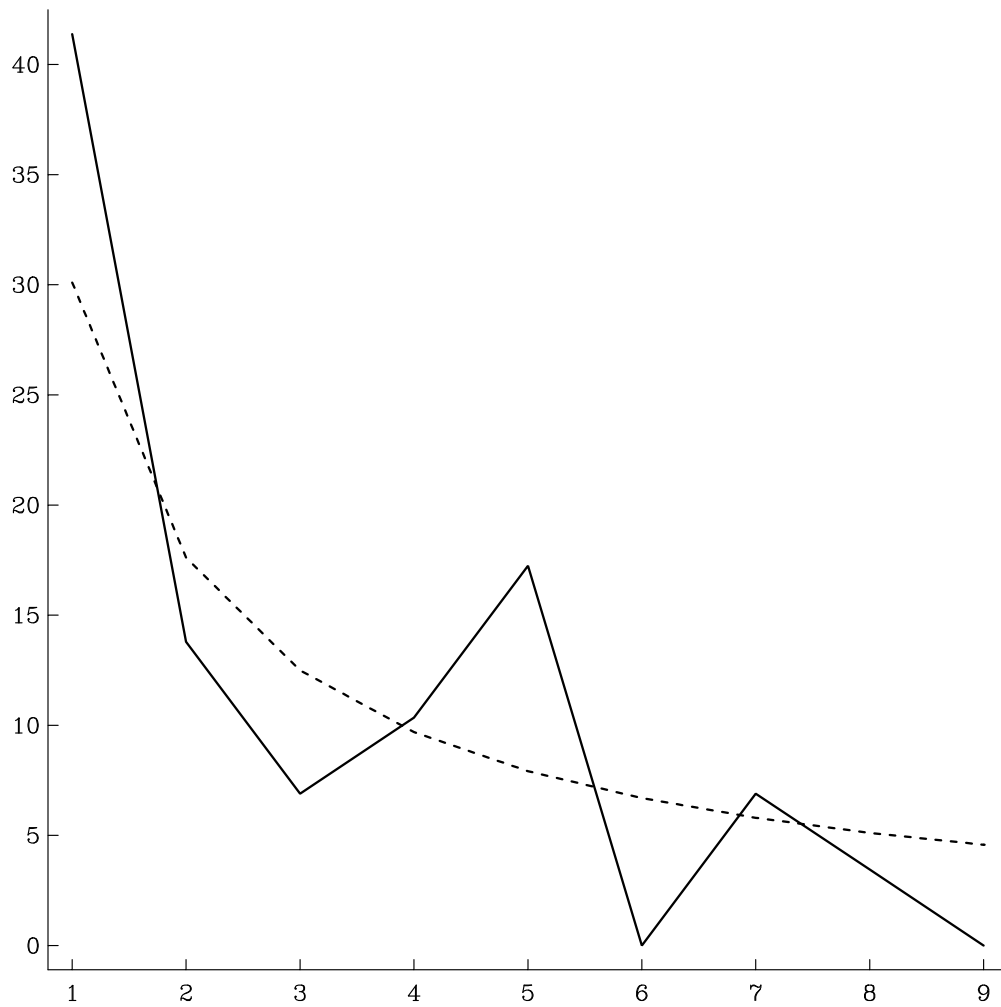Benford austria.dat Cumulative data
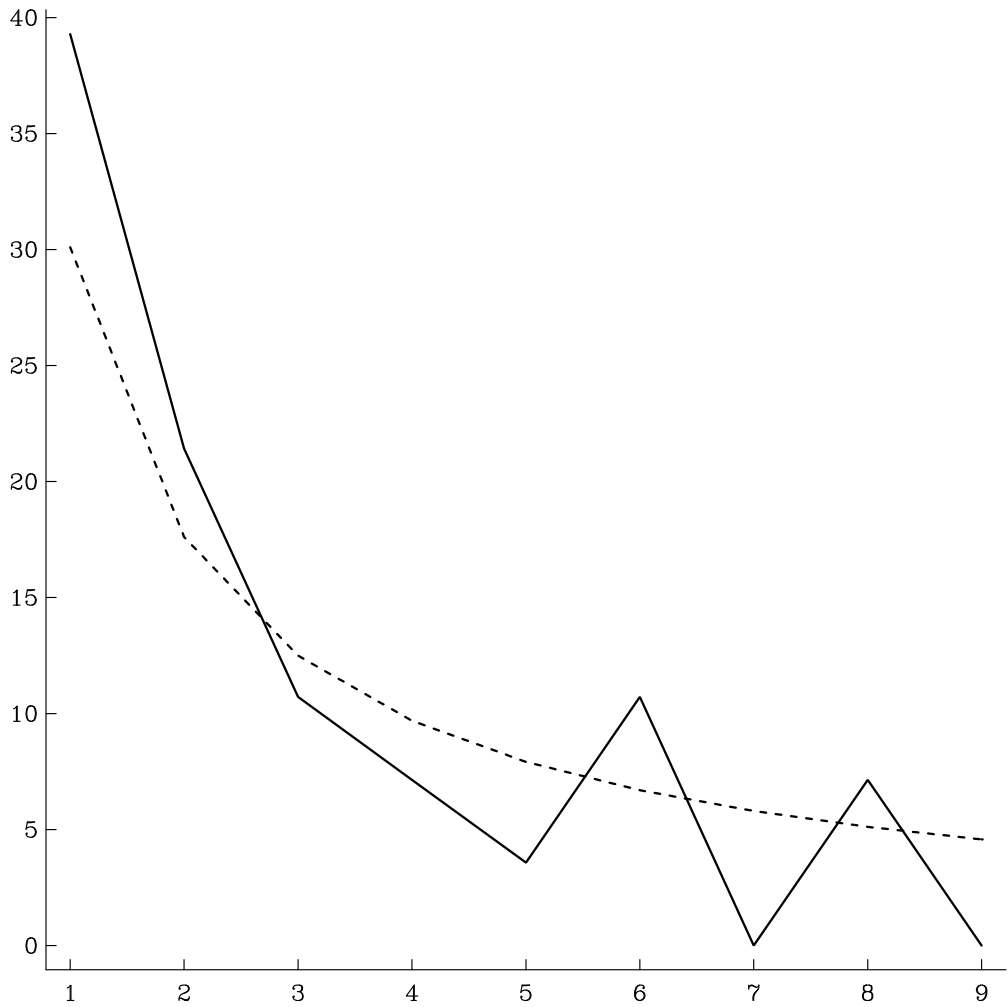
Benford belgium.dat Cumulative data
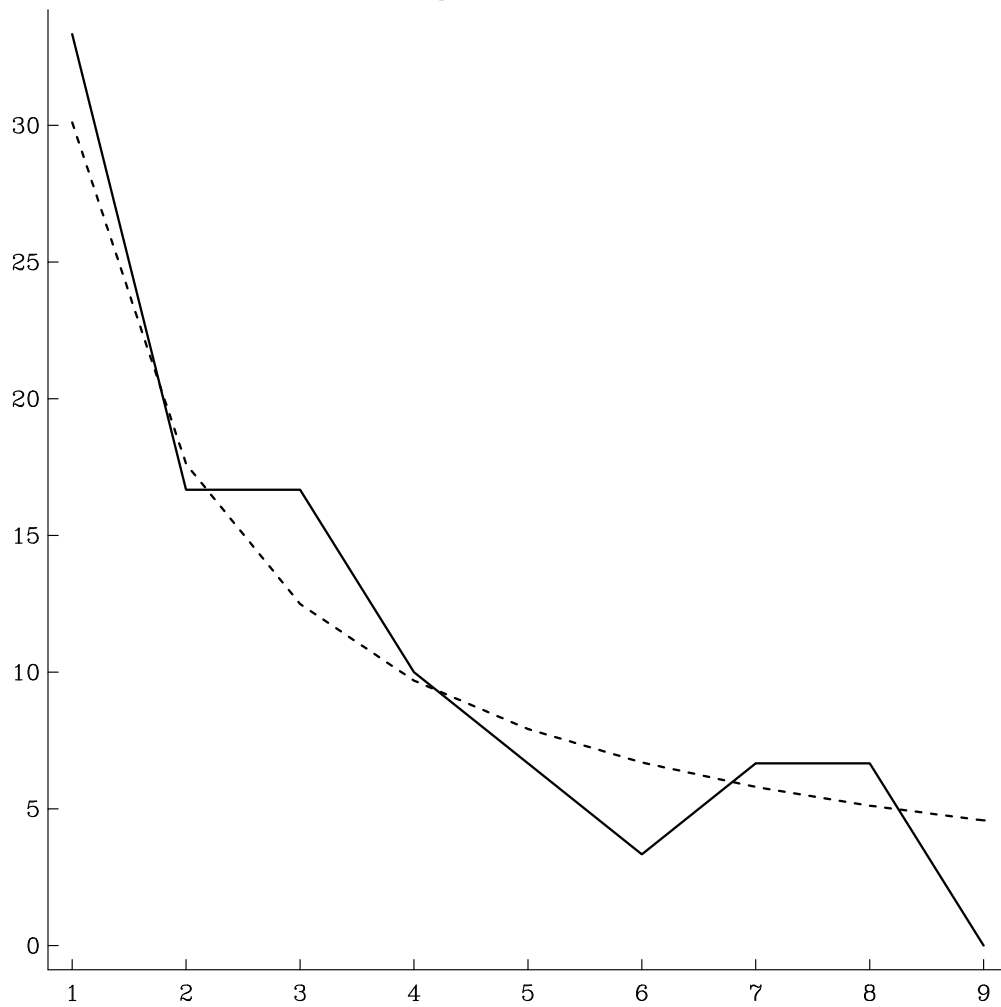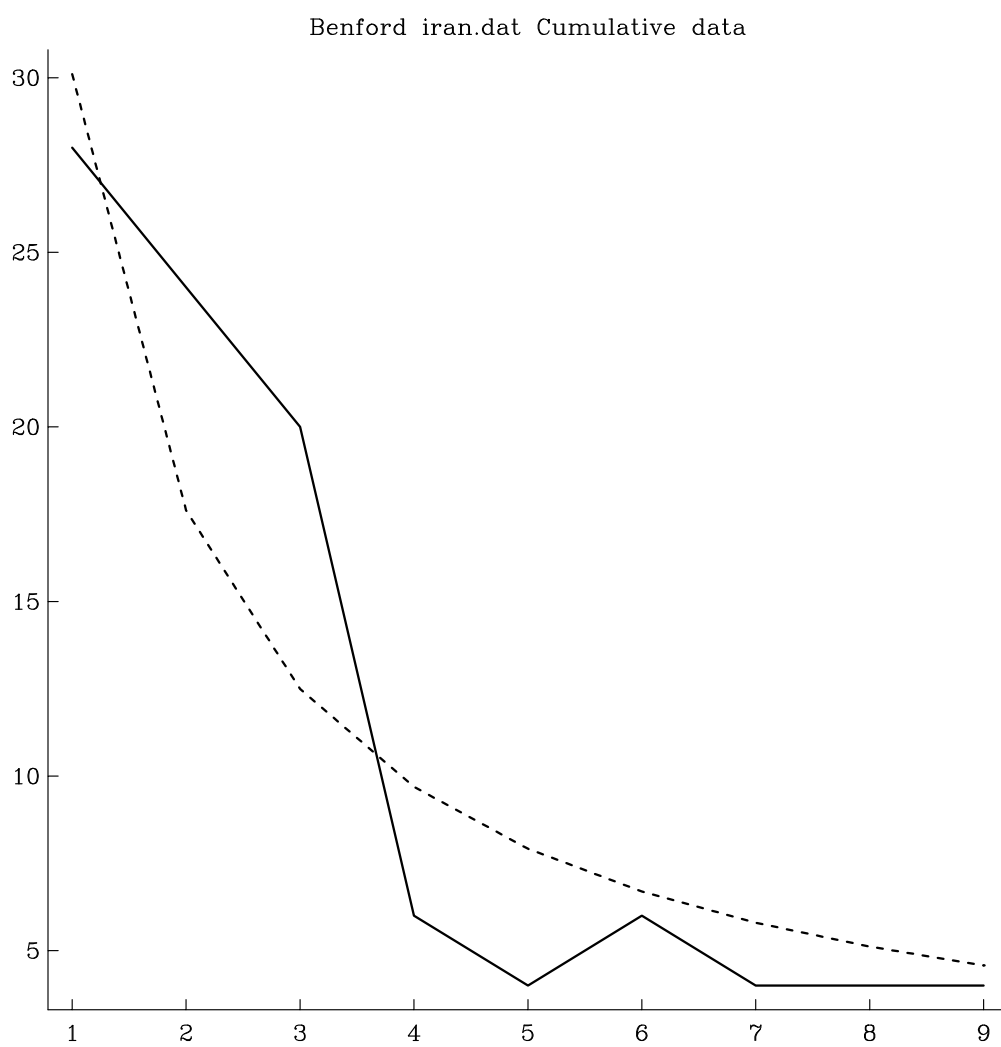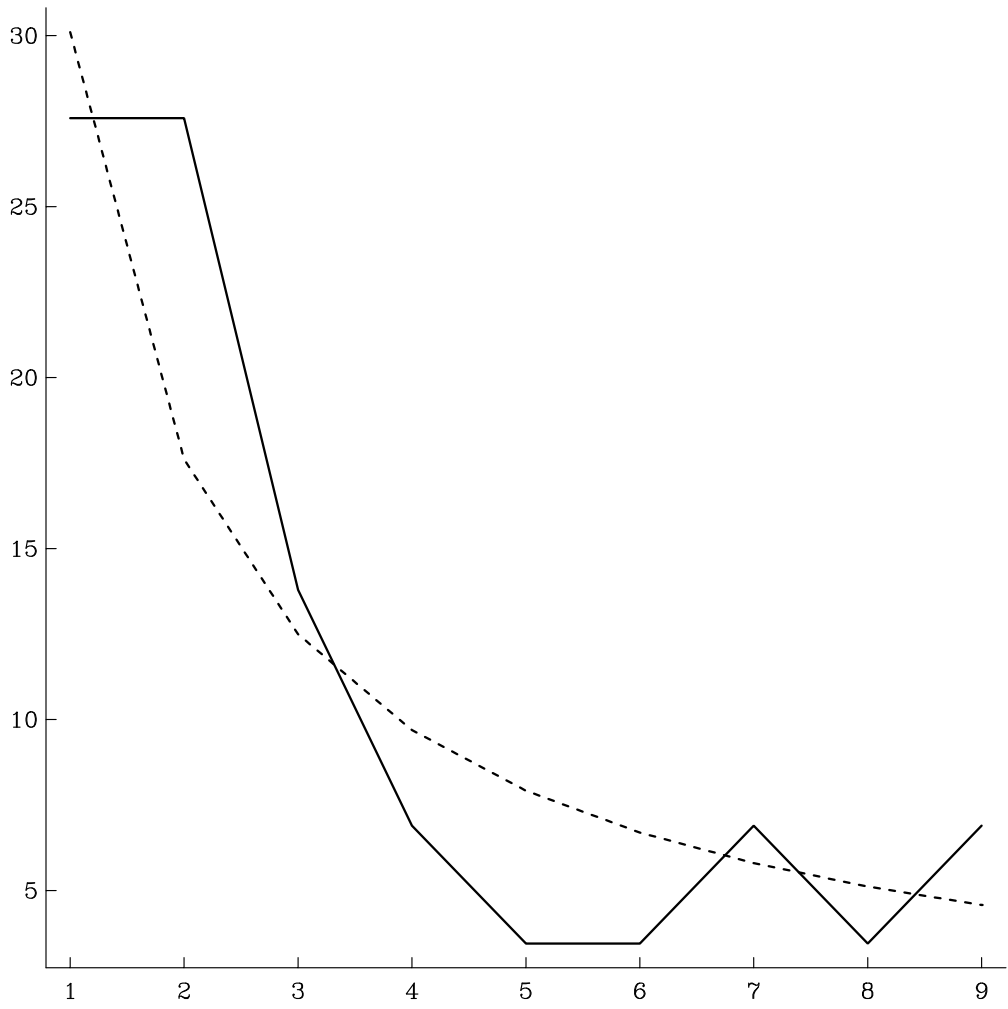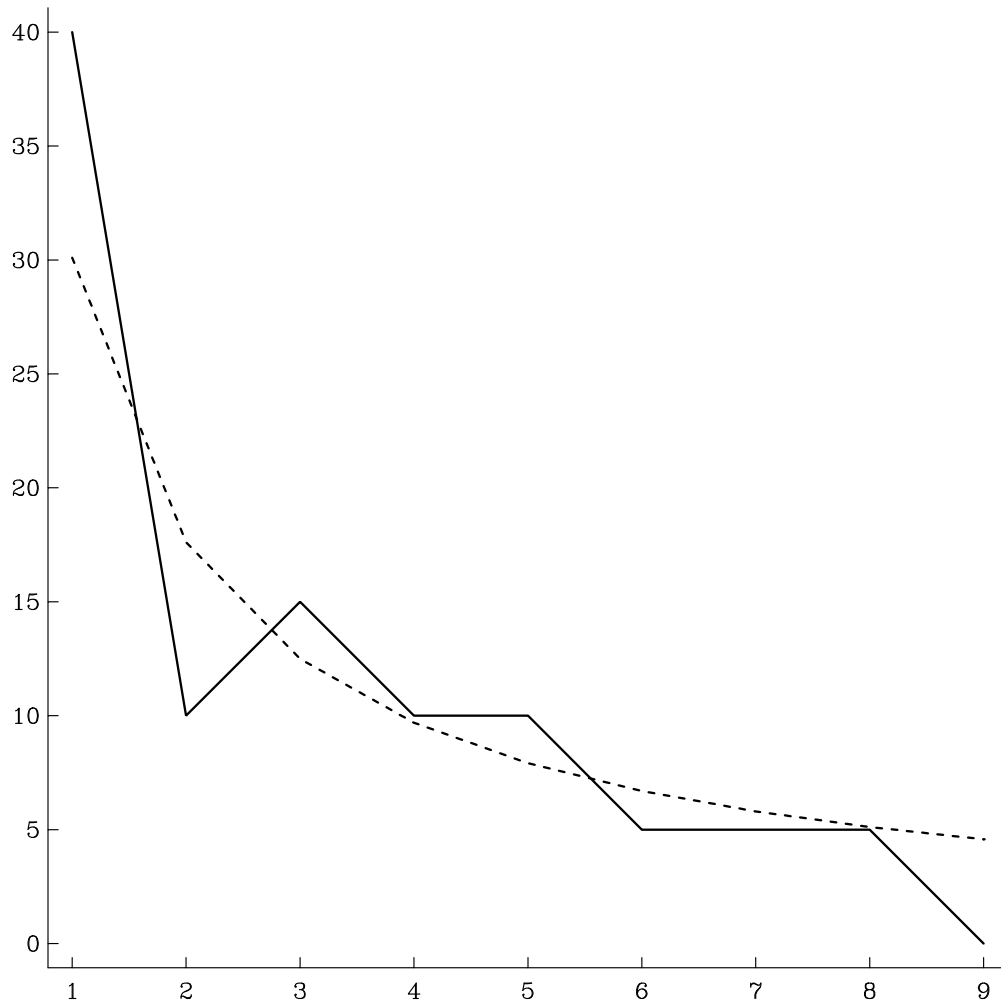
Benford brazil.dat Cumulative data
