Benford analysis of Covid-19 data

Andrew Jackson (ETH Zürich, Switzerland) and Malcolm 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 succumbed 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.

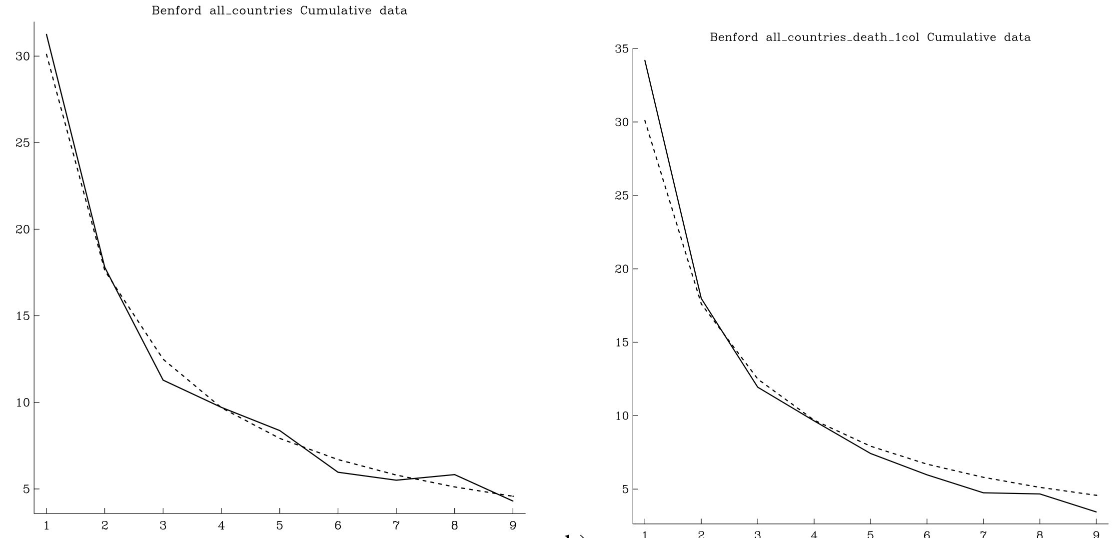
We see a clear adherence from all the data to the expected distribution. We can see this also at a country level (Figure 2). We take as examples the European countries France, Germany, Spain, UK, Switzerland, Italy. All these countries have experienced large growths in numbers. Since the datasets are 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.[1]. Figure 4 shows the correlation coefficient for all countries' case data when compared to the theoretical curve.

It is generally accepted that compliance of data sets to Benford's prediction improves as both numbers of data and their dynamic range increase. However, with Covid-19 data, as cumulative cases numbers begin to flatten in each country, Benford compliance is likely to decrease because first digits tend to all become the same, significantly distorting relative digit frequencies. This is already the case with China and South Korea in the datasets examined here, and will likely become the case with many countries over time.

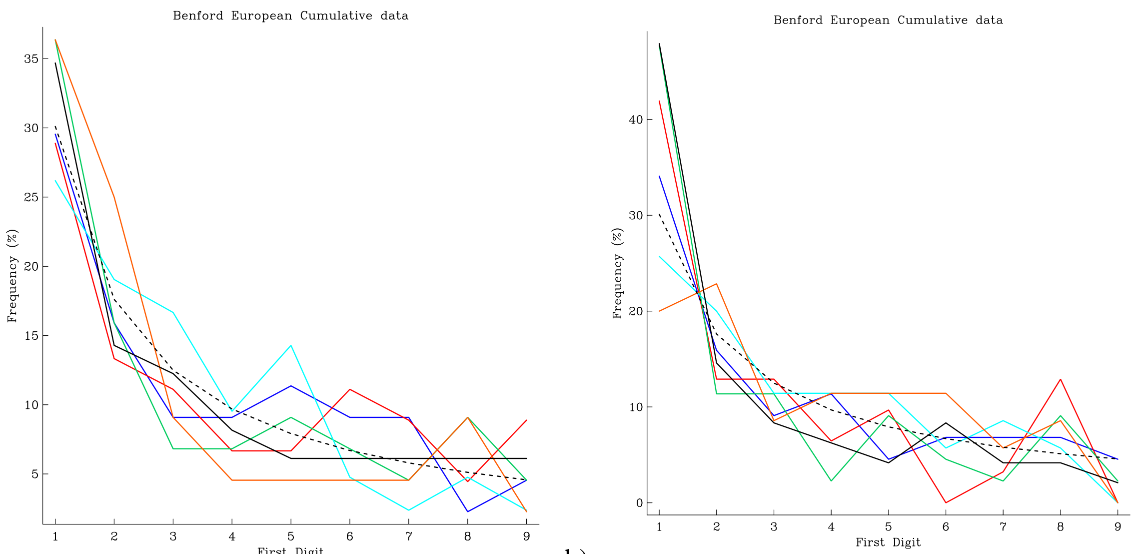
[1]Evidence of the reluctance to report figures passing a particular threshold can be found in statements to the media: *Irans daily death toll exceeded 100 for the first time since 13 April, taking the overall number of deaths to more than 8,800. "It was very painful for us to announce the triple-digit figure, the health ministry spokeswoman, Sima Sadat Lari, said. This is an unpredictable and wild virus and may surprise us at any time."* (Guardian online 15/6/20, article entitled "Global report: China detects highest daily case rise in months".)

More information is at `https://github.com/anu-ilab/Covid-19-Benford-Analysis` where updated data sets, figures and analysis can be found.



a) b)

Figure 1: Comparison of data from all countries together with the expected distribution (dashed). a) Cases. b) Deaths.



a) b)

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.

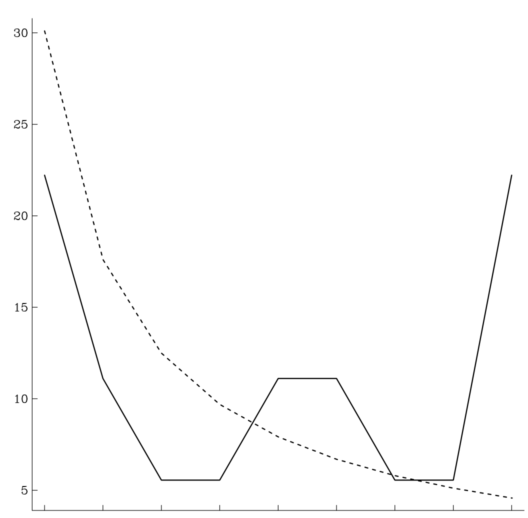


Figure 3: Death data reported by the Czech Republic as of 9/4/20.

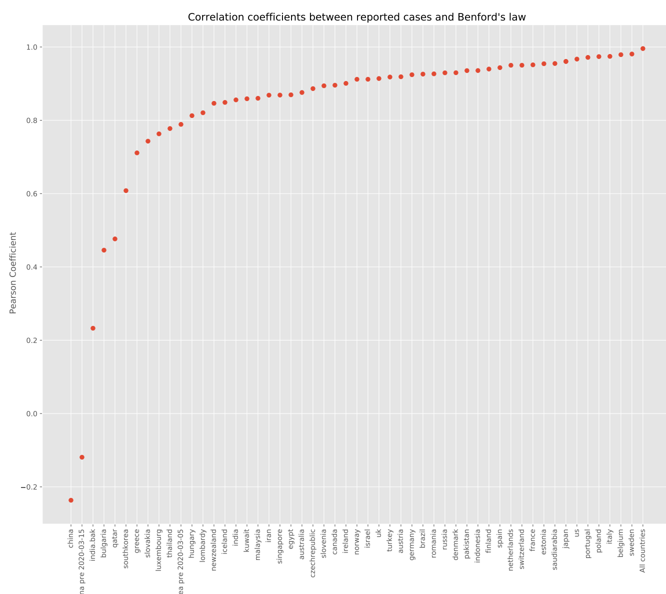
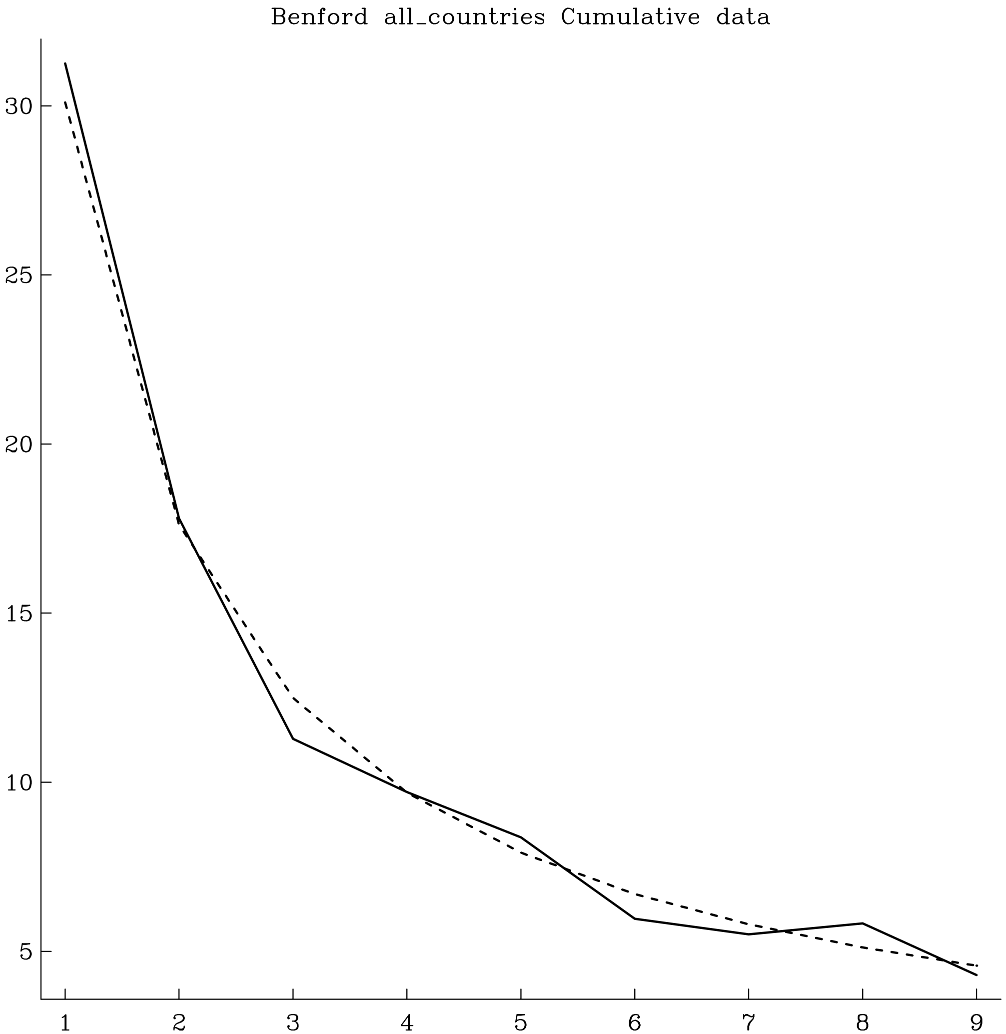
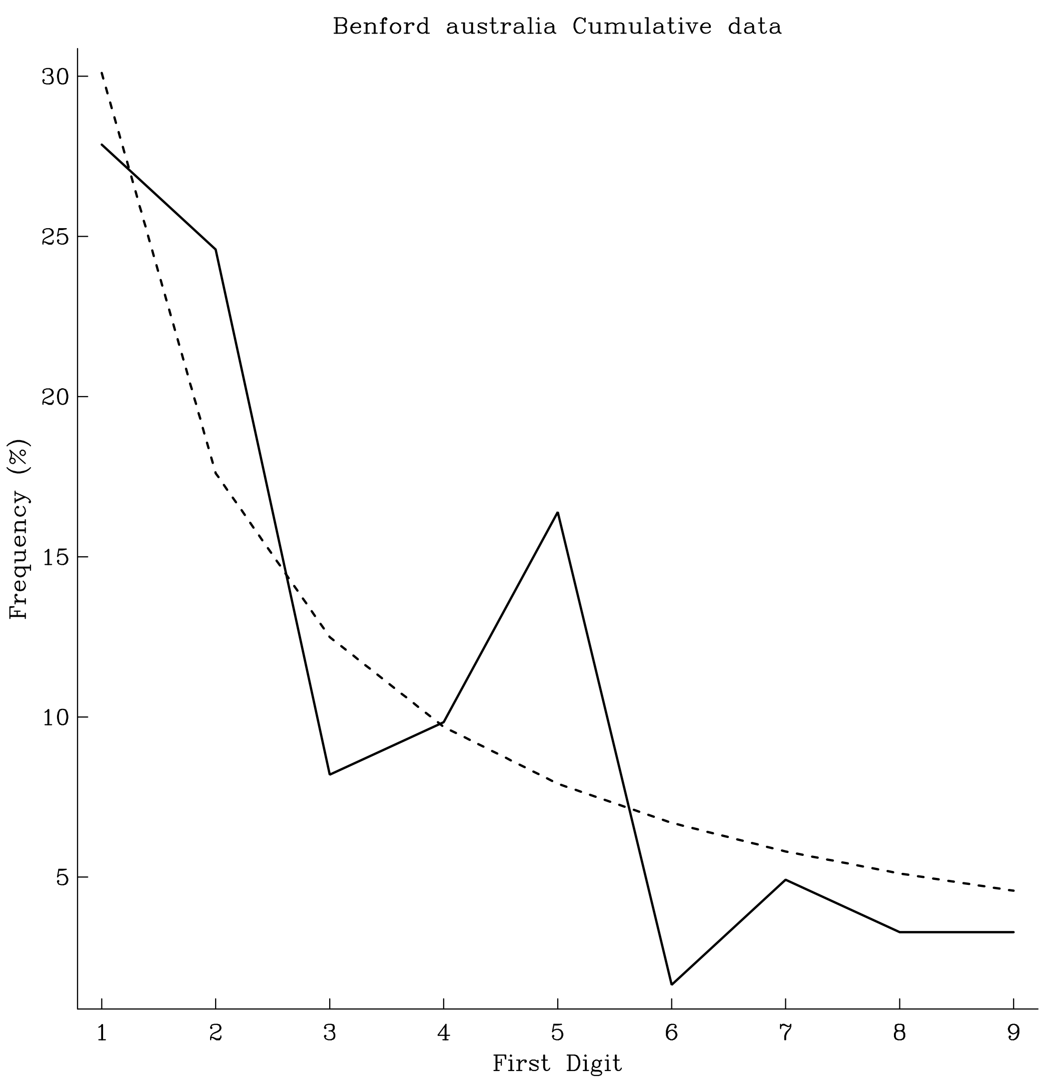


Figure 4: Correlation coefficient for each of the countries' case data that was analysed.