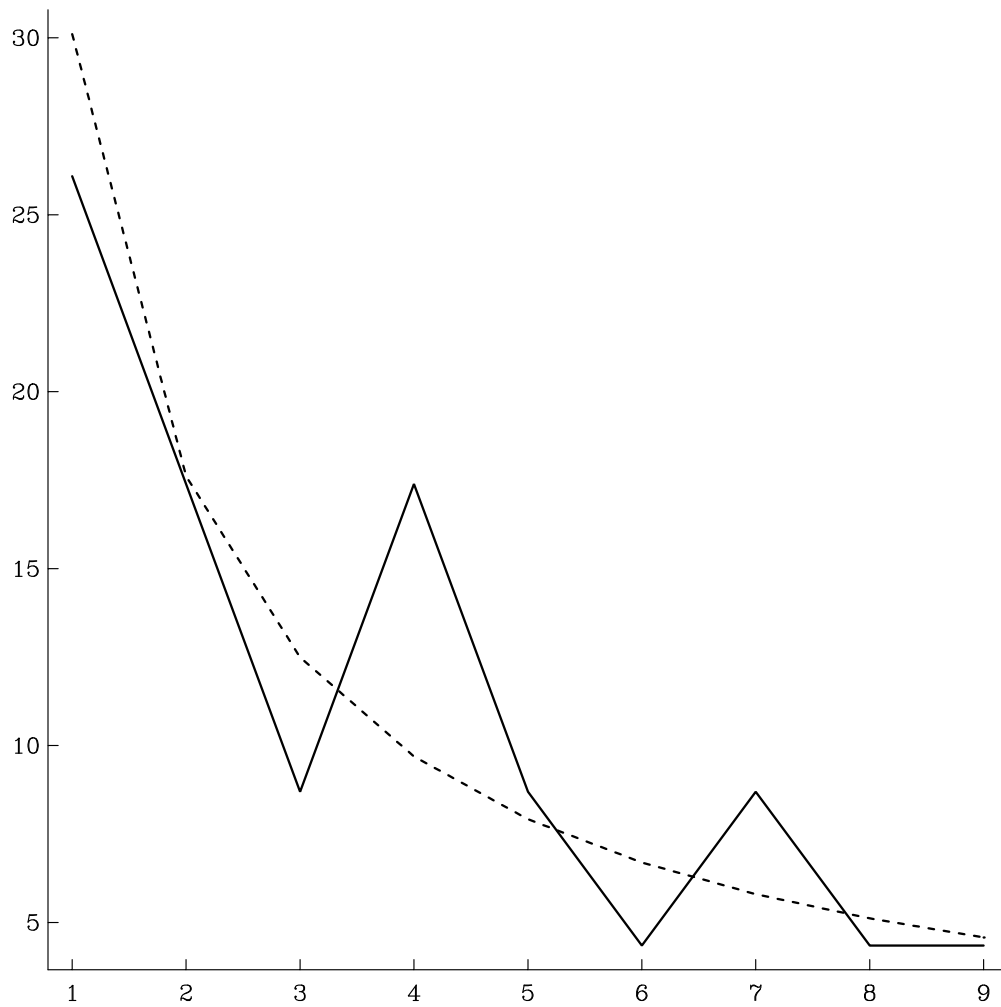
Benford bulgaria.dat Cumulative data
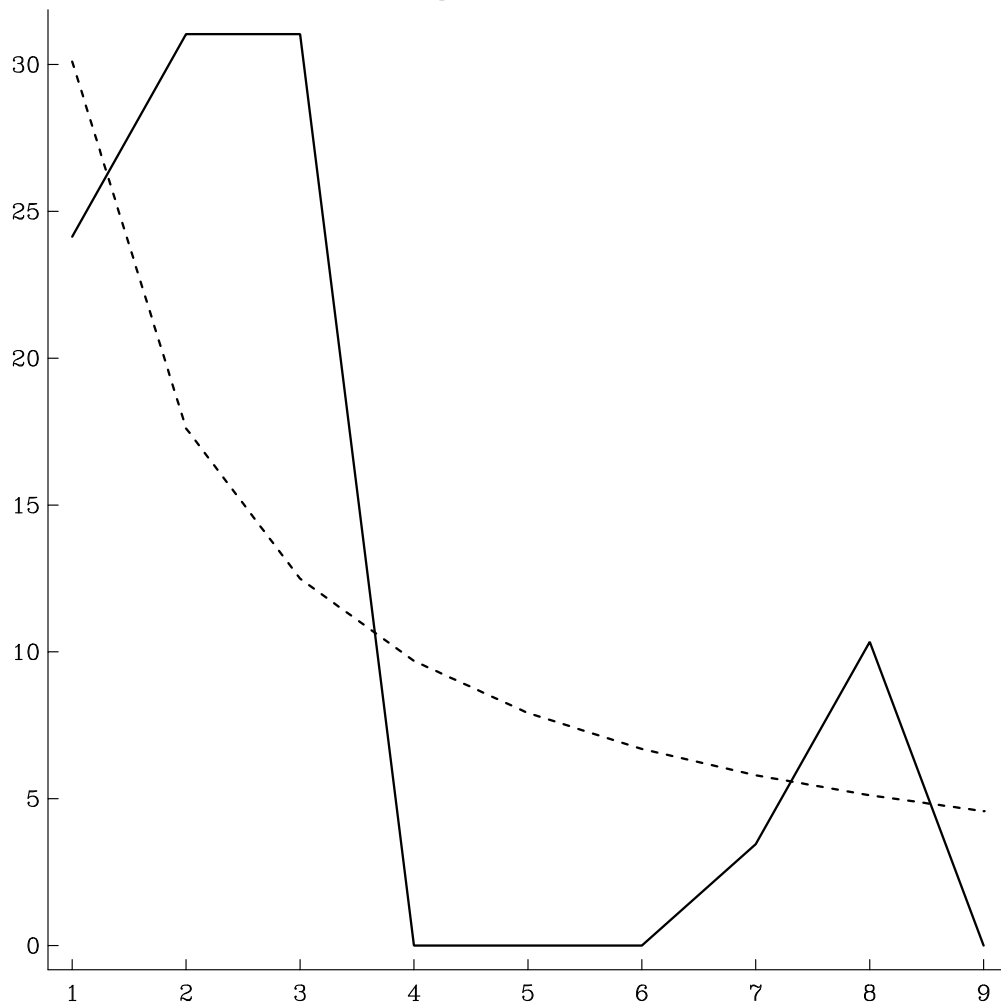
Benford canada.dat Cumulative data
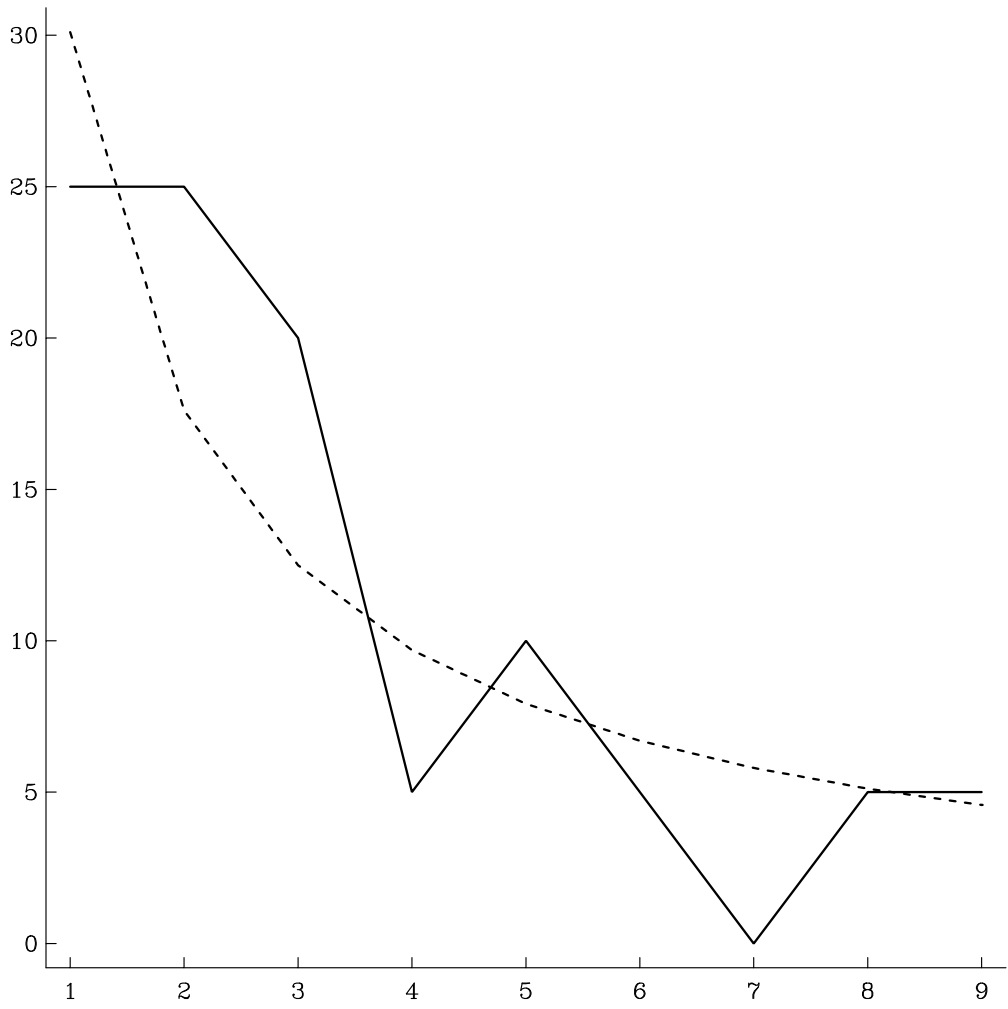
Benford czechrepublic.dat Cumulative data
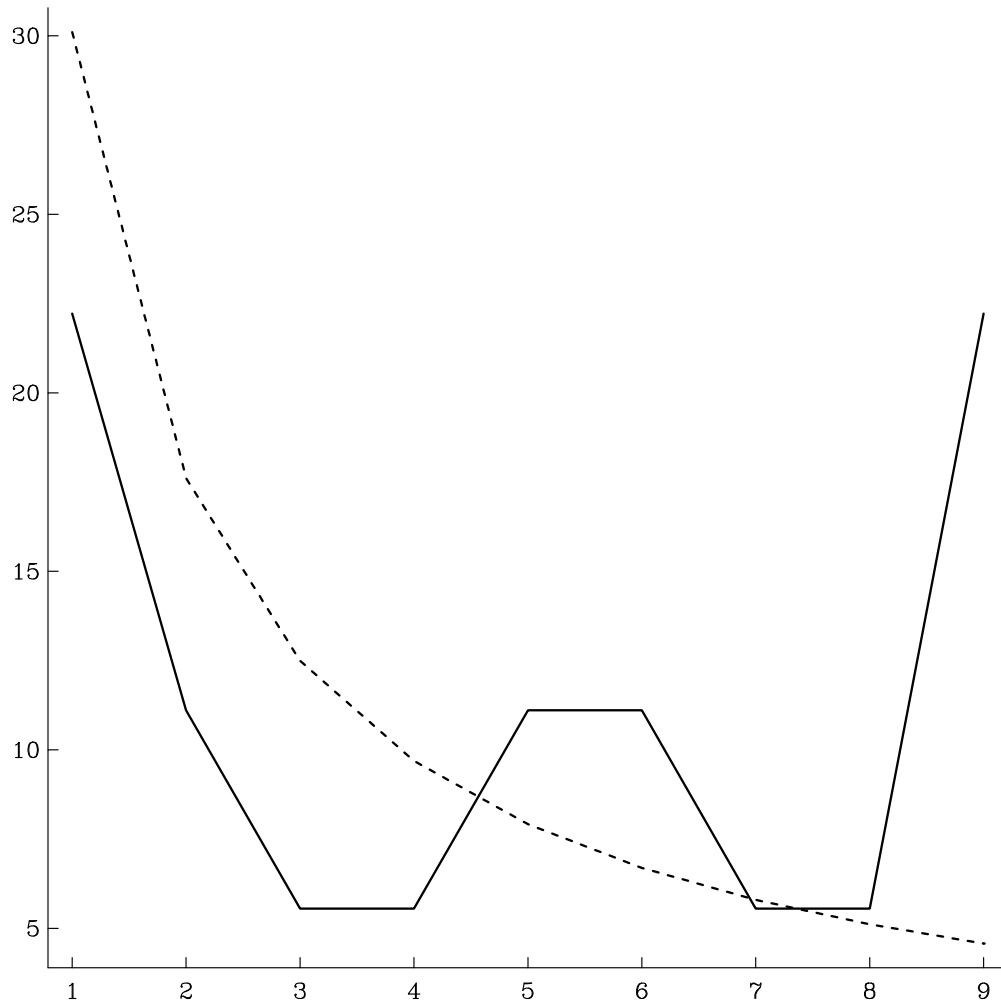
Benford denmark.dat Cumulative data