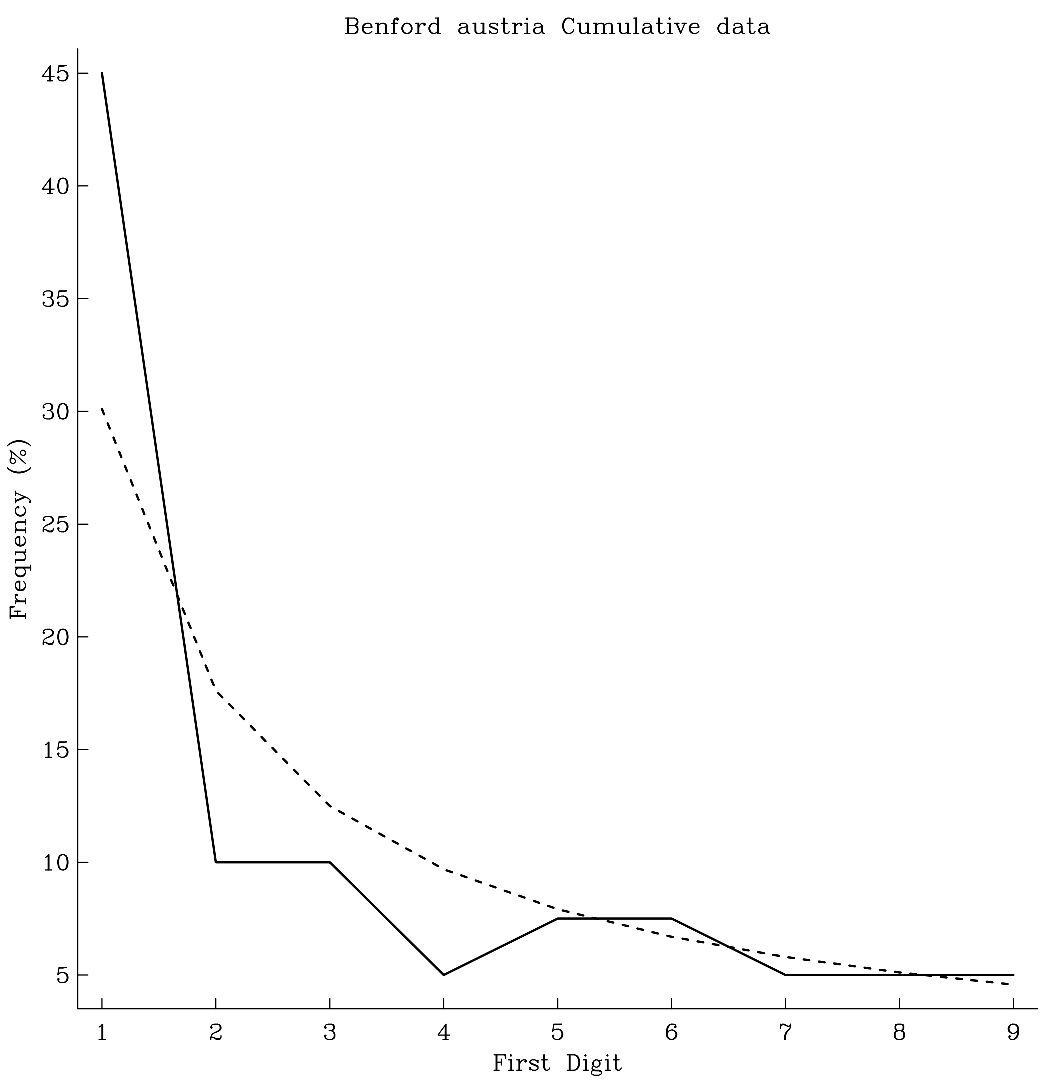
Cases by country

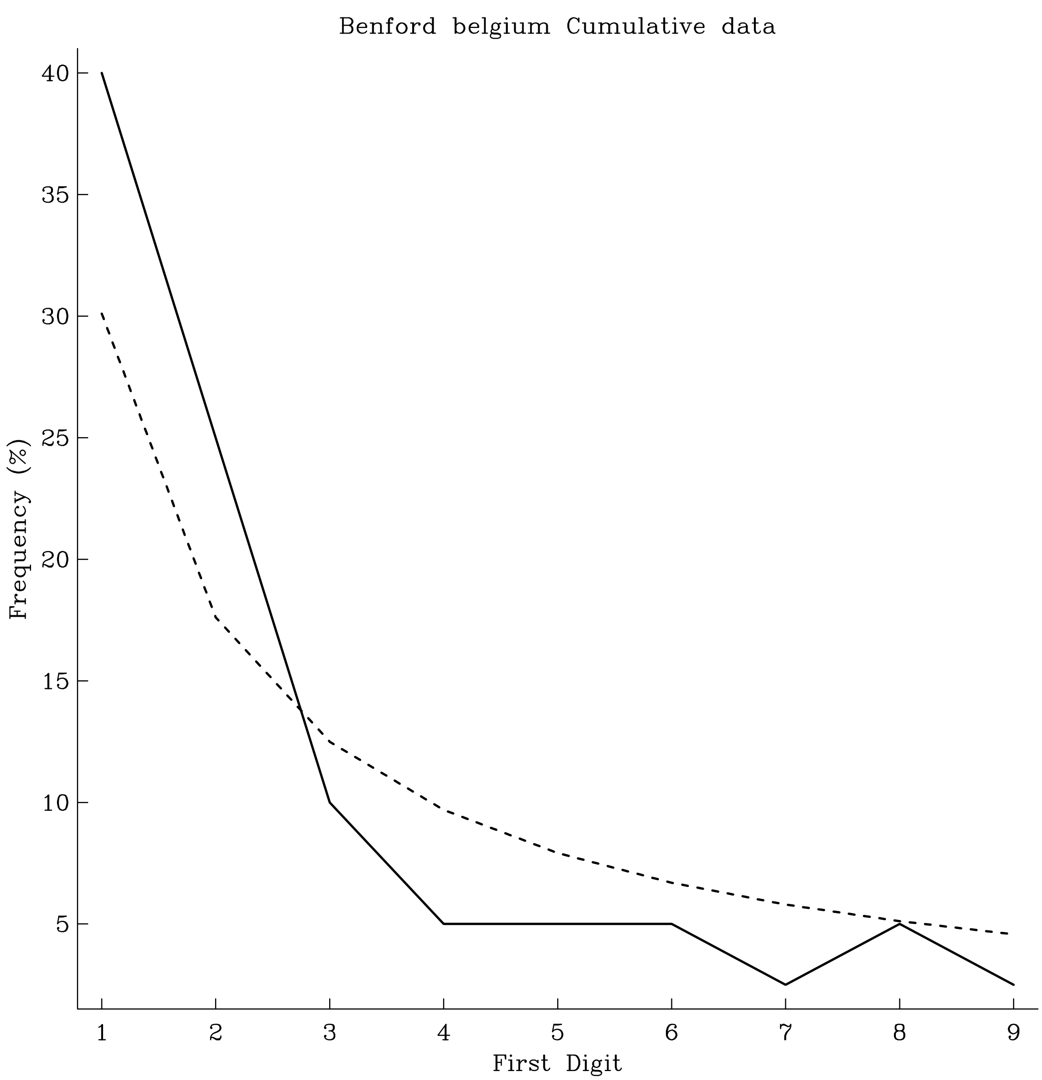


Benford all_countries Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

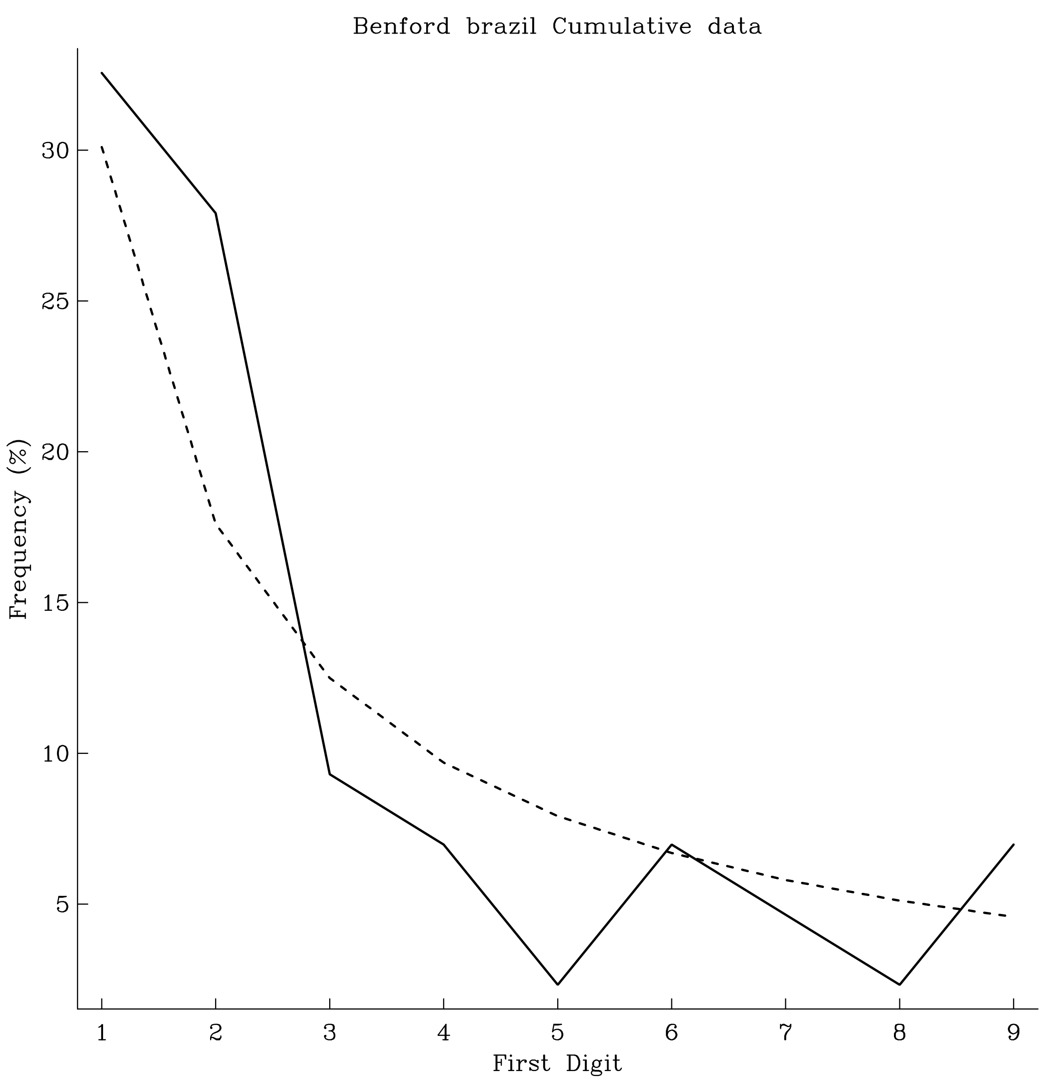


Benford australia Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