Benford egypt.dat Cumulative data
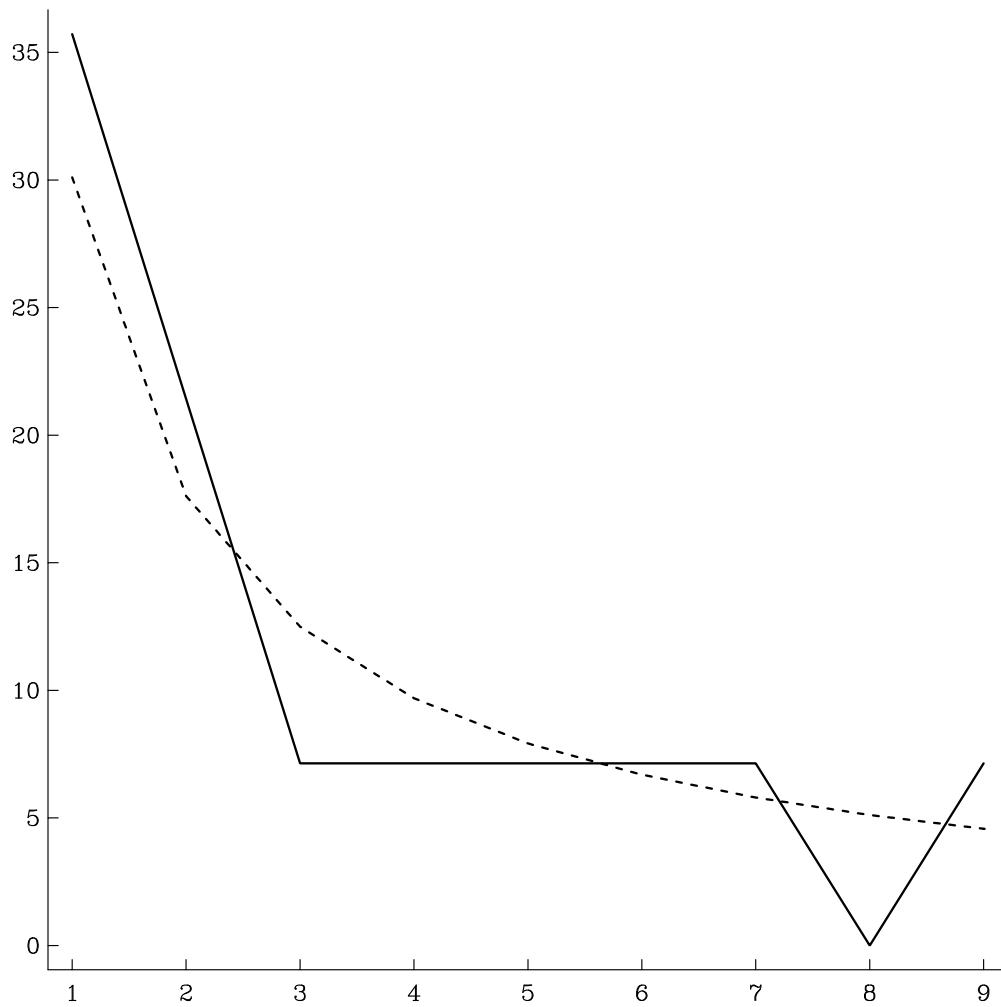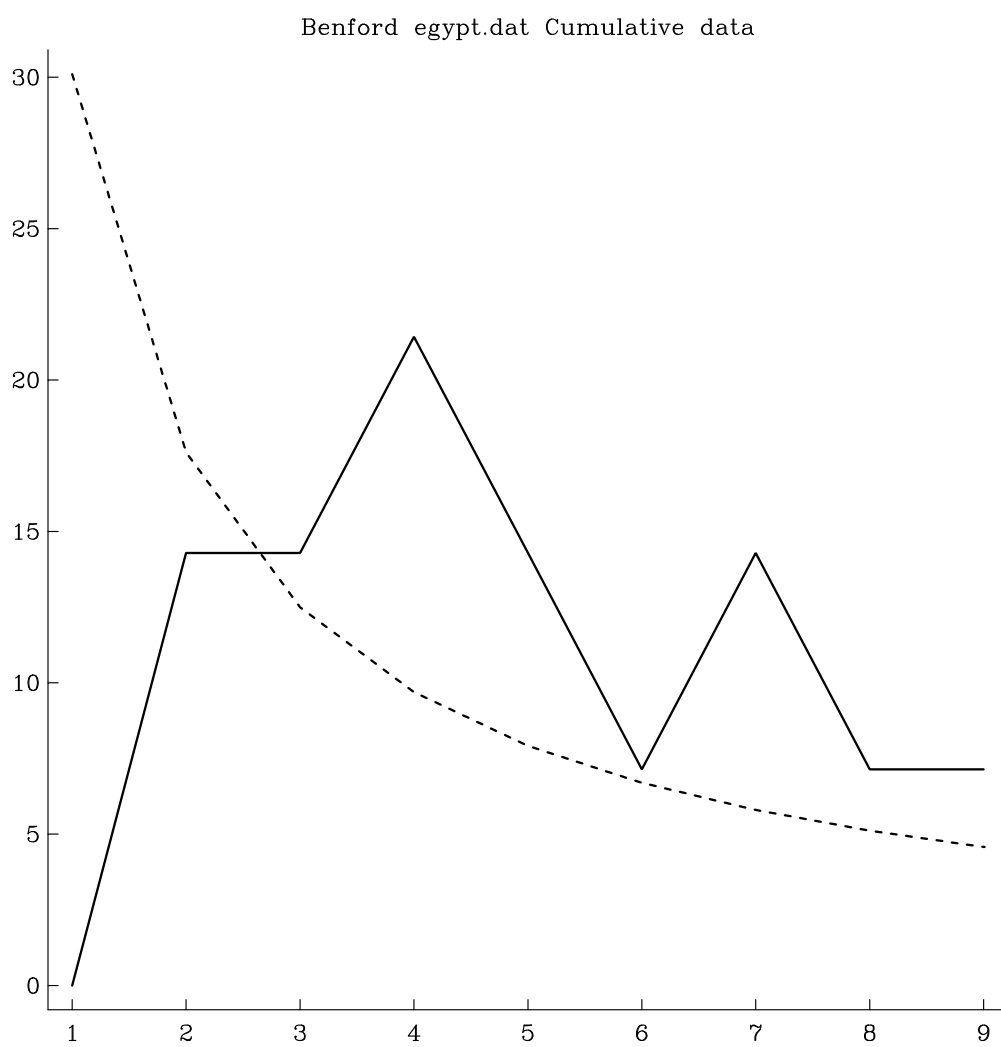
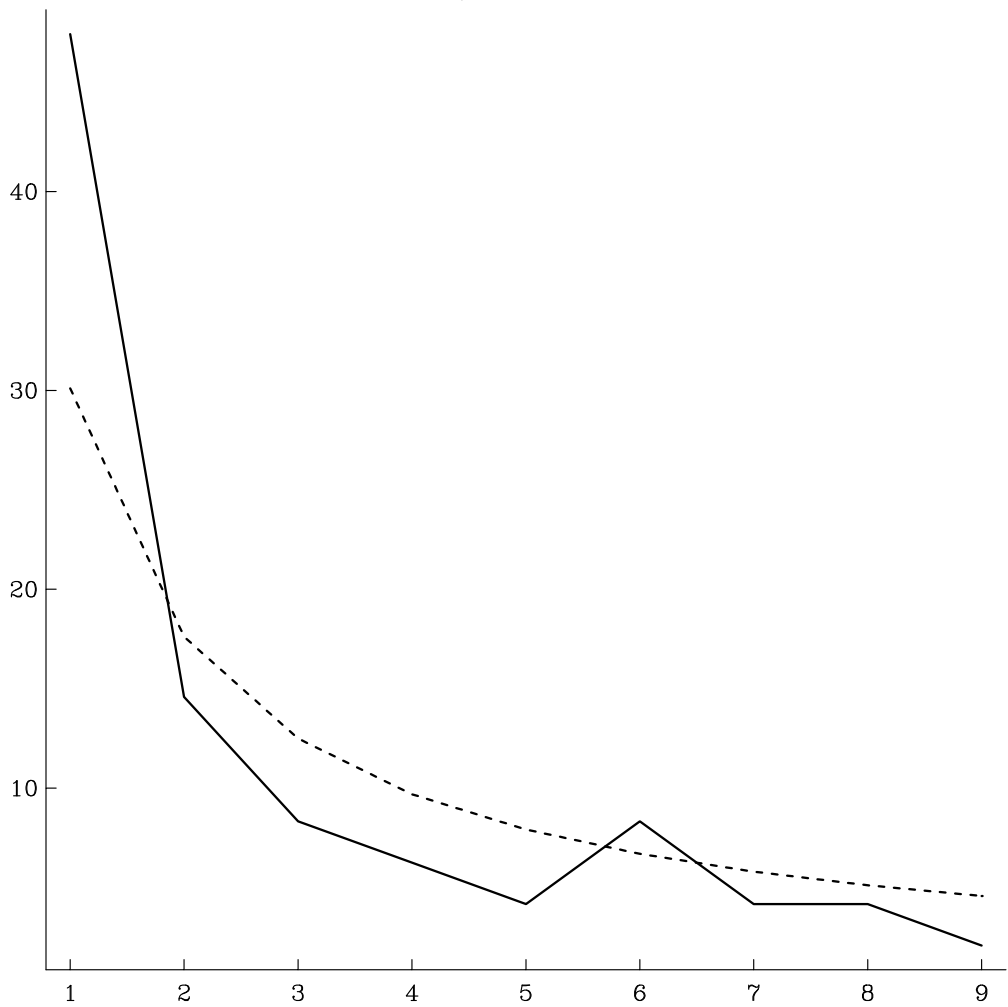Benford estonia.dat Cumulative data
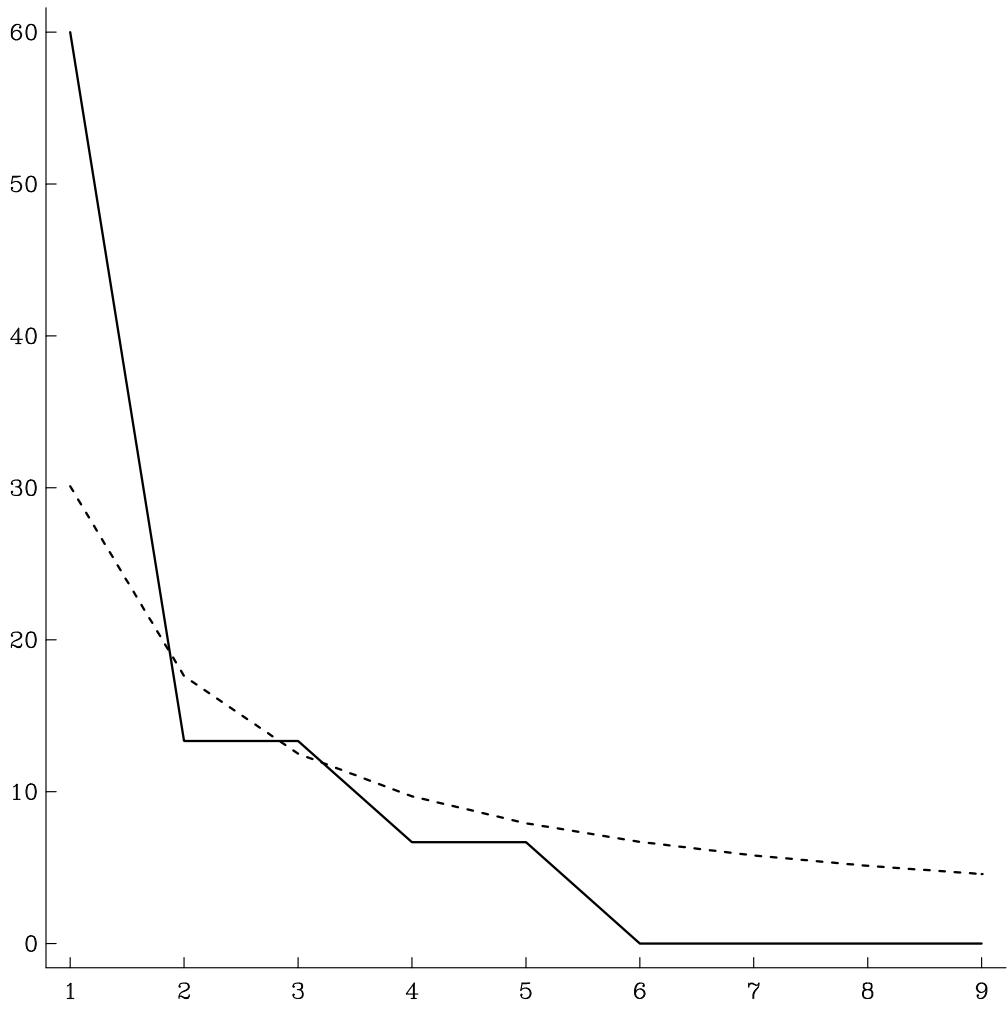
Benford france.dat Cumulative data
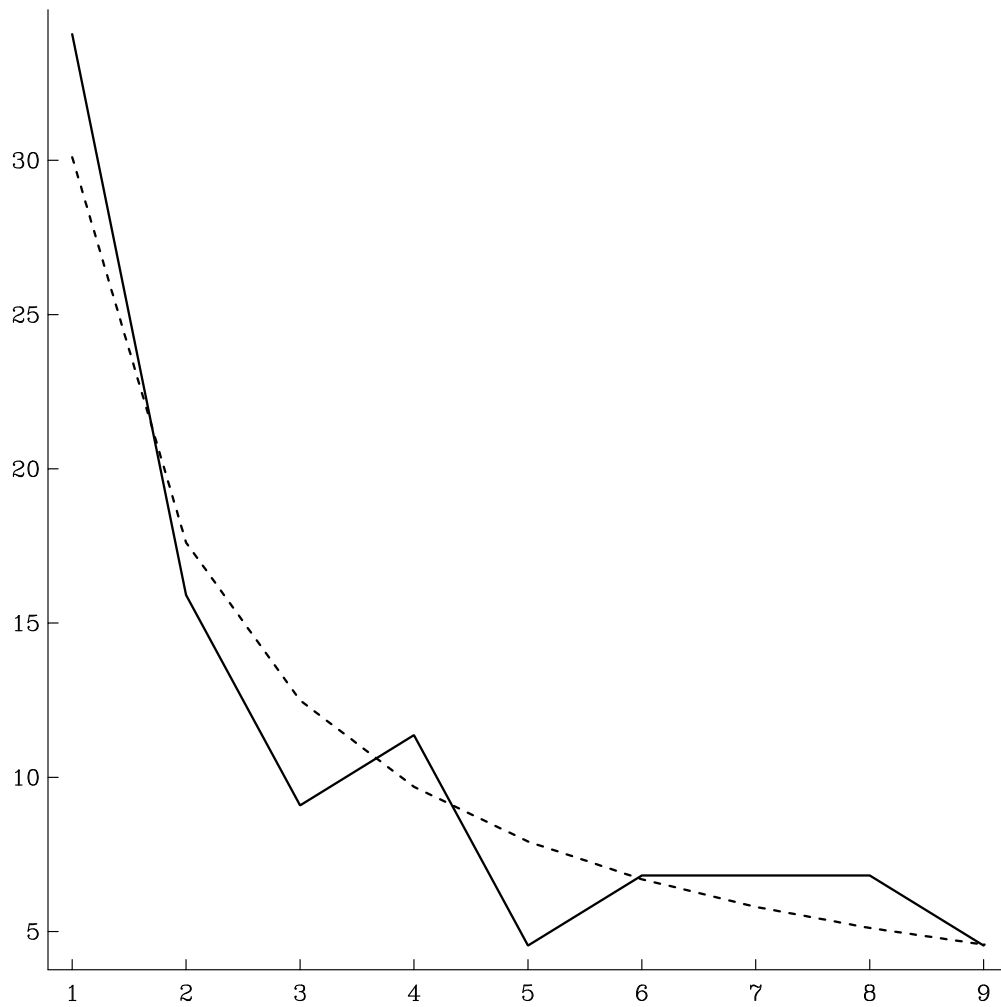
Benford germany.dat Cumulative data
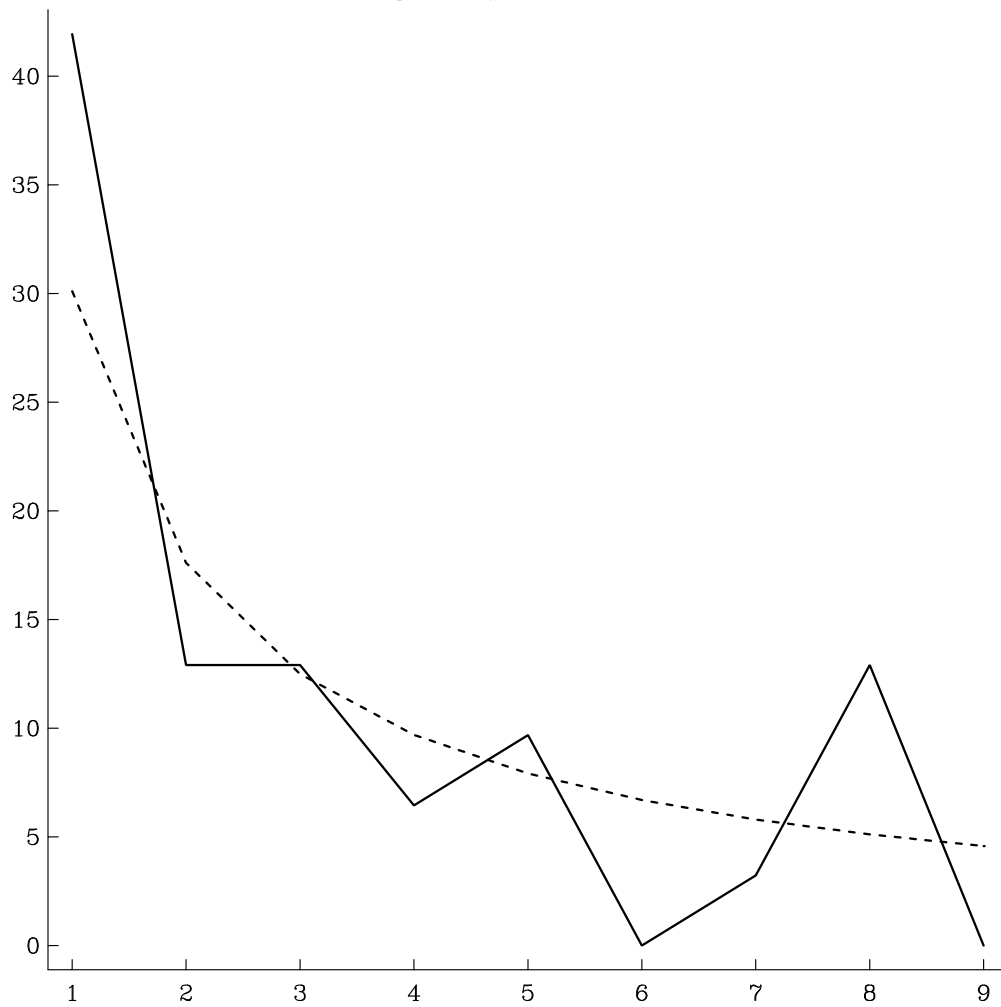
Benford greece.dat Cumulative data
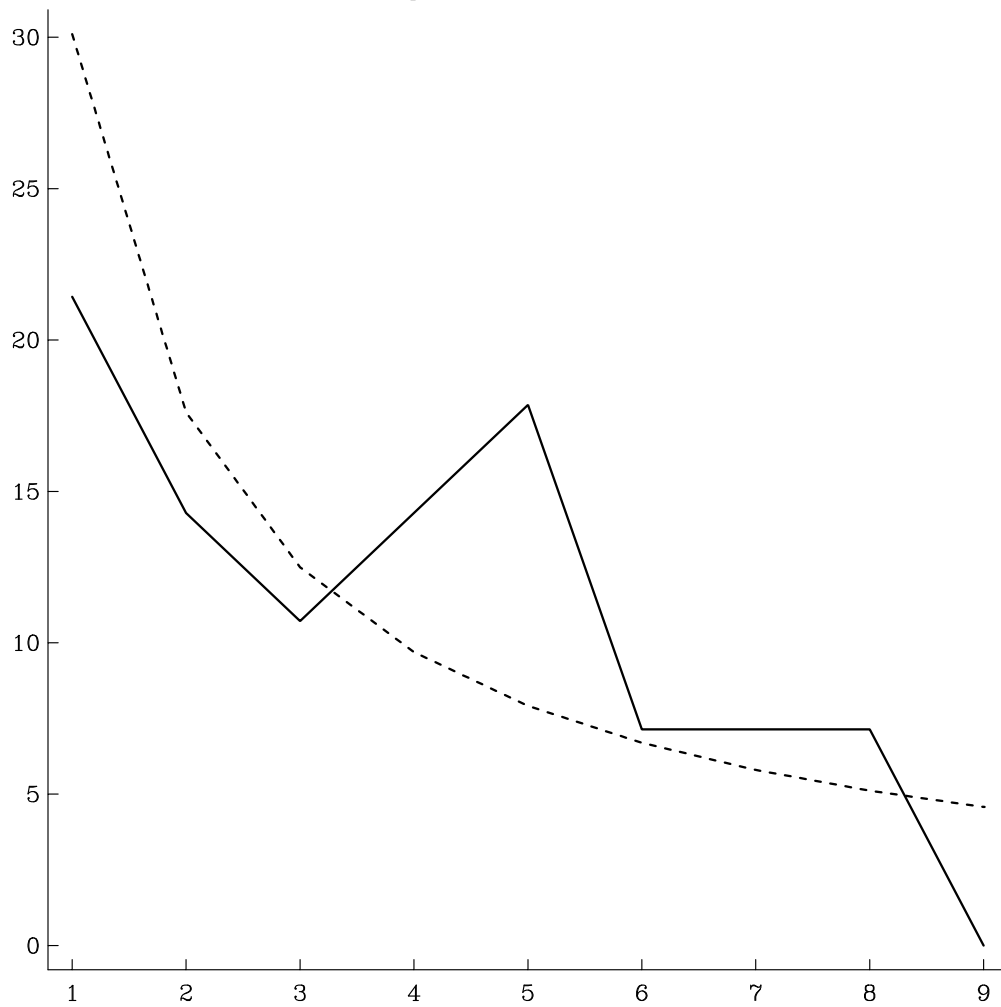
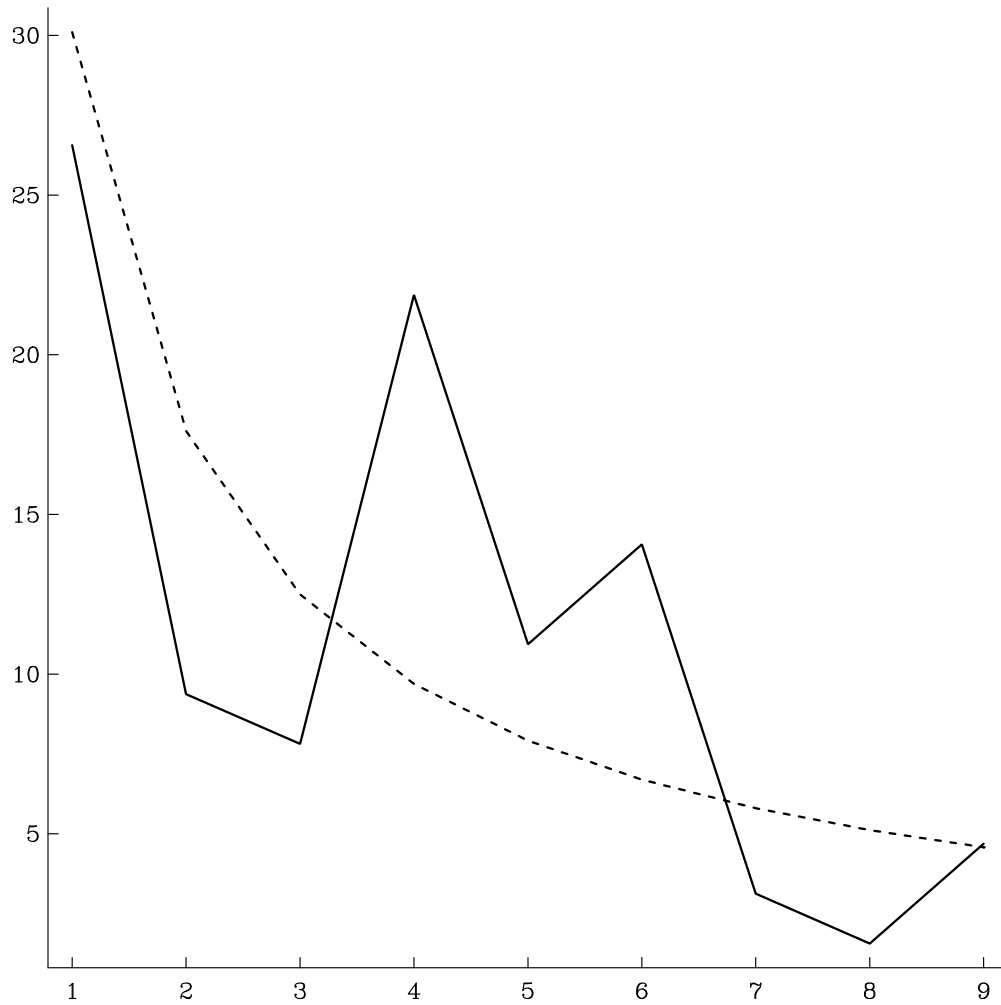Benford hungary.dat Cumulative data
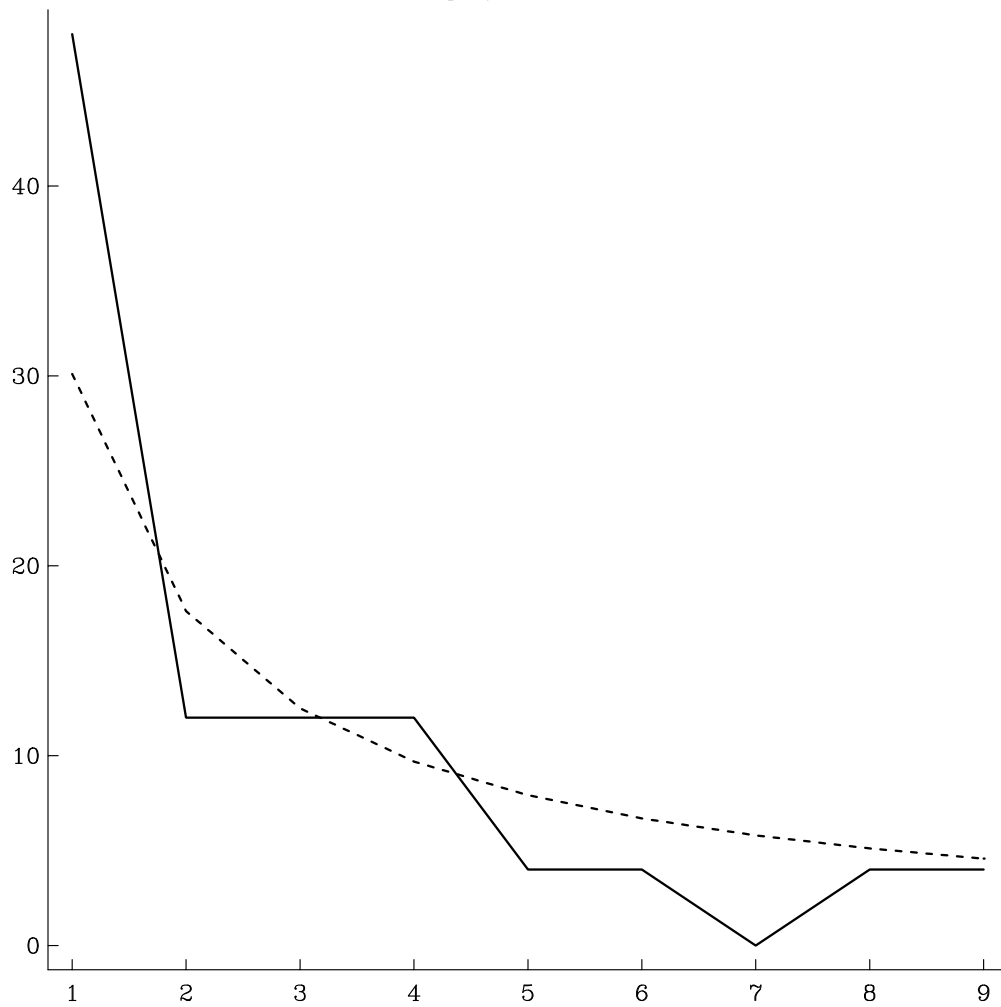
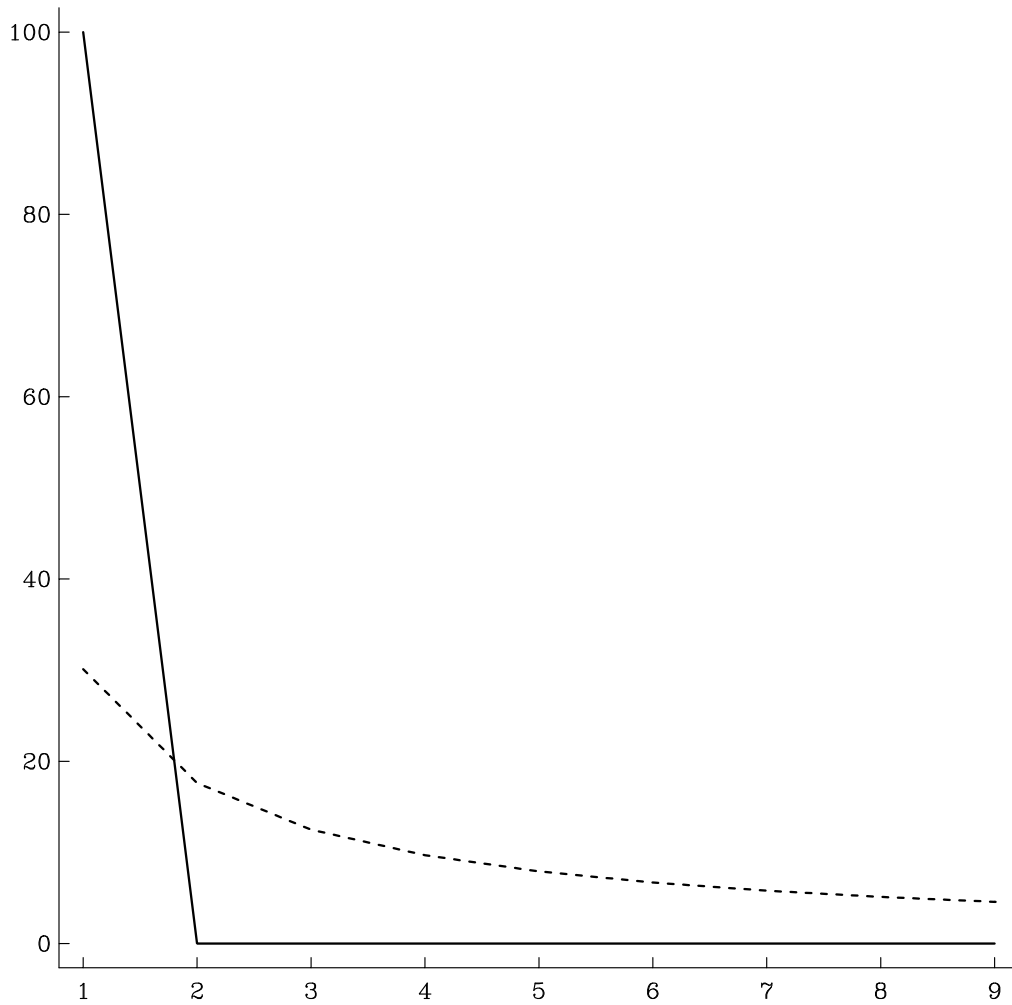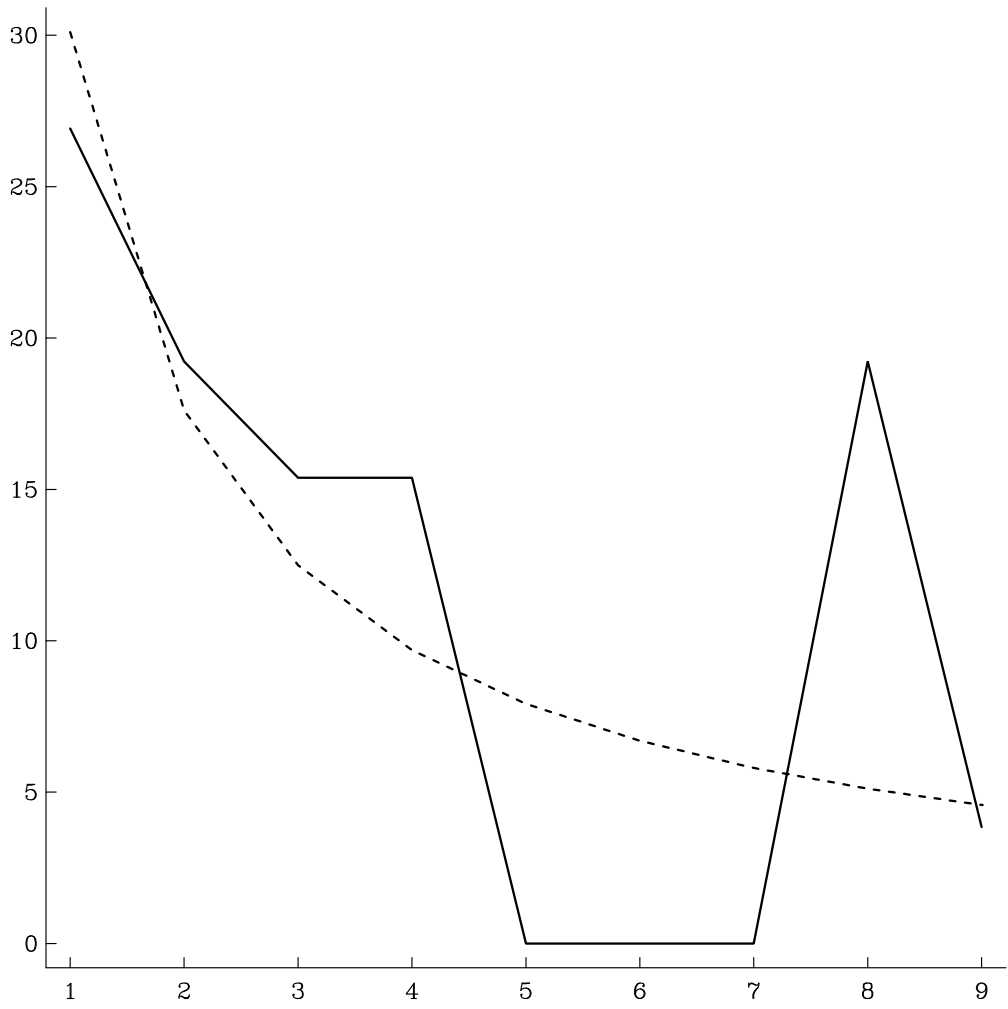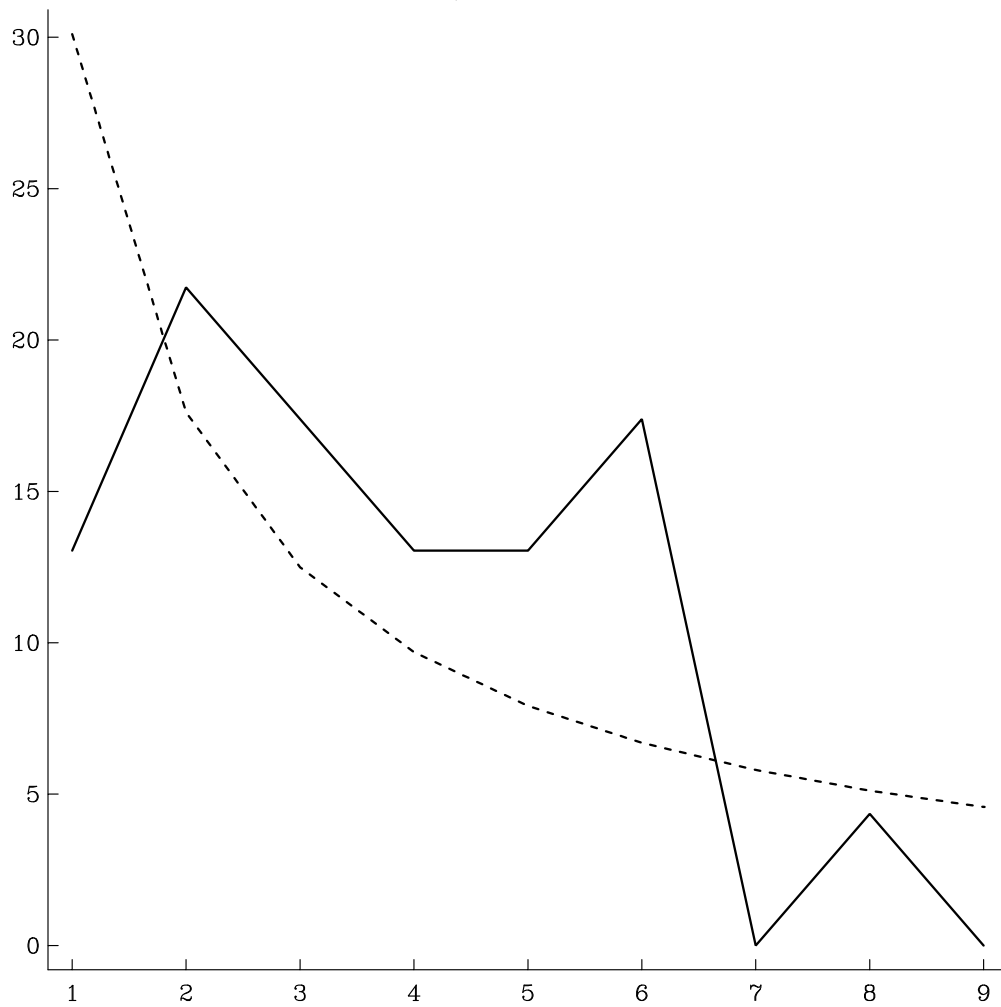Benford iceland.dat Cumulative data
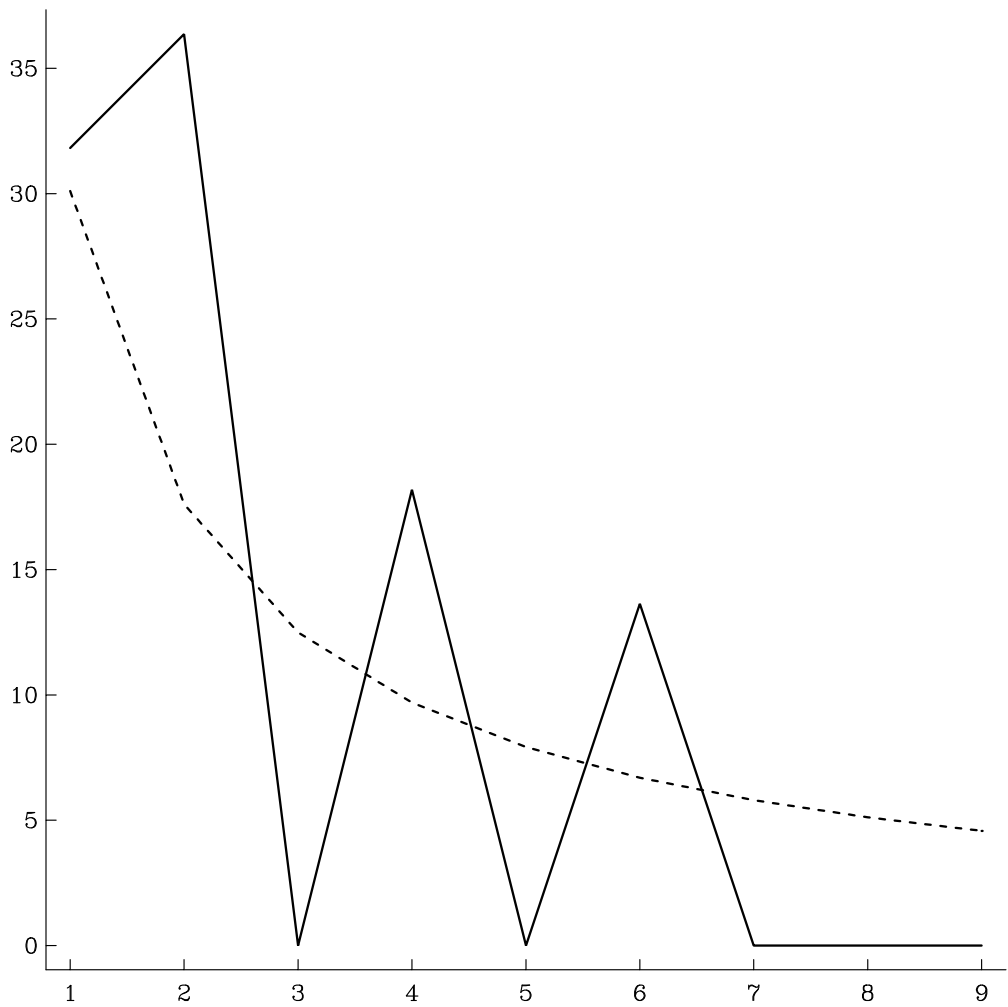
Benford india.dat Cumulative data