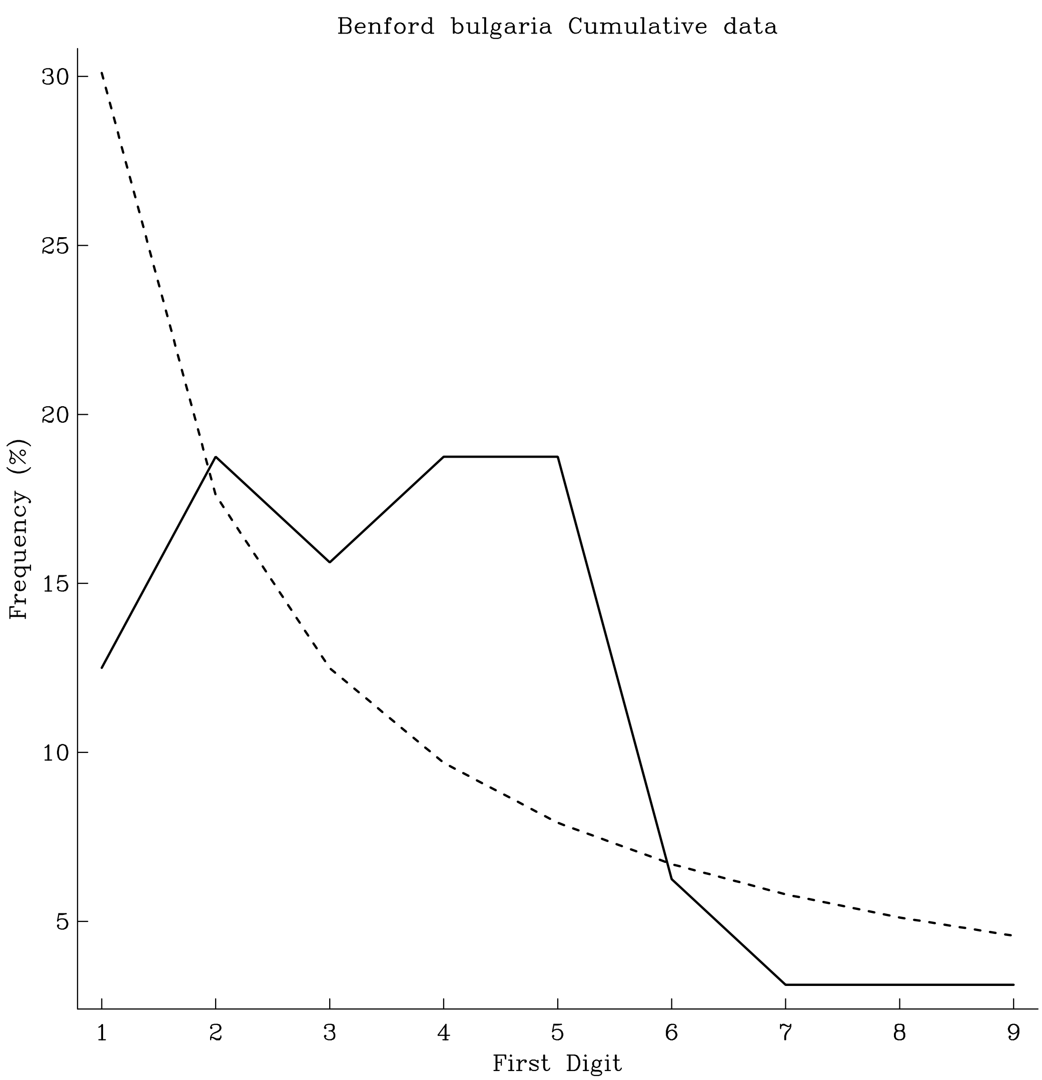

Benford austria Cumulative data
Frequency (%)
45
40
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit



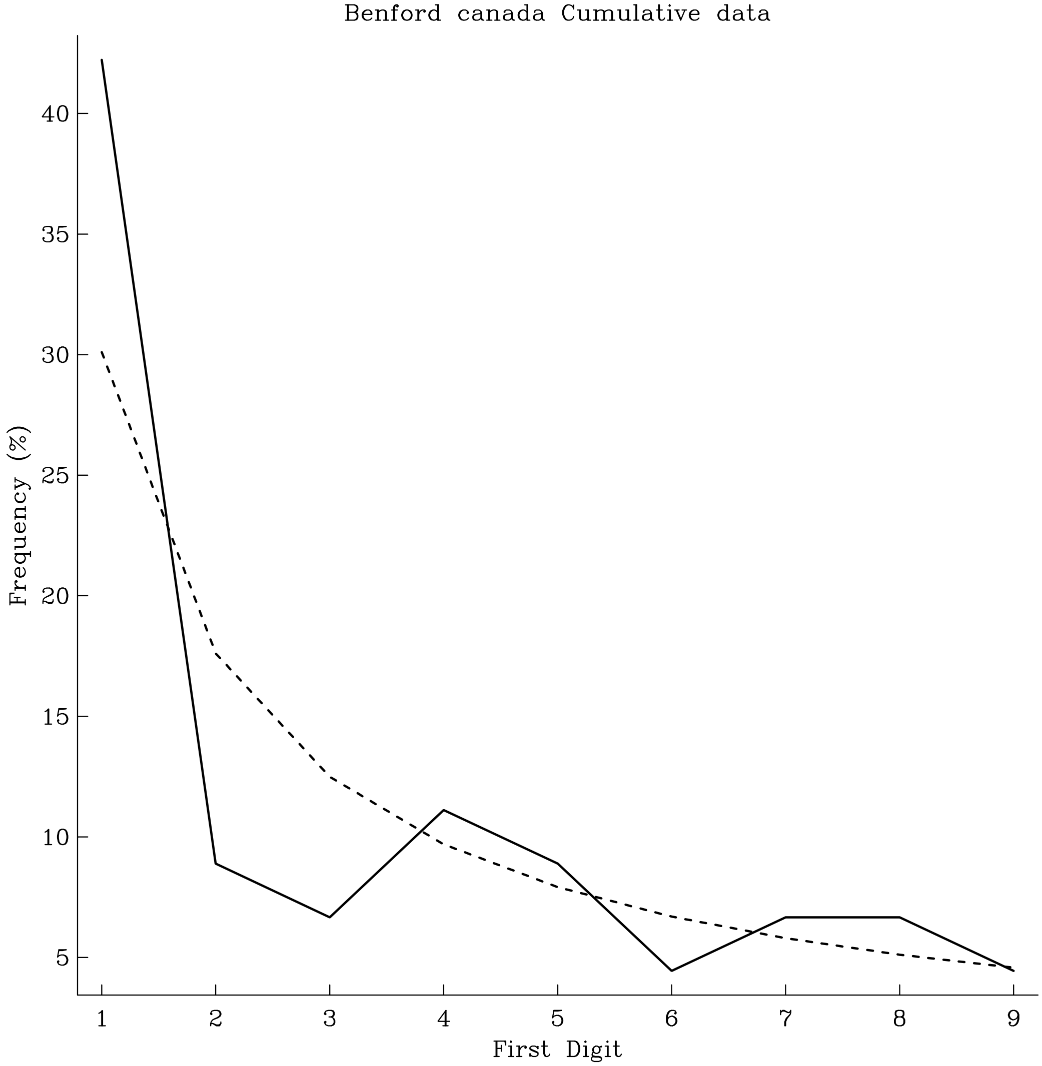
Benford belgium Cumulative data
Frequency (%)
40
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