Benford indonesia.dat Cumulative data

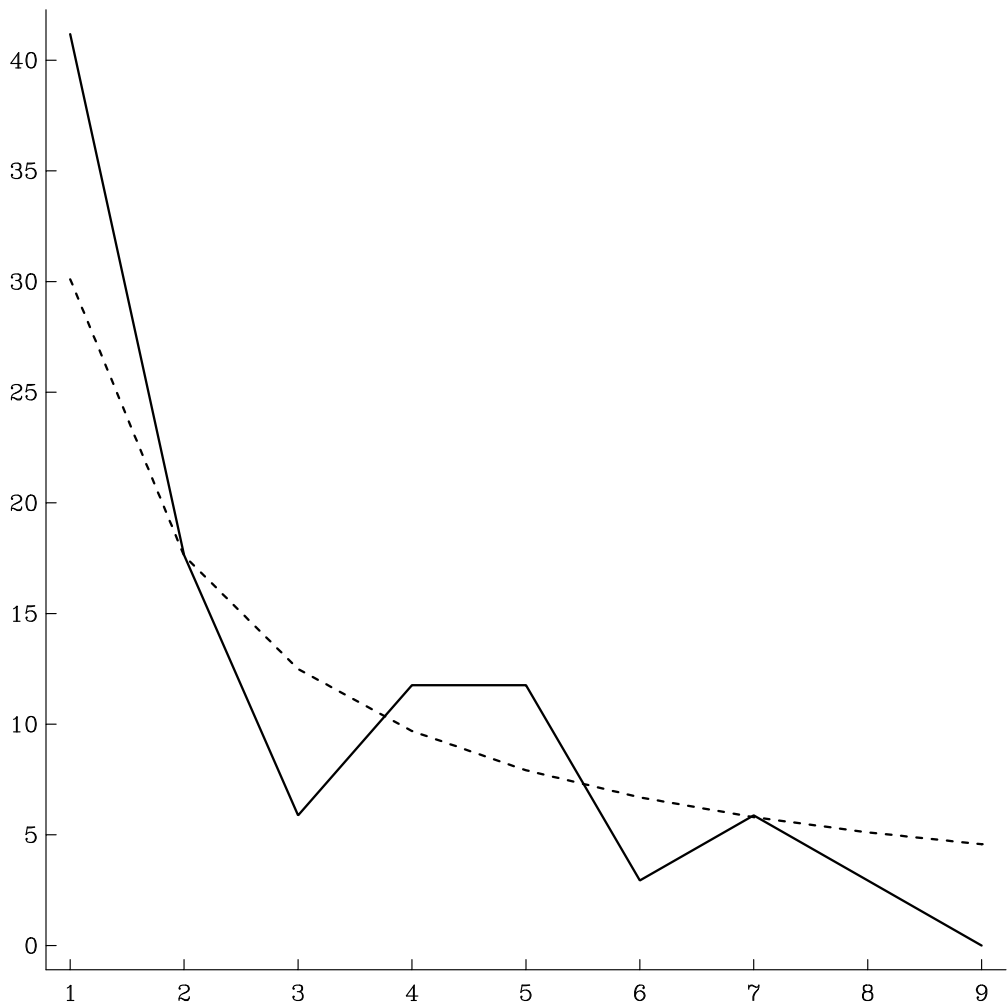Benford iran.dat Cumulative data

Benford ireland.dat Cumulative data

Benford israel.dat Cumulative data

Benford italy.dat Cumulative data

Benford japan.dat Cumulative data

Benford kuwait.dat Cumulative data

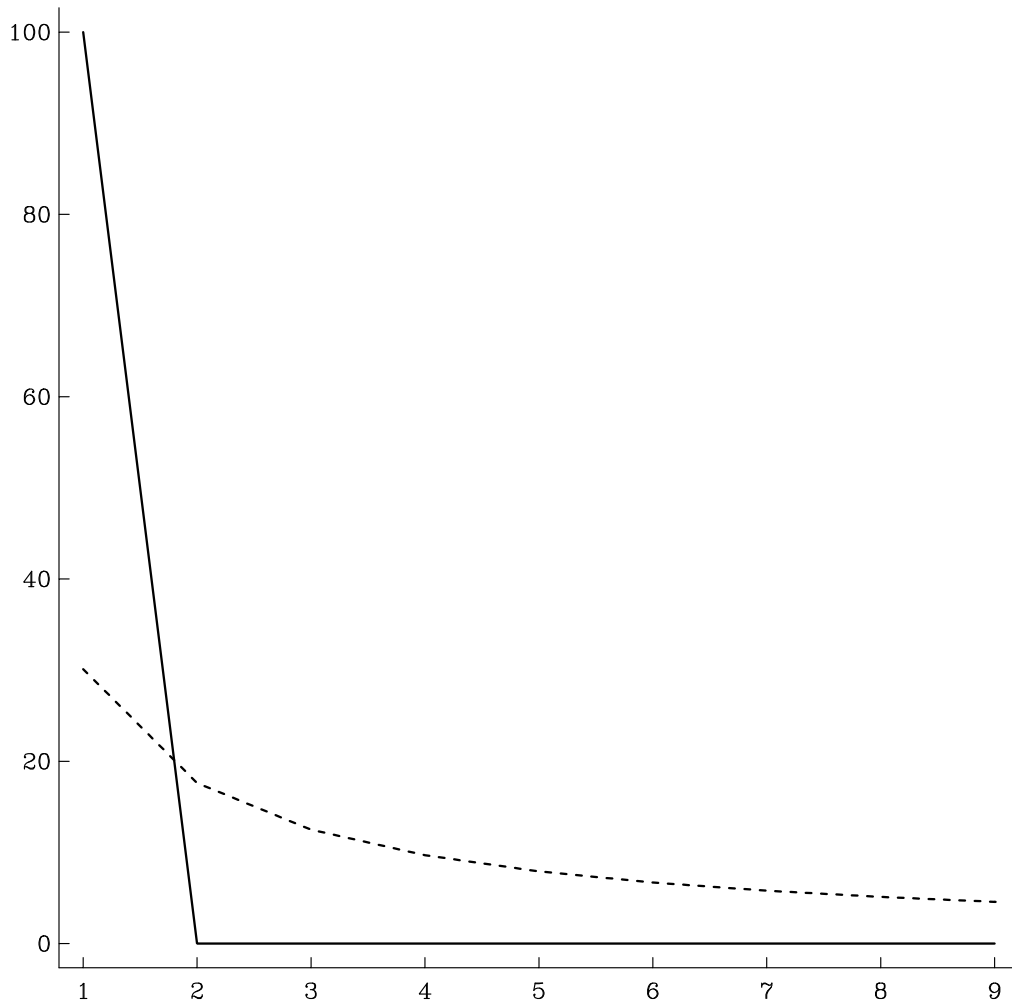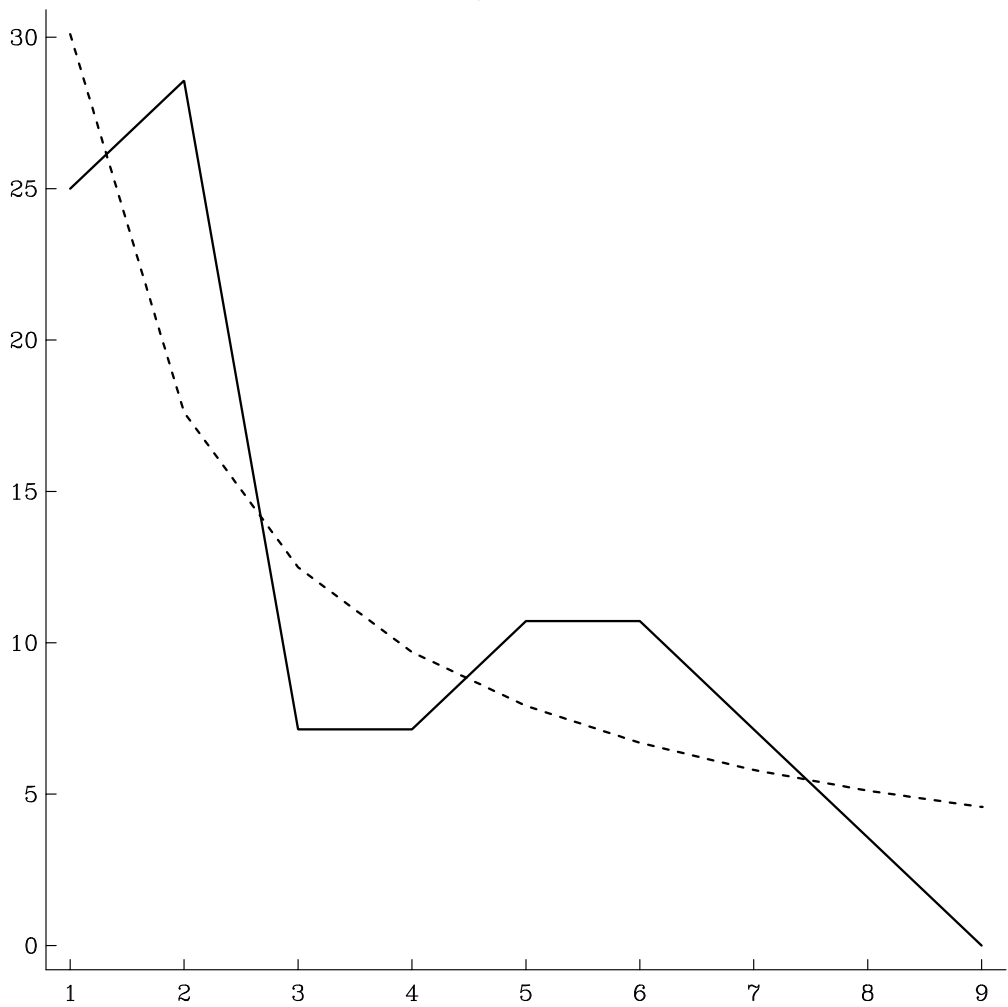Benford luxembourg.dat Cumulative data

Benford malaysia.dat Cumulative data

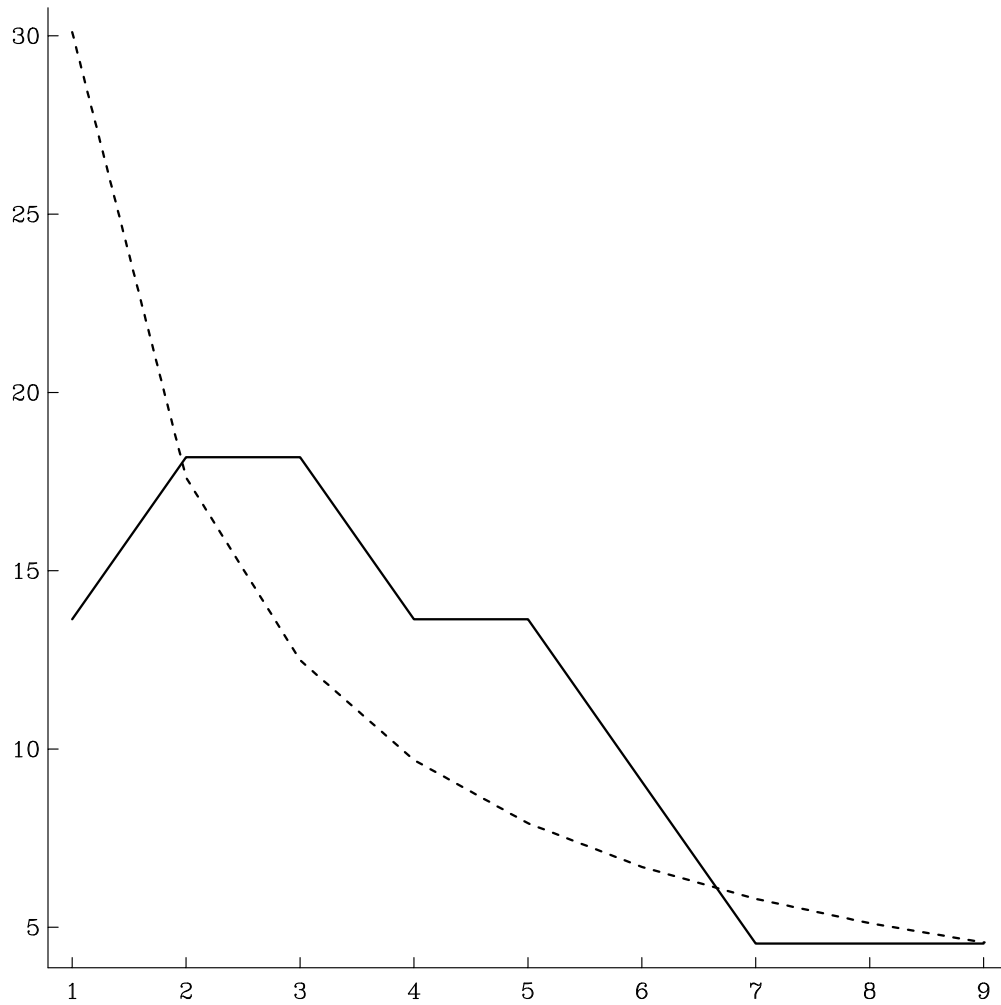Benford netherlands.dat Cumulative data
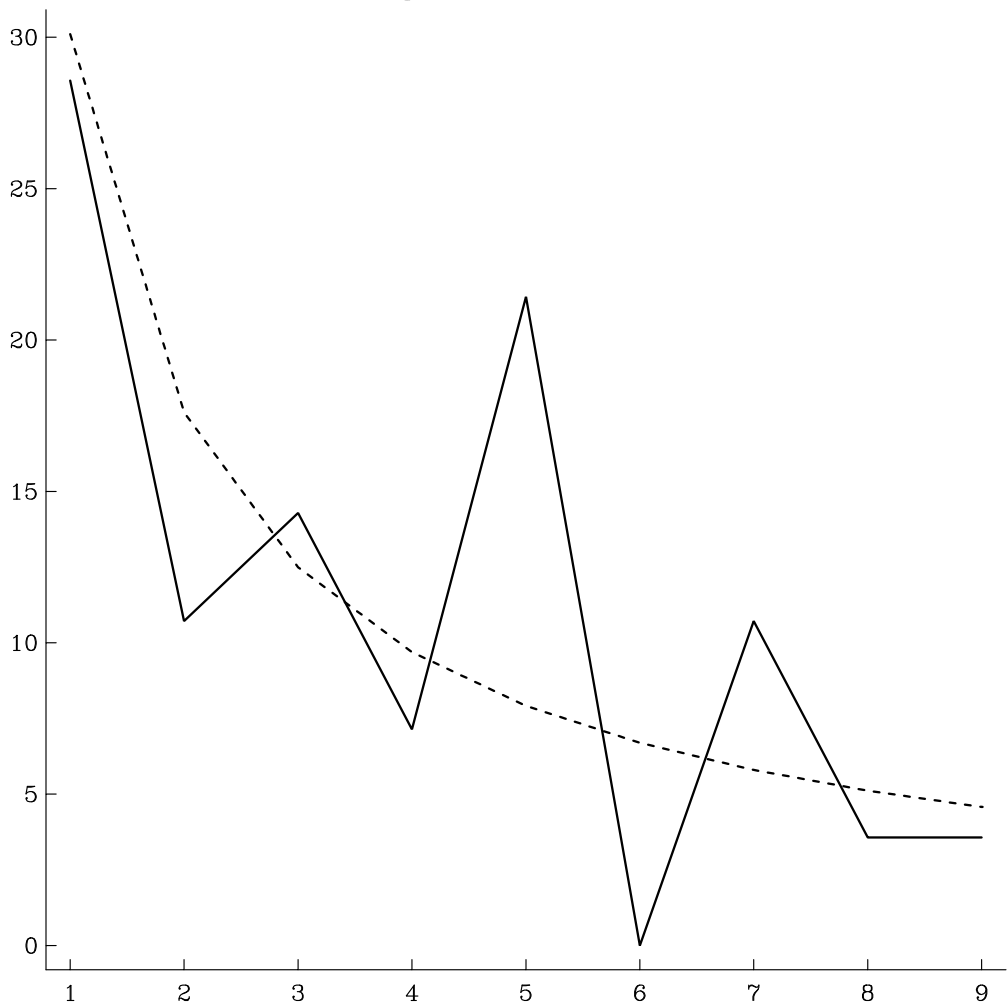
Benford newzealand.dat Cumulative data
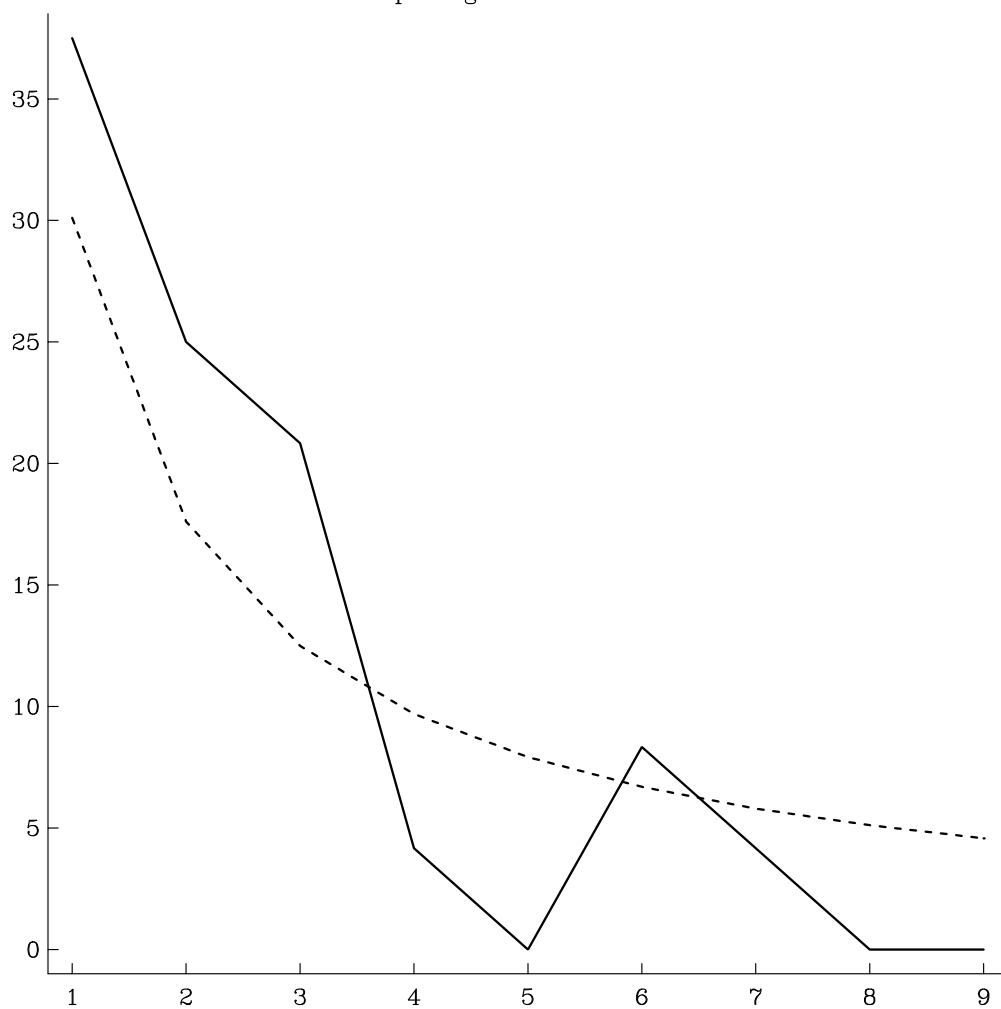
Benford norway.dat Cumulative data
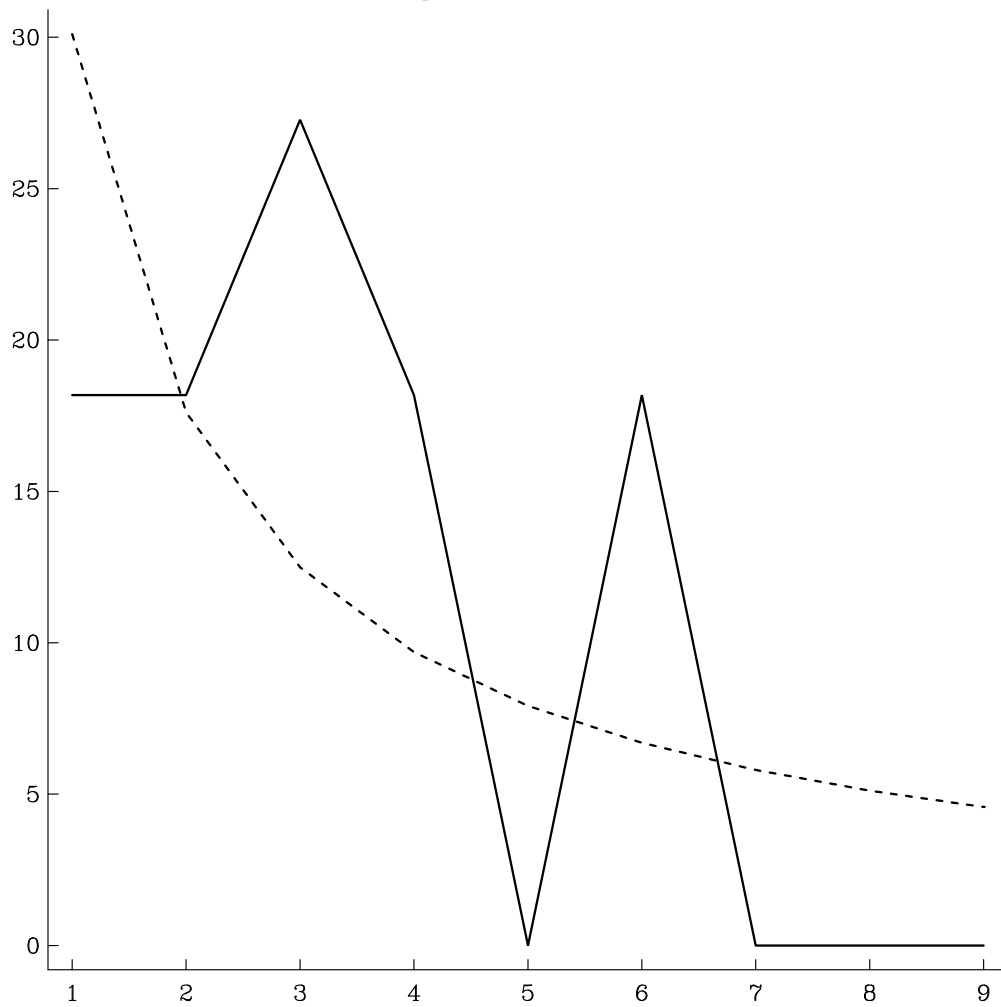
Benford pakistan.dat Cumulative data
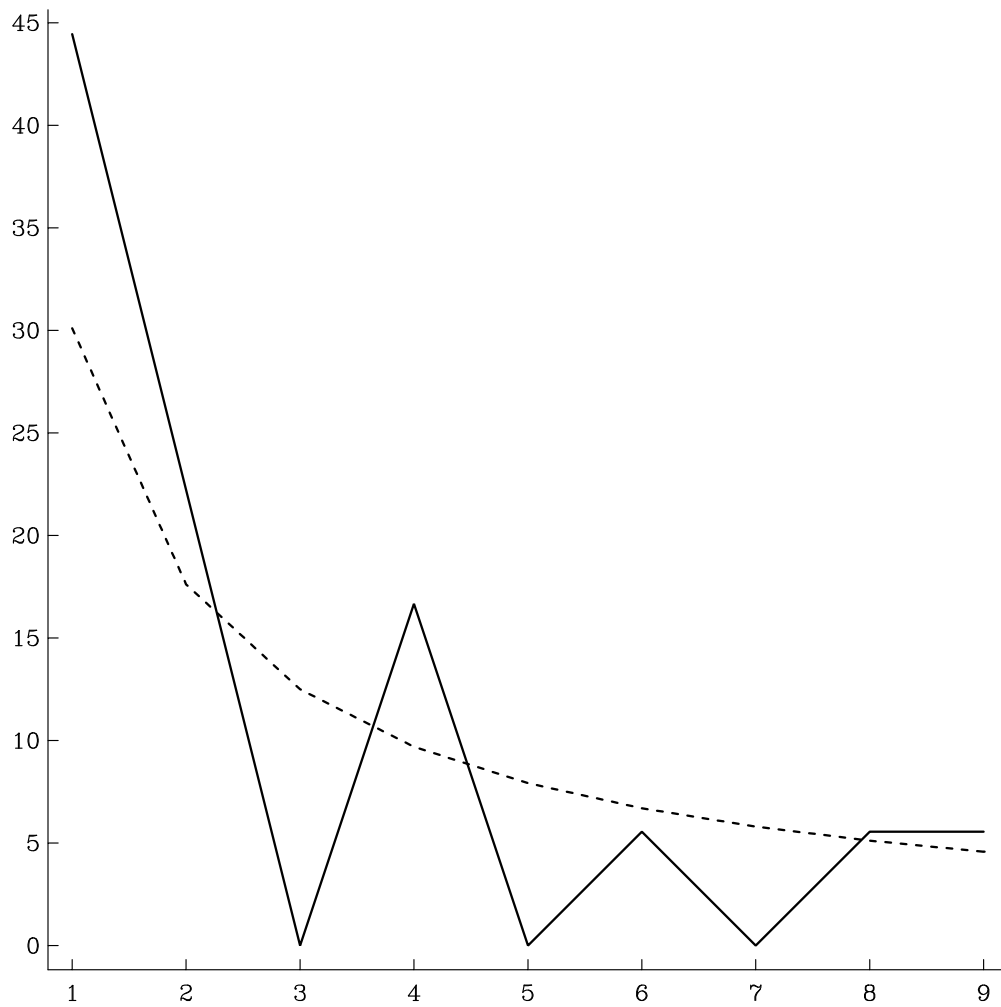
Benford poland.dat Cumulative data

Benford portugal.dat Cumulative data
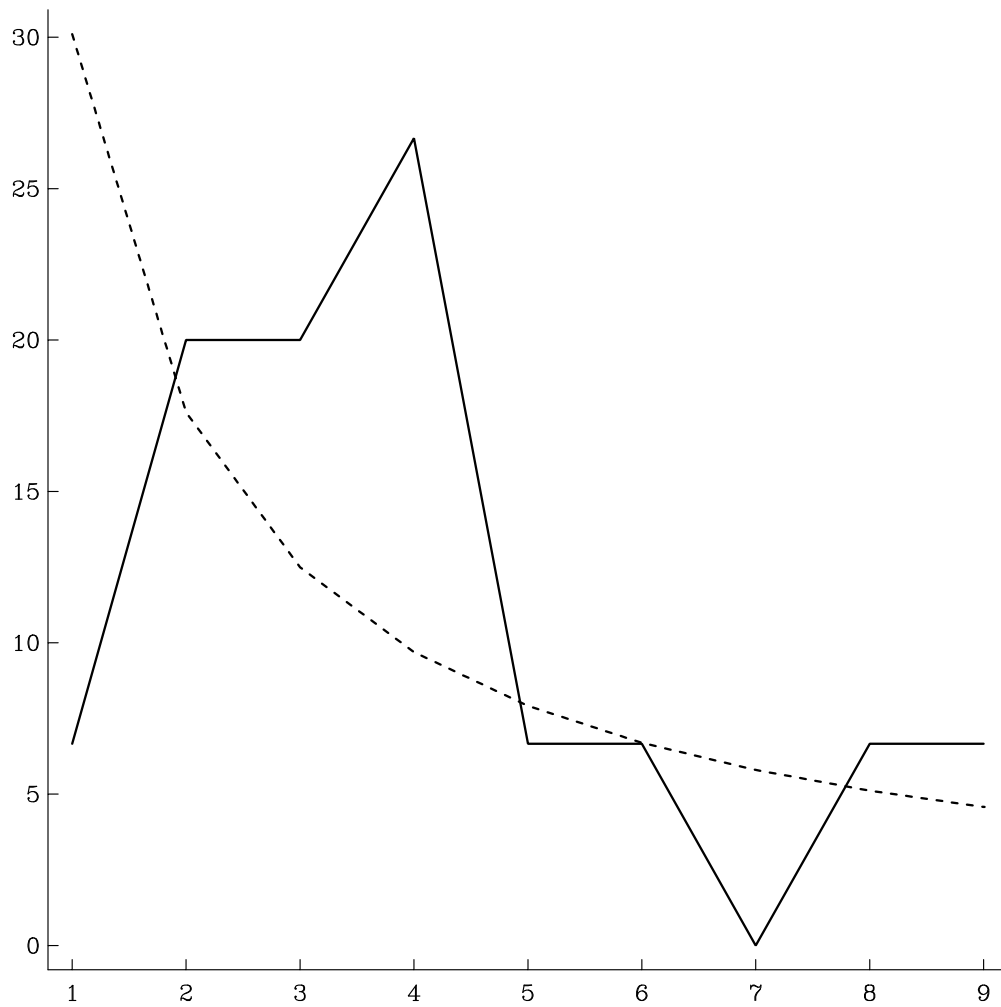
Benford qatar.dat Cumulative data