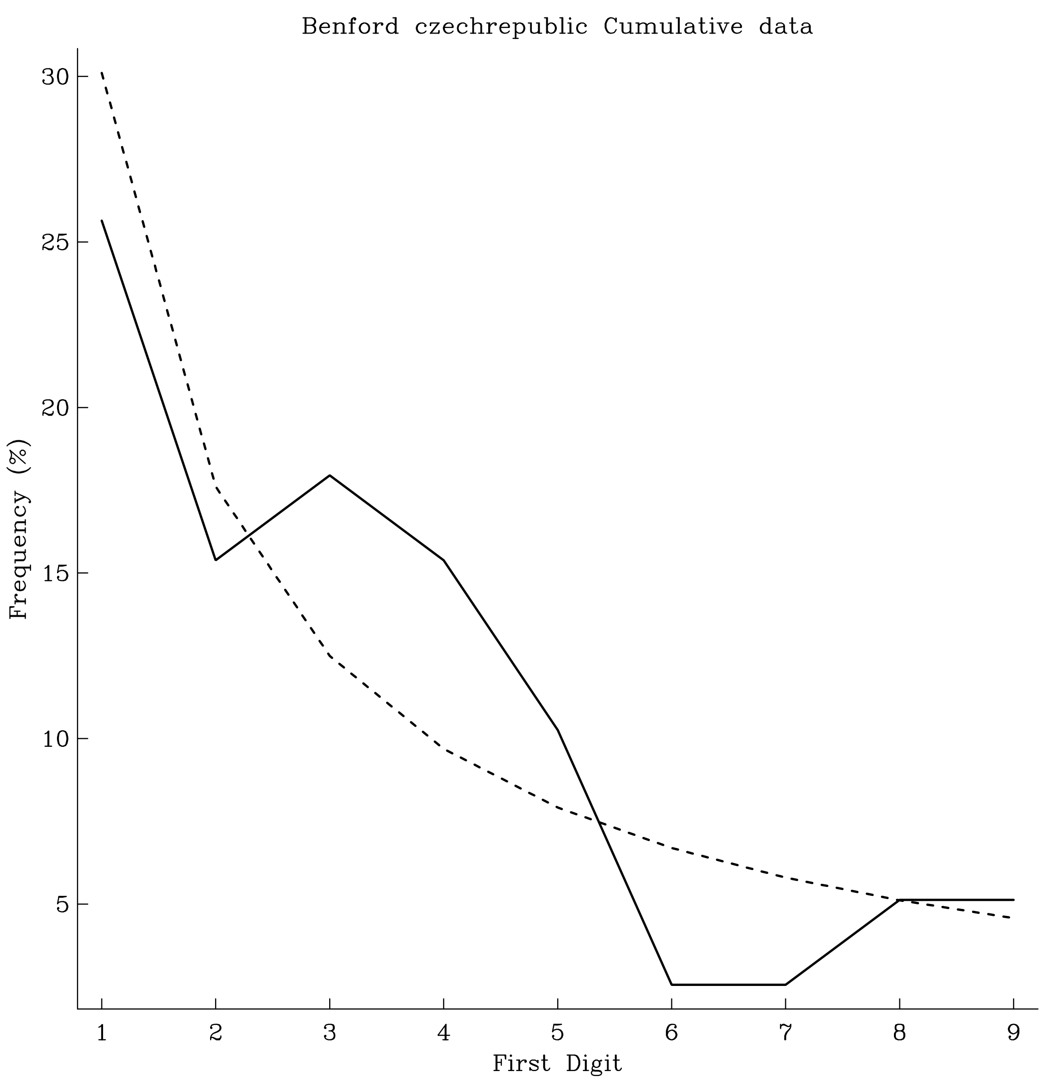


Benford brazil Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

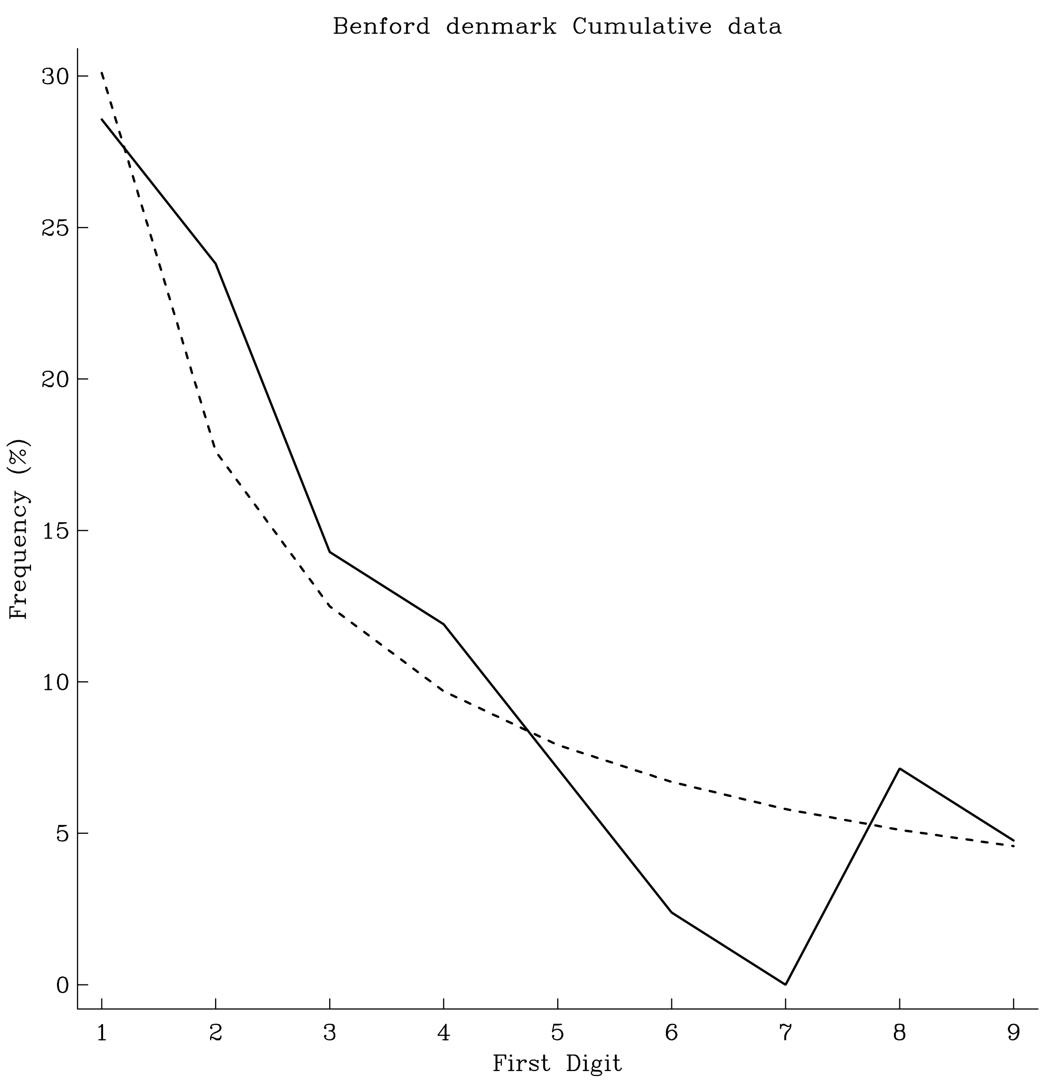

Benford bulgaria Cumulative data
Frequency (%)
First Digit
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