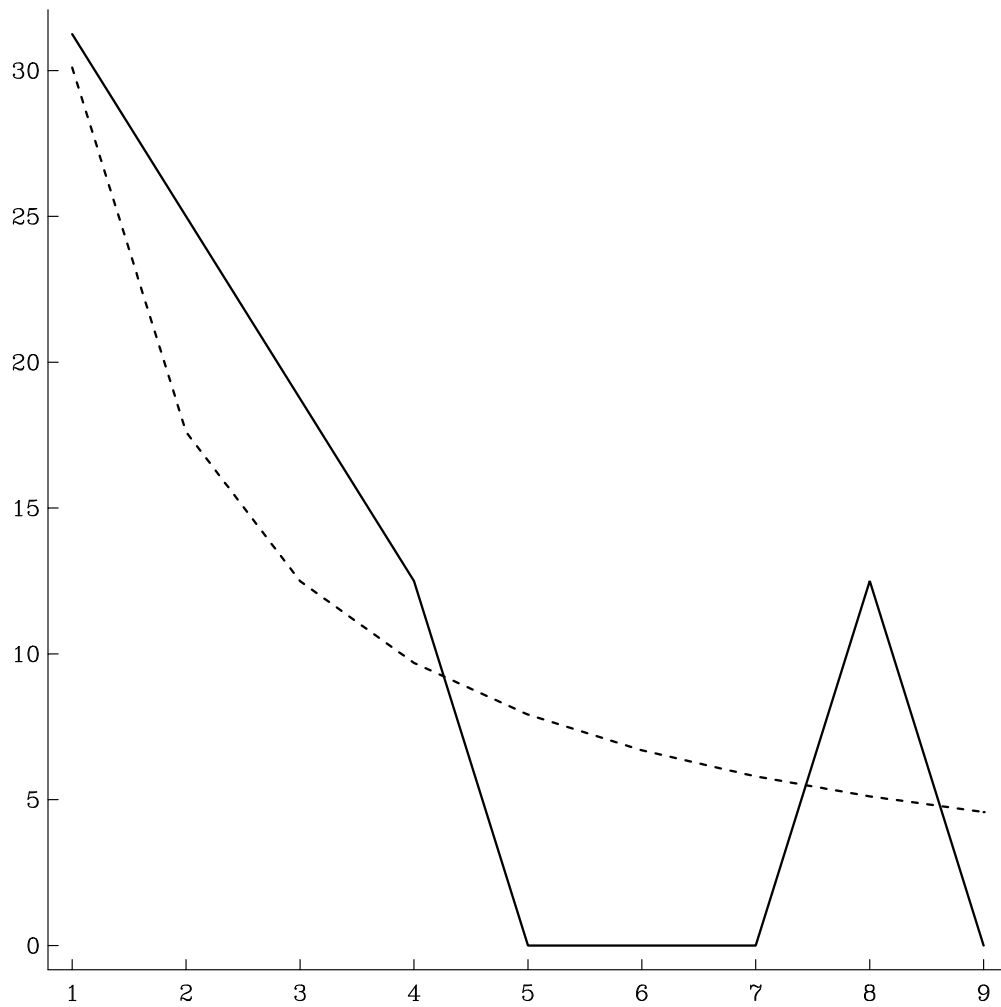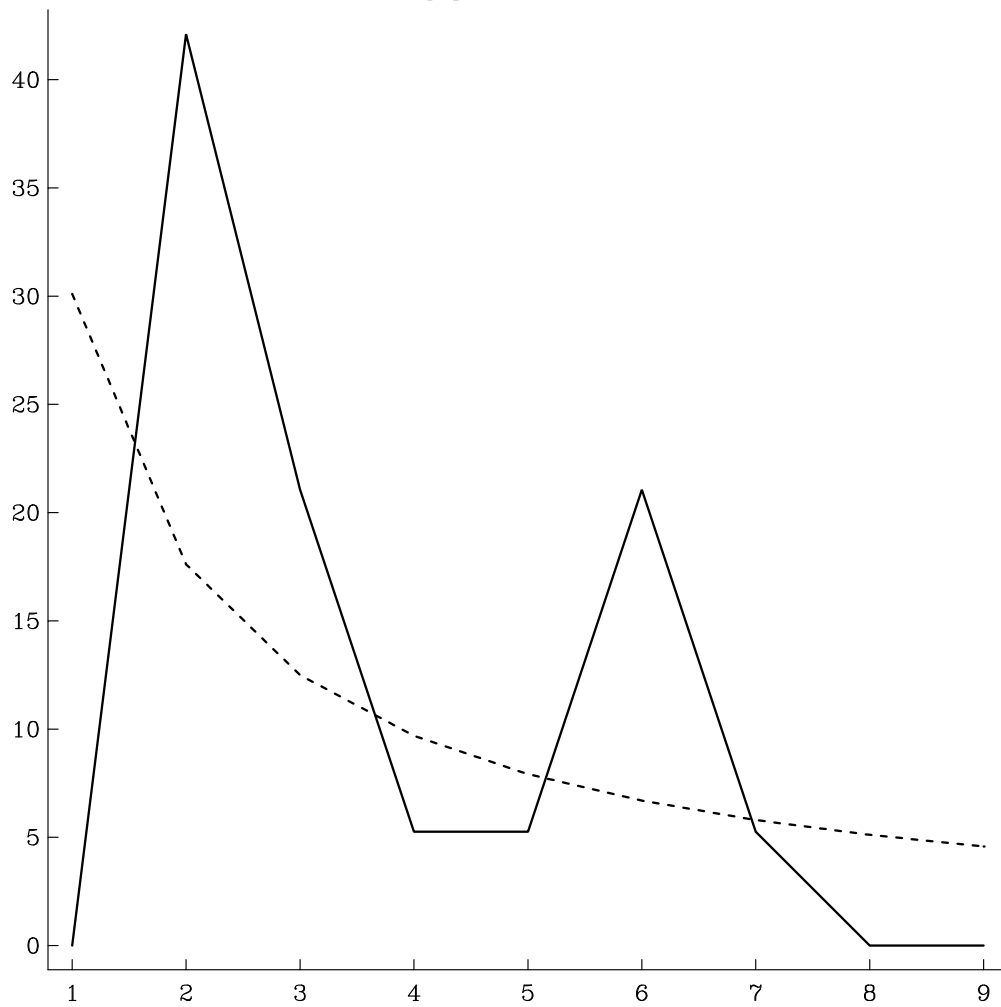Benford romania.dat Cumulative data
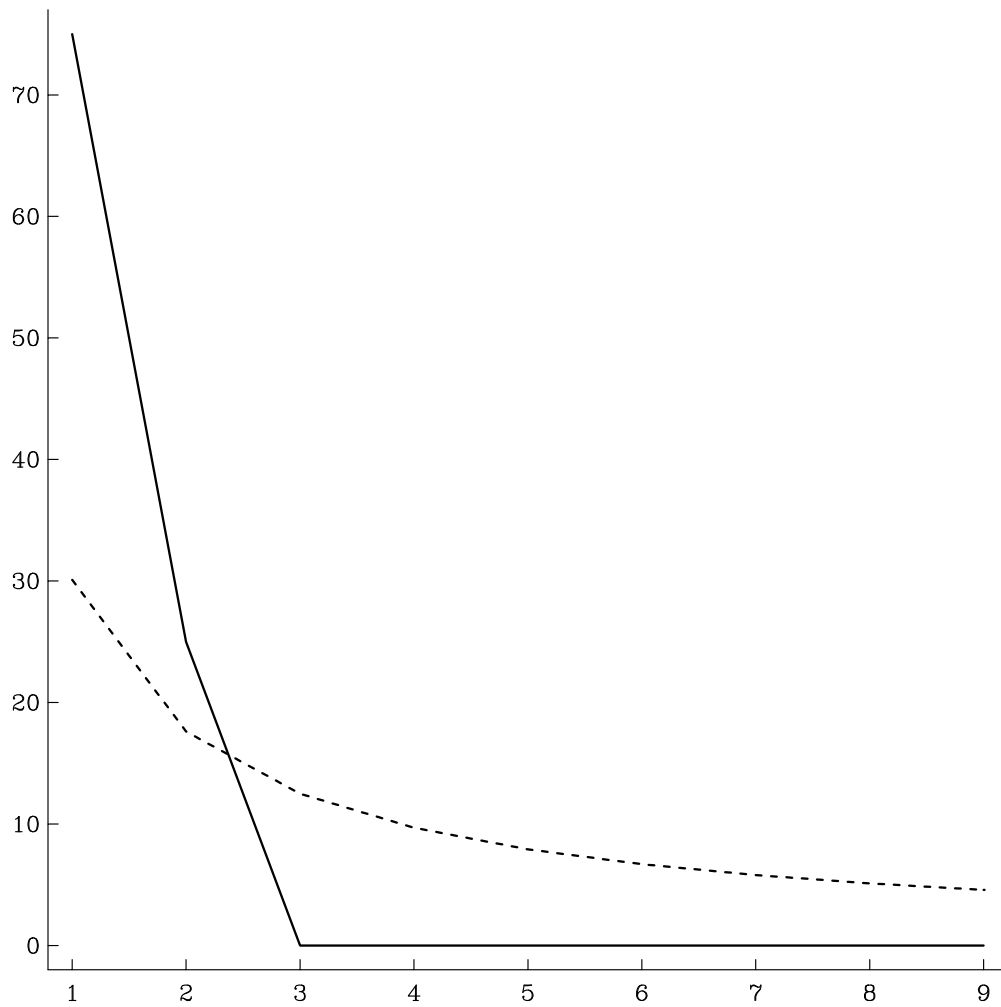
Benford russia.dat Cumulative data

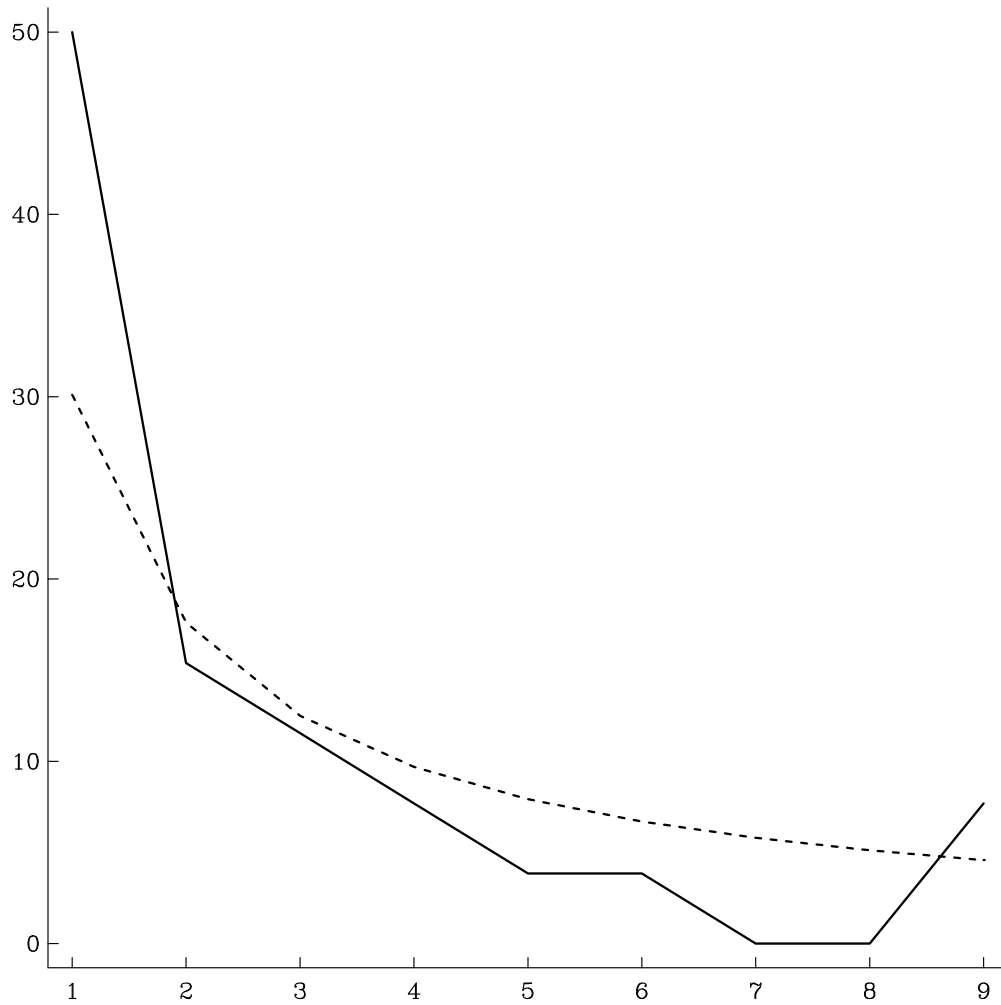Benford saudiarabia.dat Cumulative data
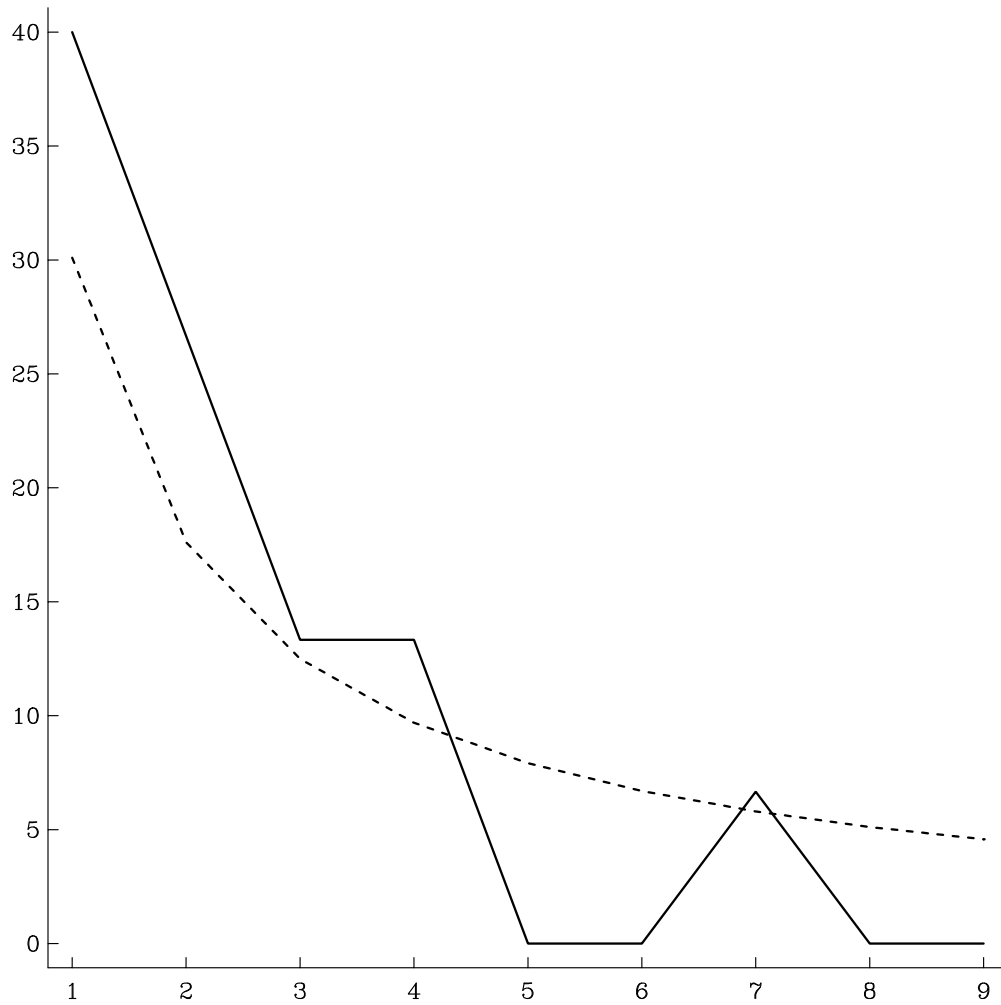
Benford singapore.dat Cumulative data
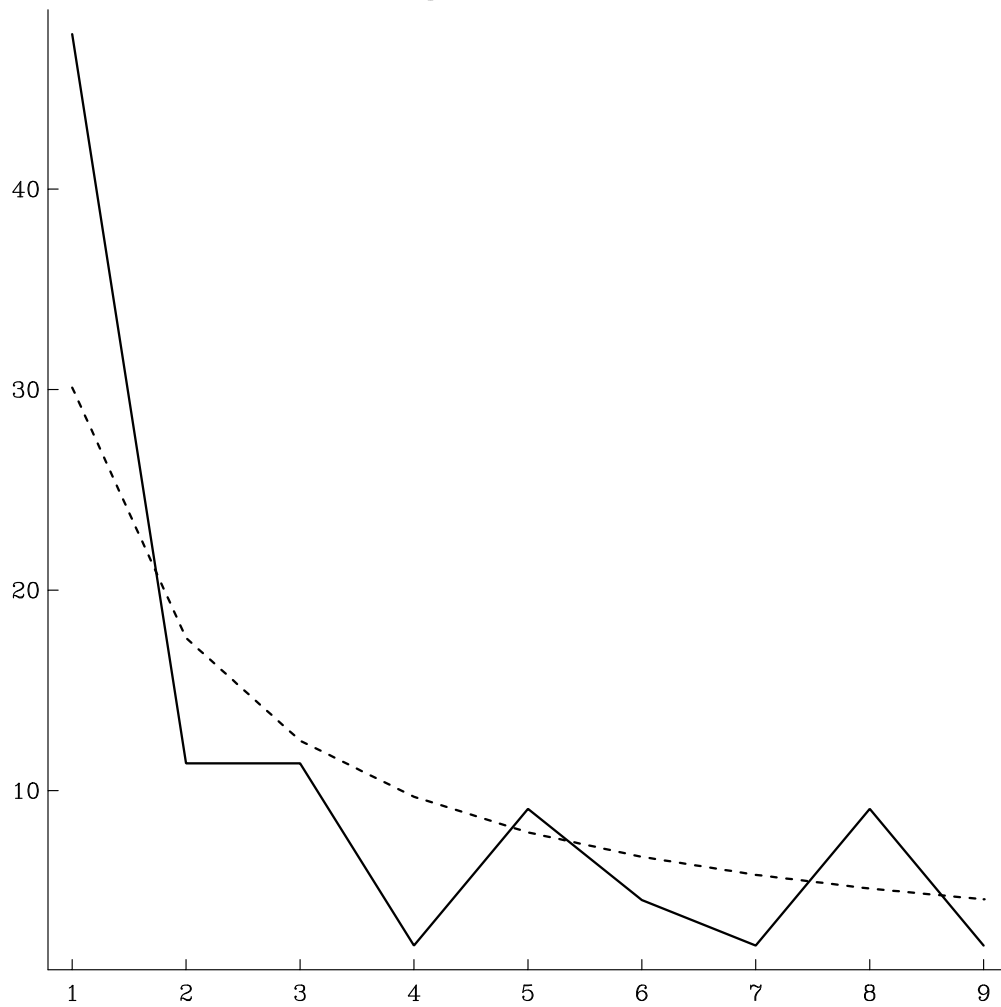
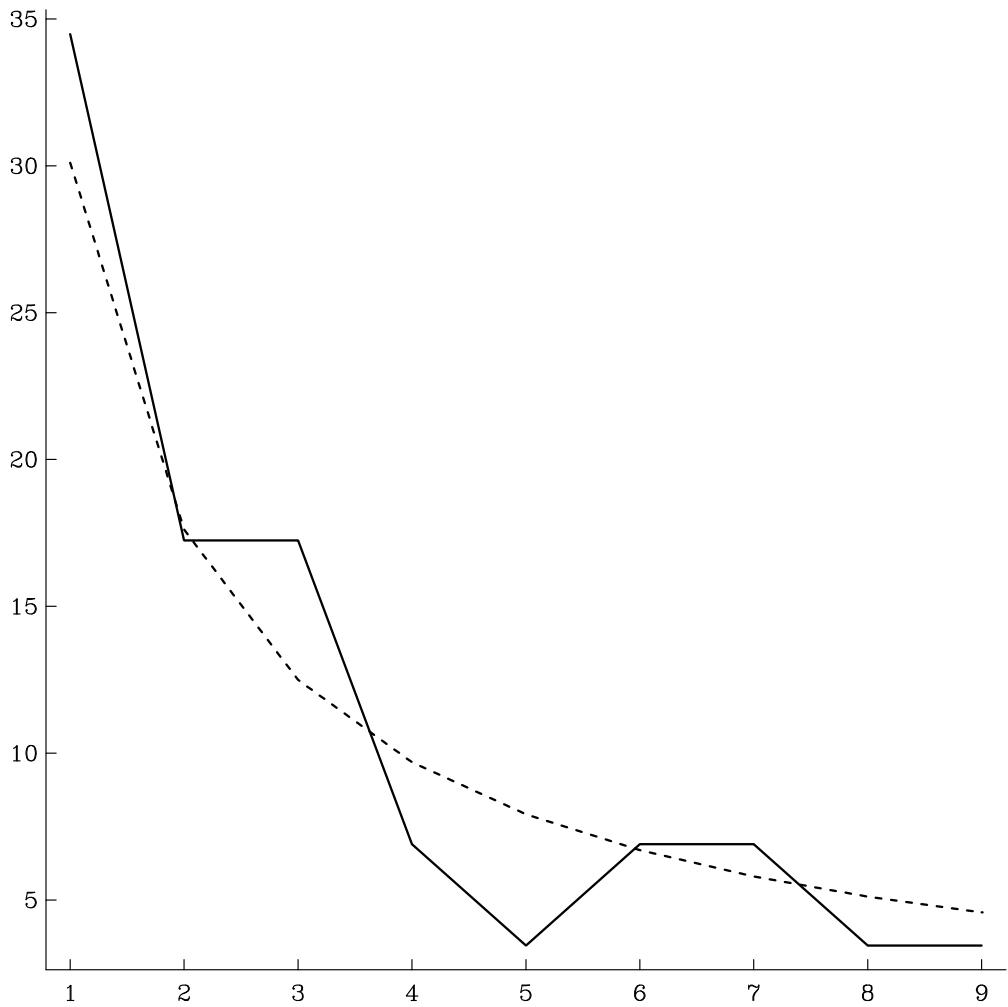Benford slovakia.dat Cumulative data

Benford slovenia.dat Cumulative data

Benford southkorea_pre_mar_5.dat Cumulative data

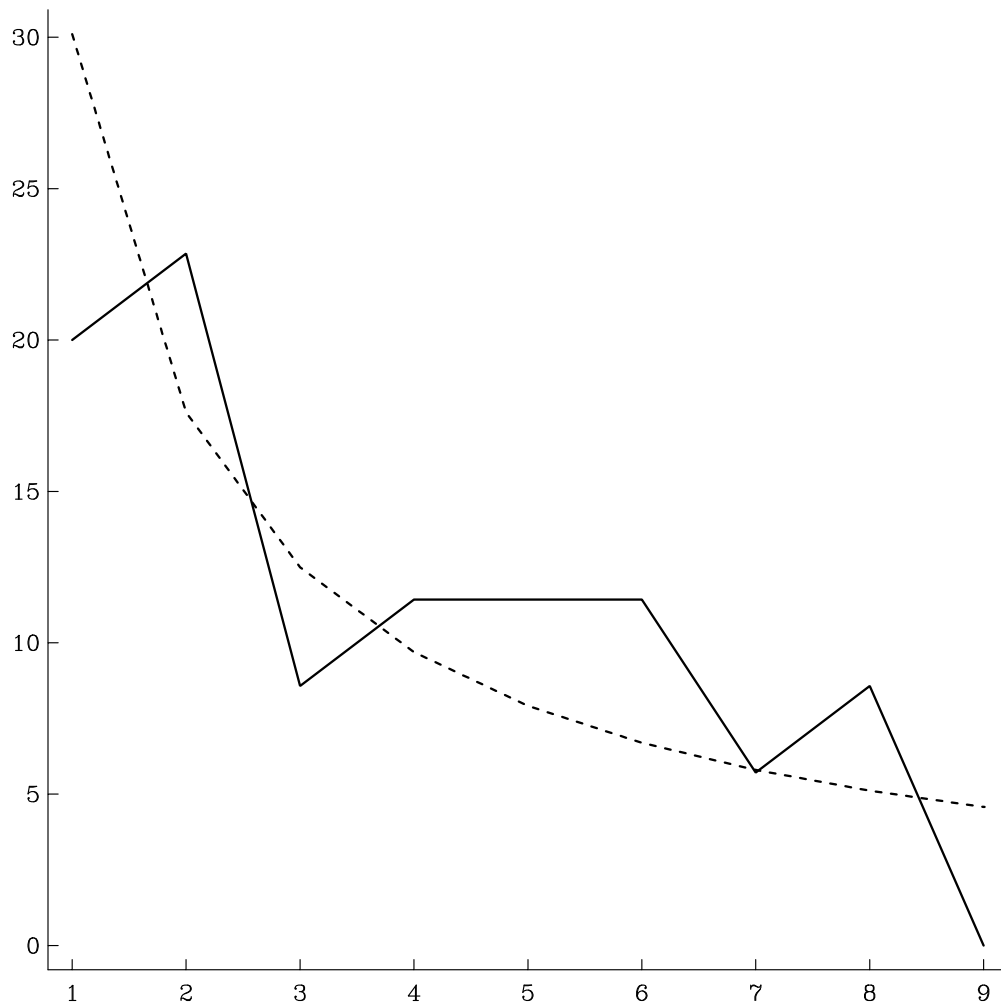Benford spain.dat Cumulative data
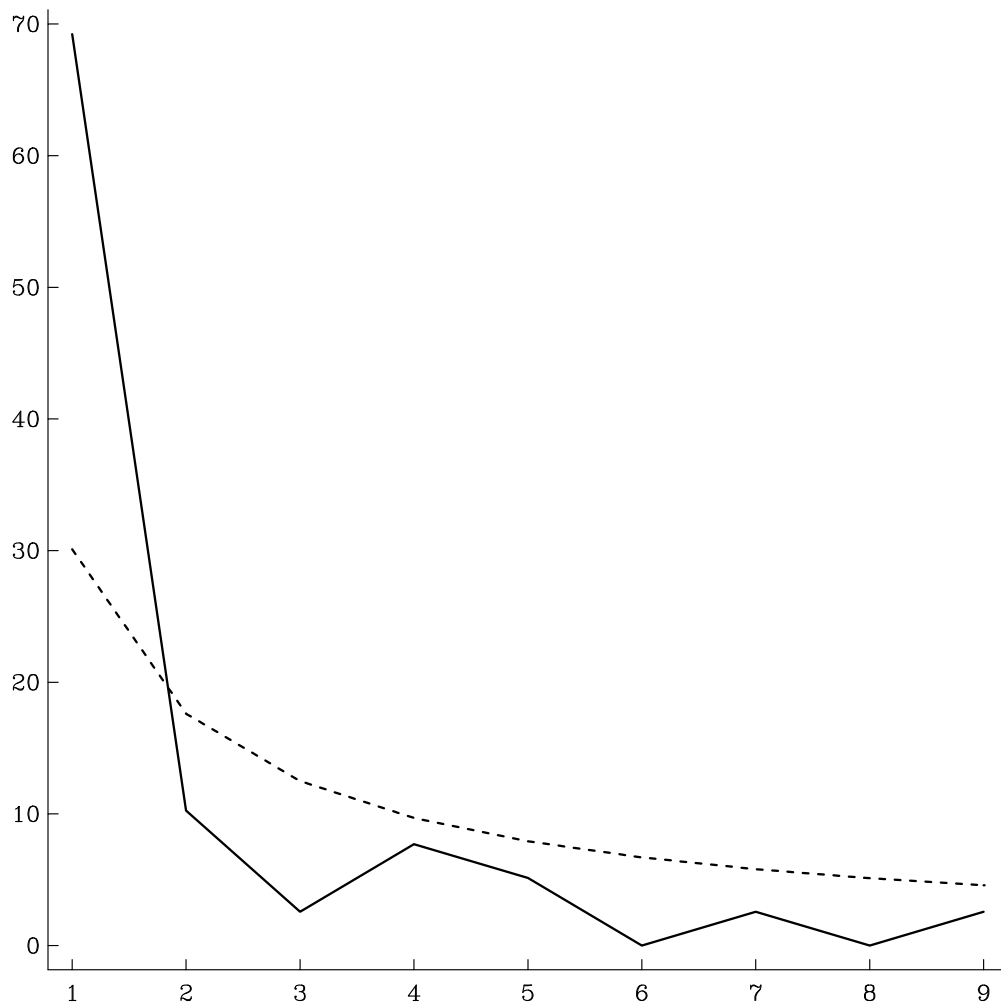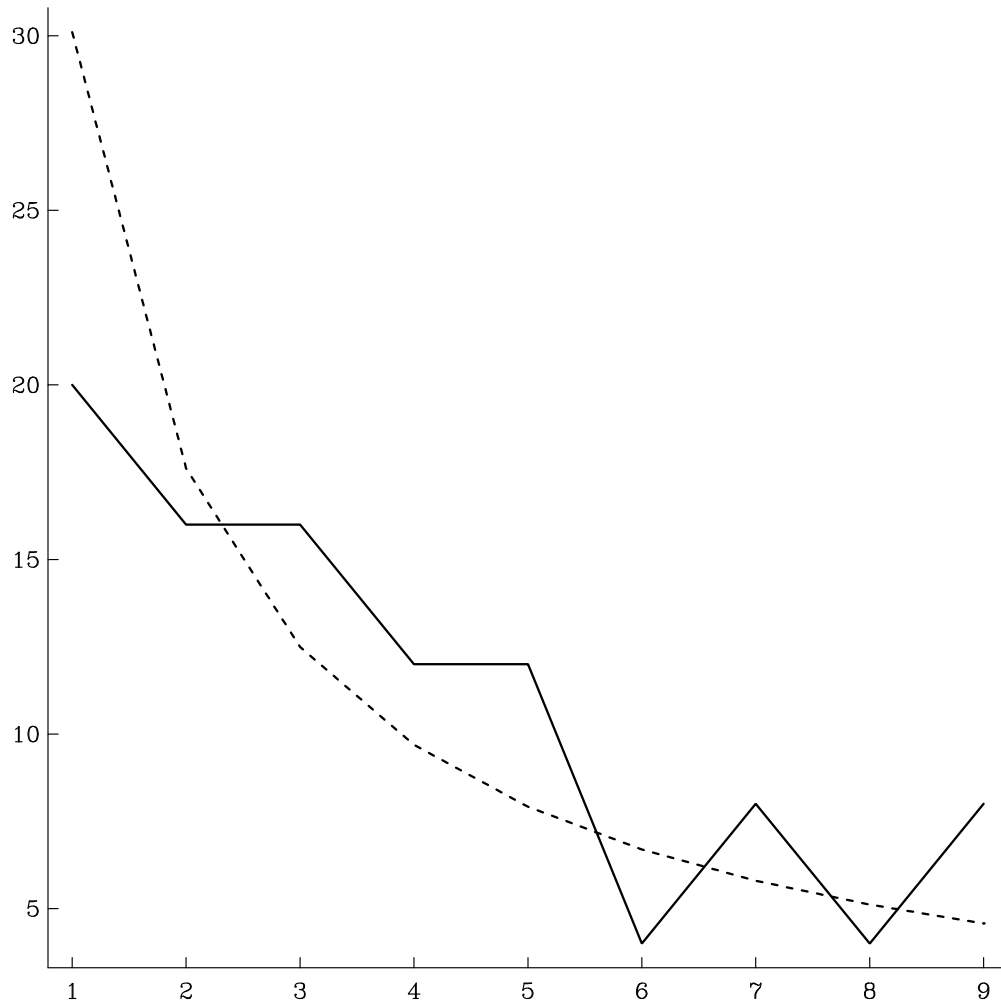
Benford sweden.dat Cumulative data

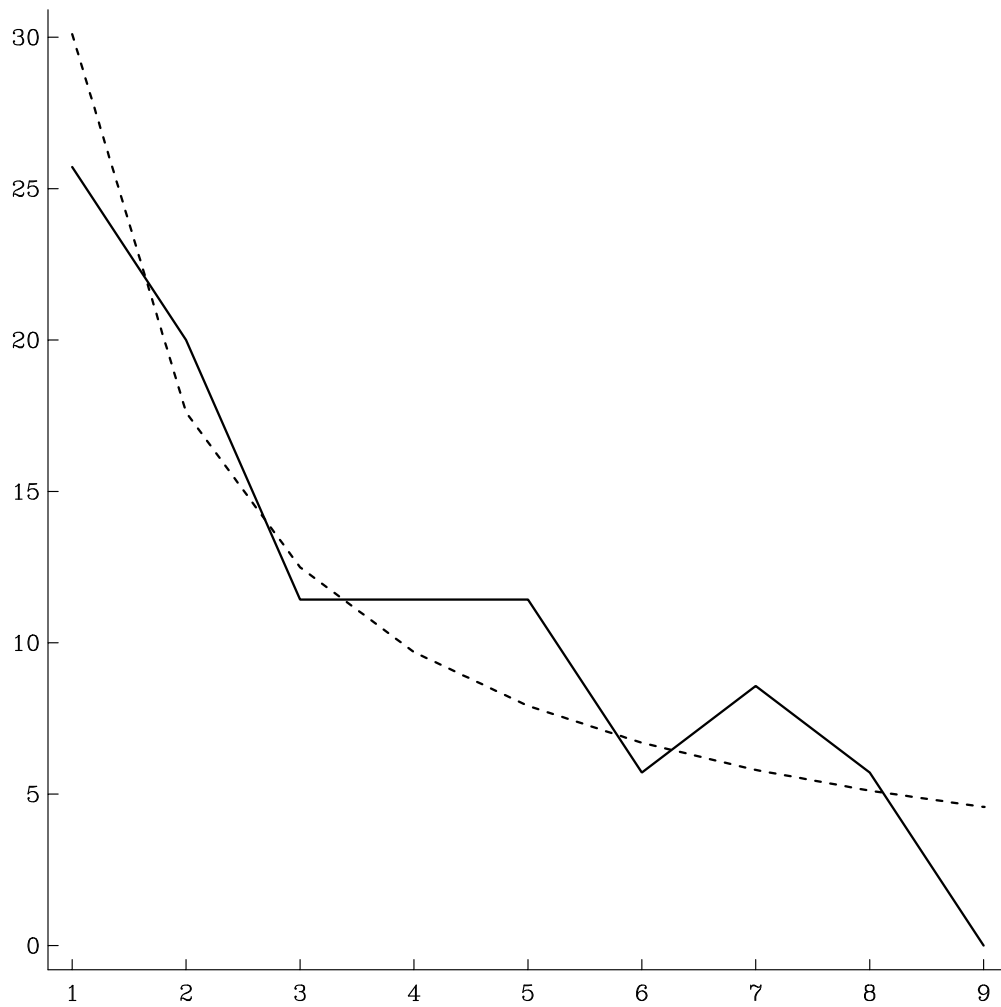Benford switzerland.dat Cumulative data

Benford thailand.dat Cumulative data

Benford turkey.dat Cumulative data

Benford uk.dat Cumulative data

Benford us.dat Cumulative data