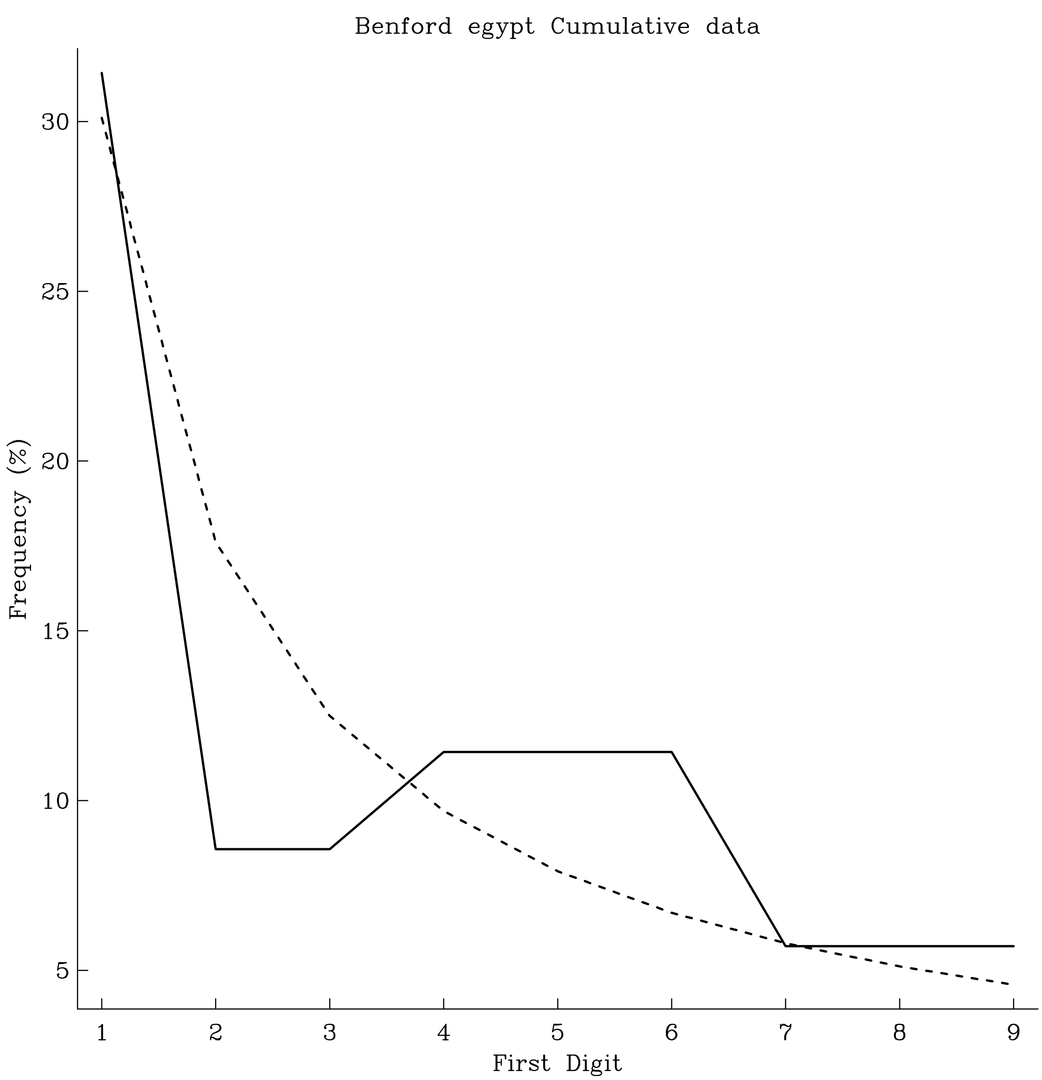
Benford canada Cumulative data



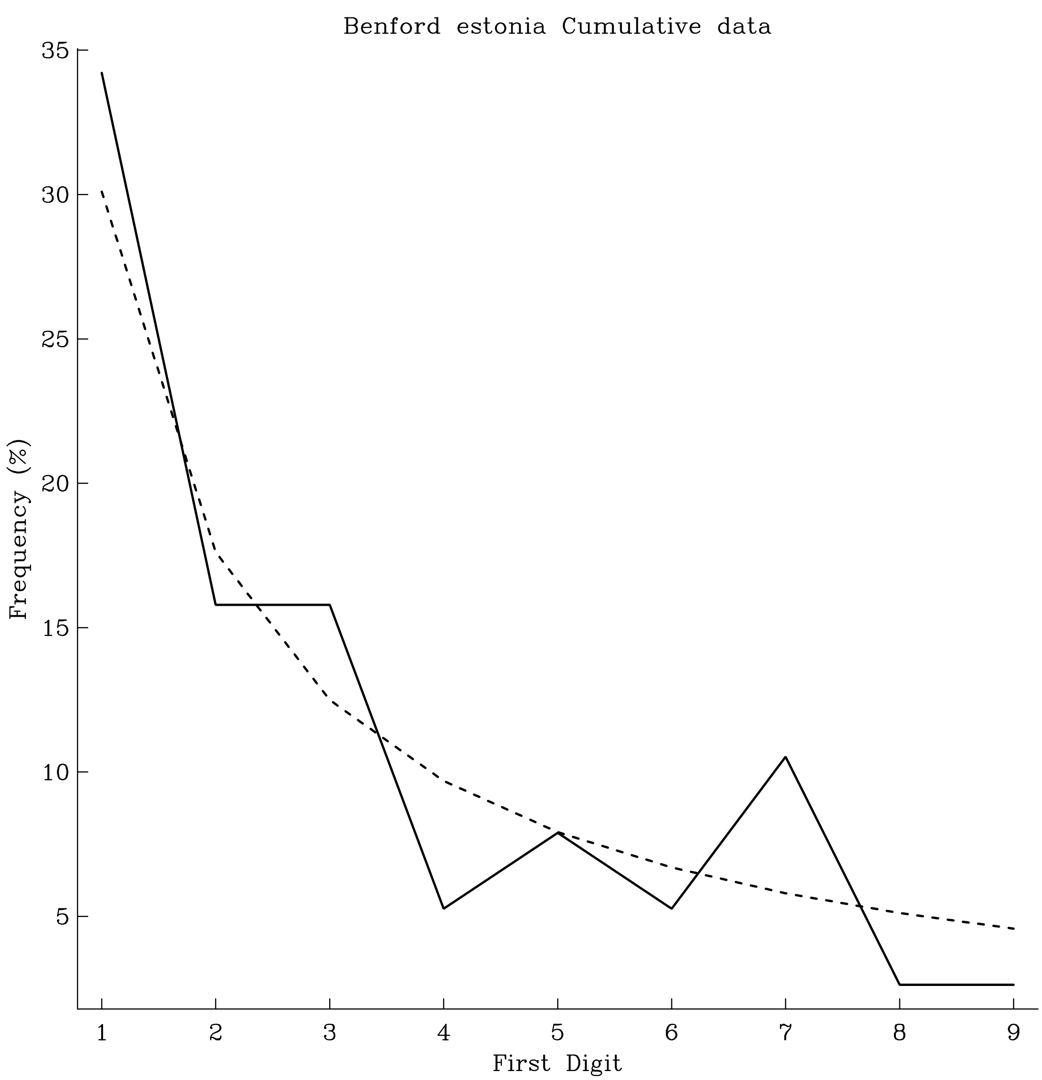


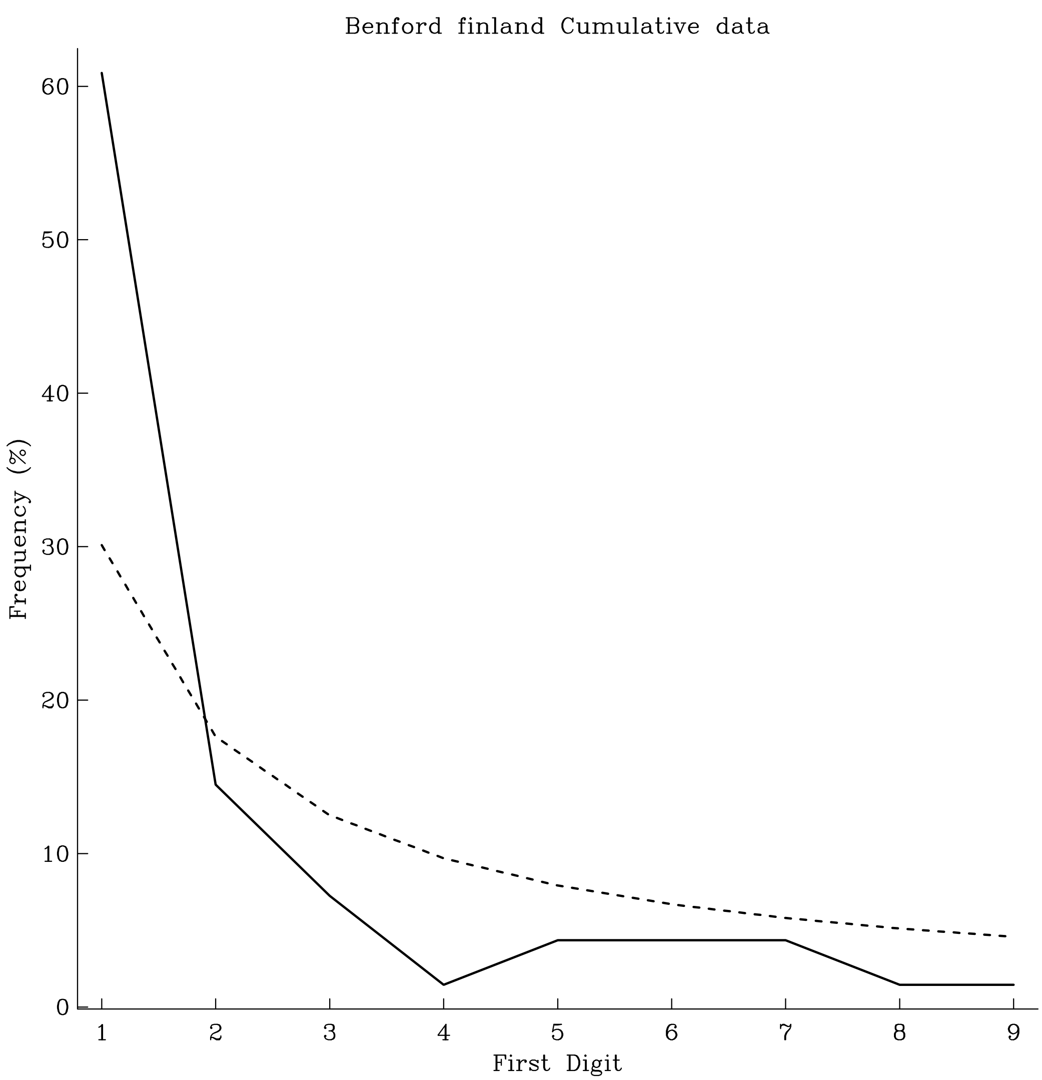
Benford czechrepublic Cumulative data
30
25
20
15
10
5
Frequency (%)
1
2
3
4
5
6
7
8
9
First Digit


Benford denmark Cumulative data
Frequency (%)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First Digit

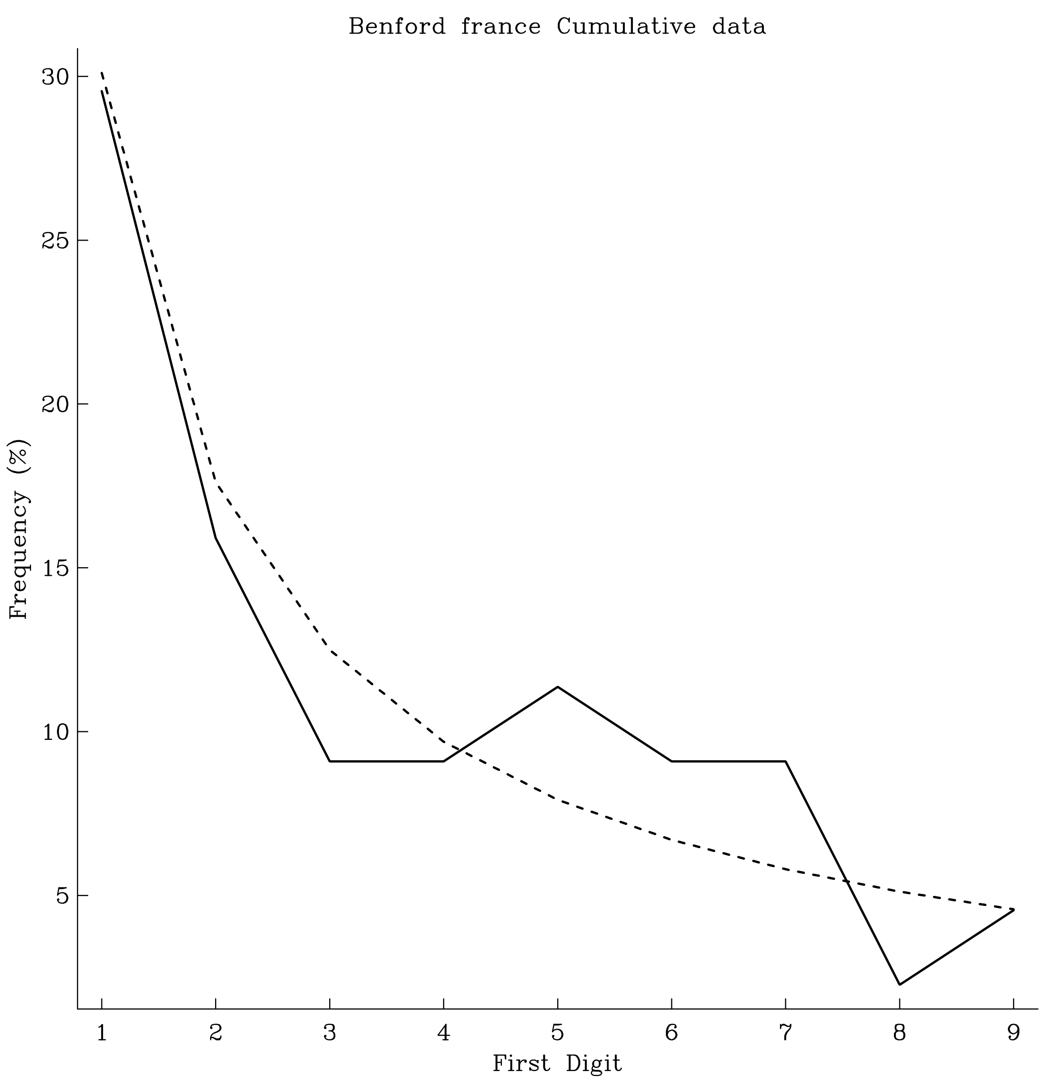


Benford egypt Cumulative data
30
25
20
15
10
5
Frequency (%)
1
2
3
4
5
6
7
8
9
First Digit

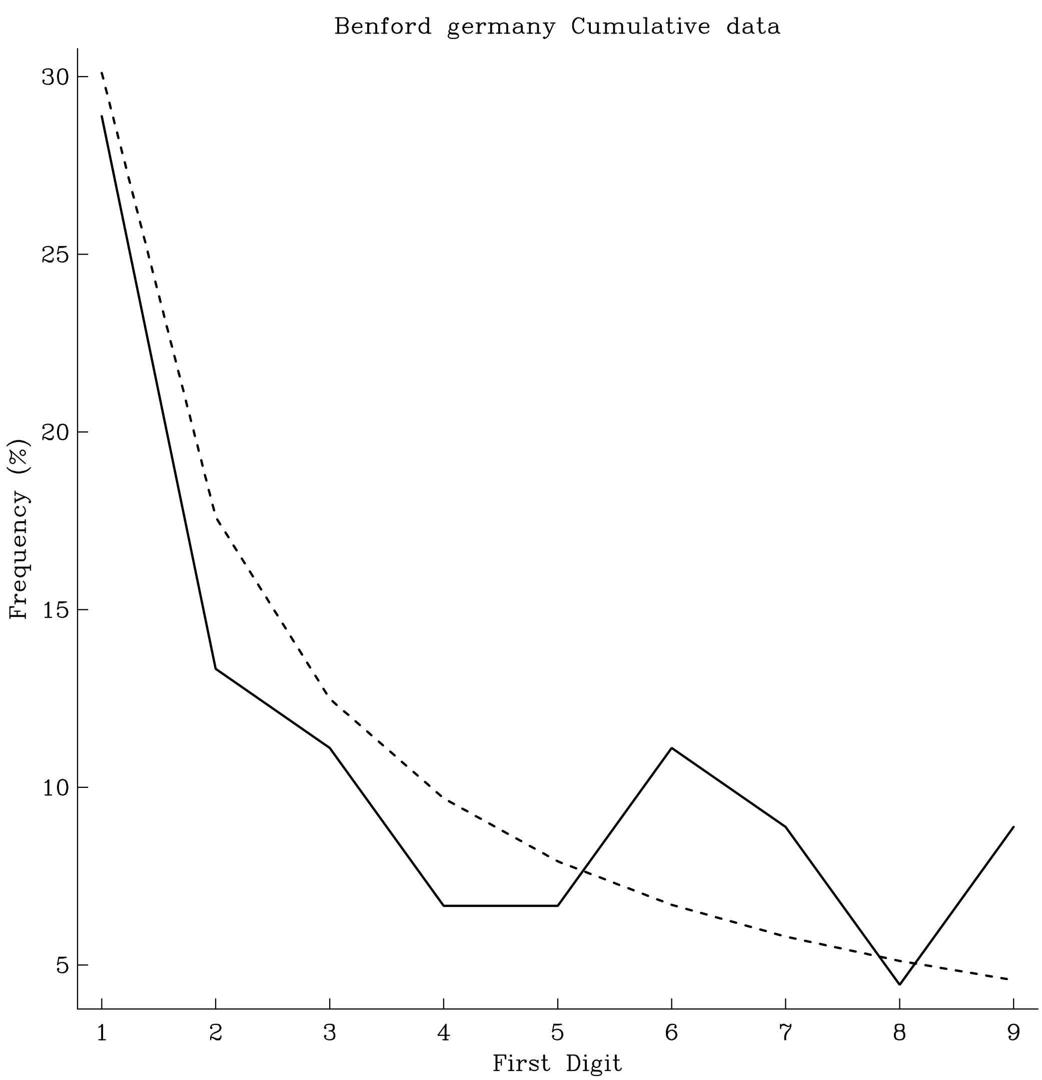

Benford estonia Cumulative data
35
30
25
20
15
10
5
Frequency (%)
1
2
3
4
5
6
7
8
9
First Digit



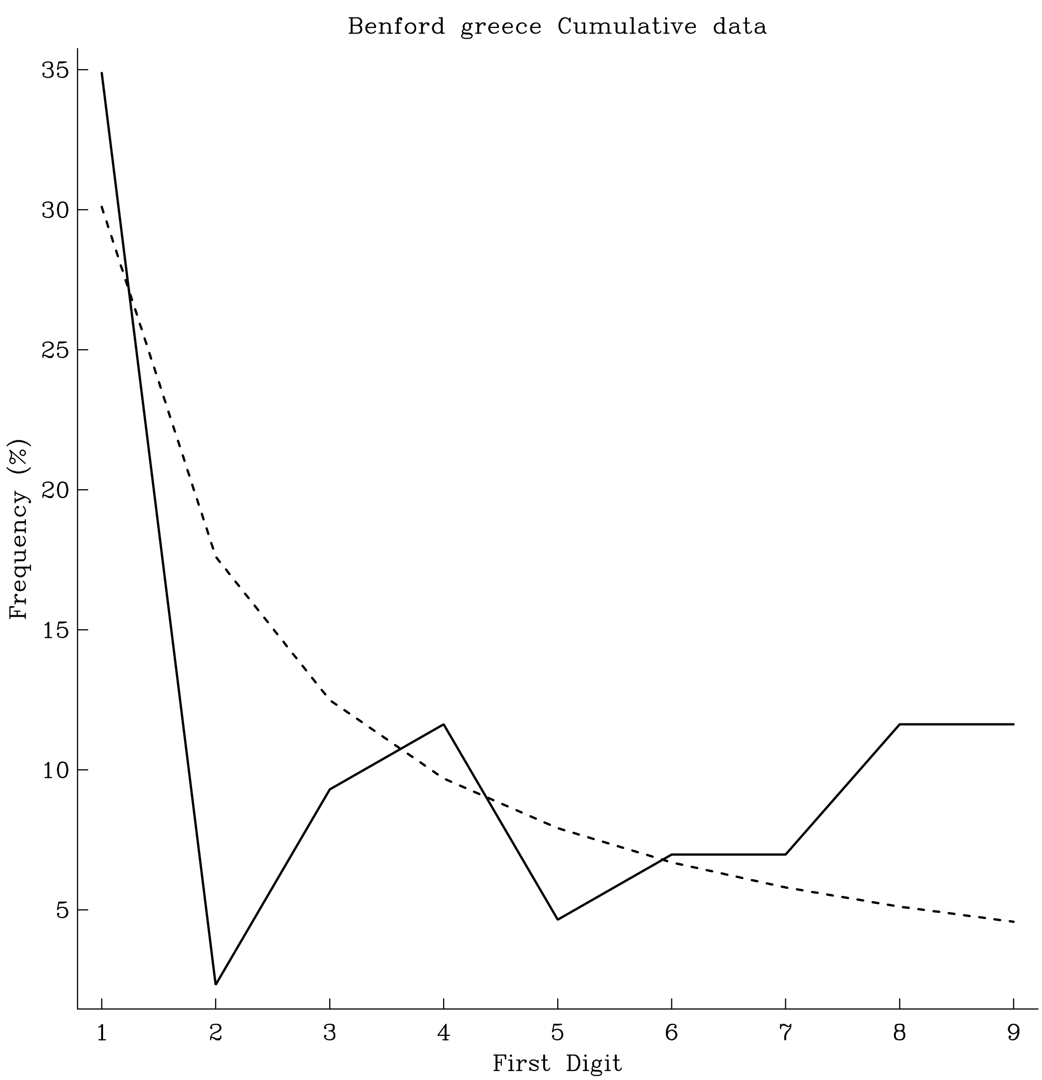
Benford finland Cumulative data
Frequency (%)
First Digit
0
10
20
30
40
50
60
1
2
3
4
5
6
7
8
9



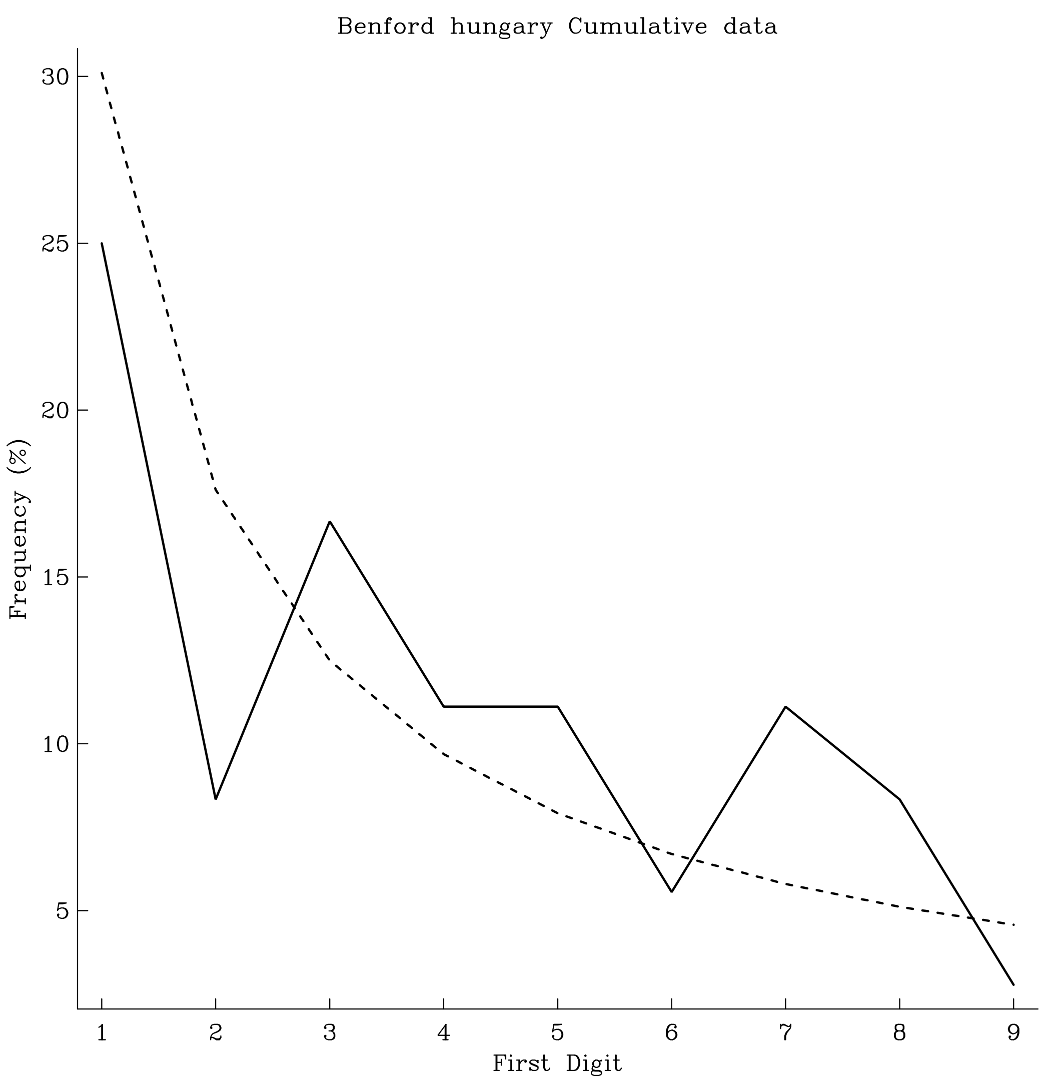
Benford france Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

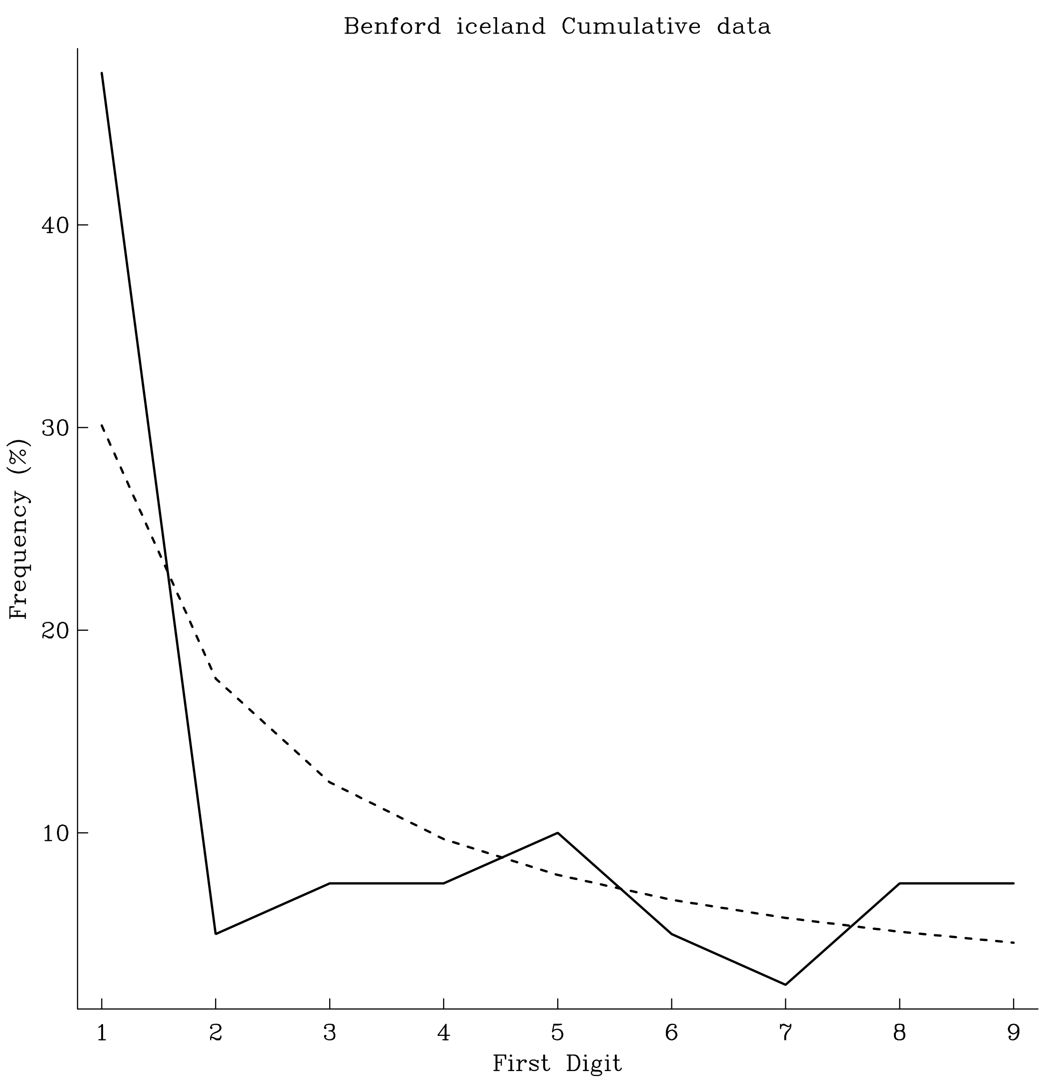


Benford germany Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

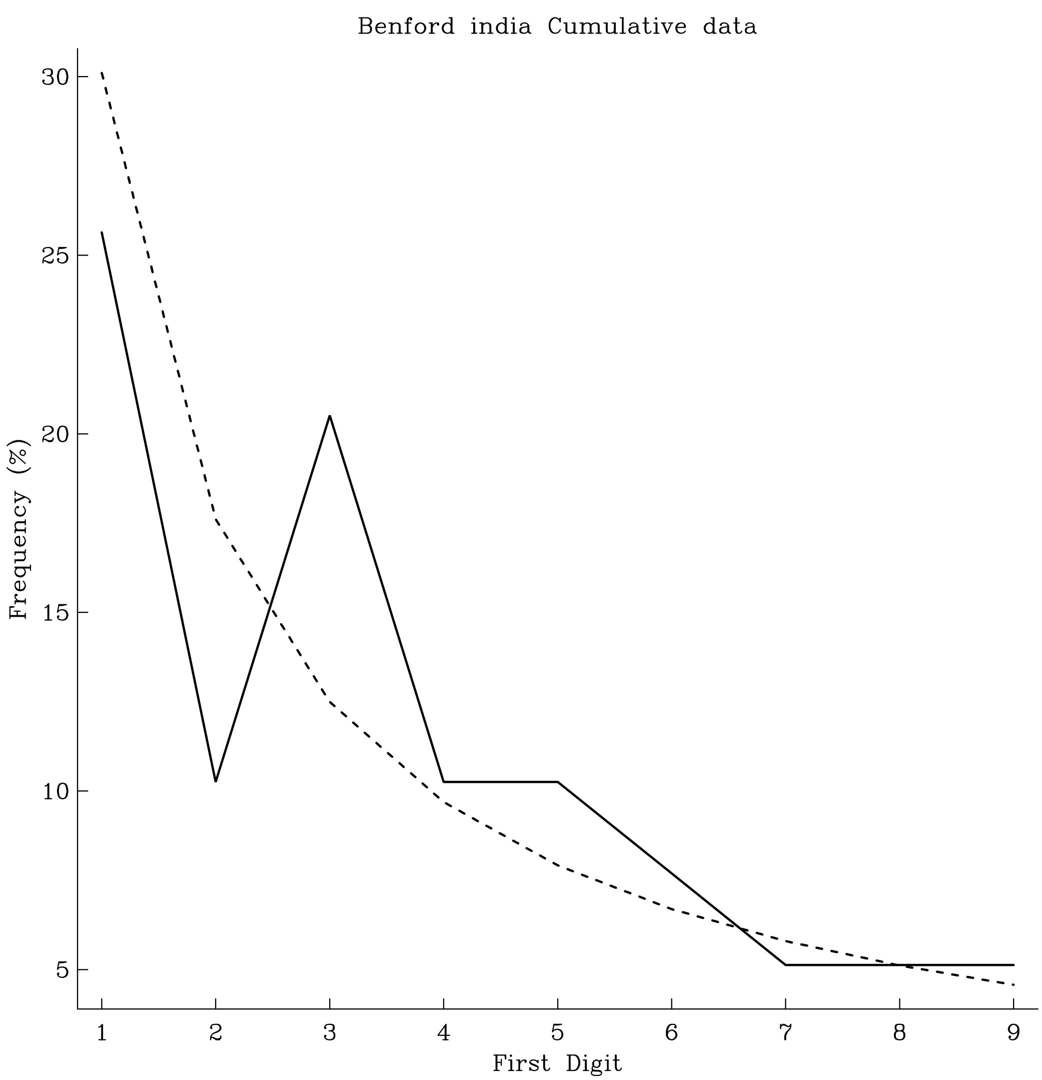


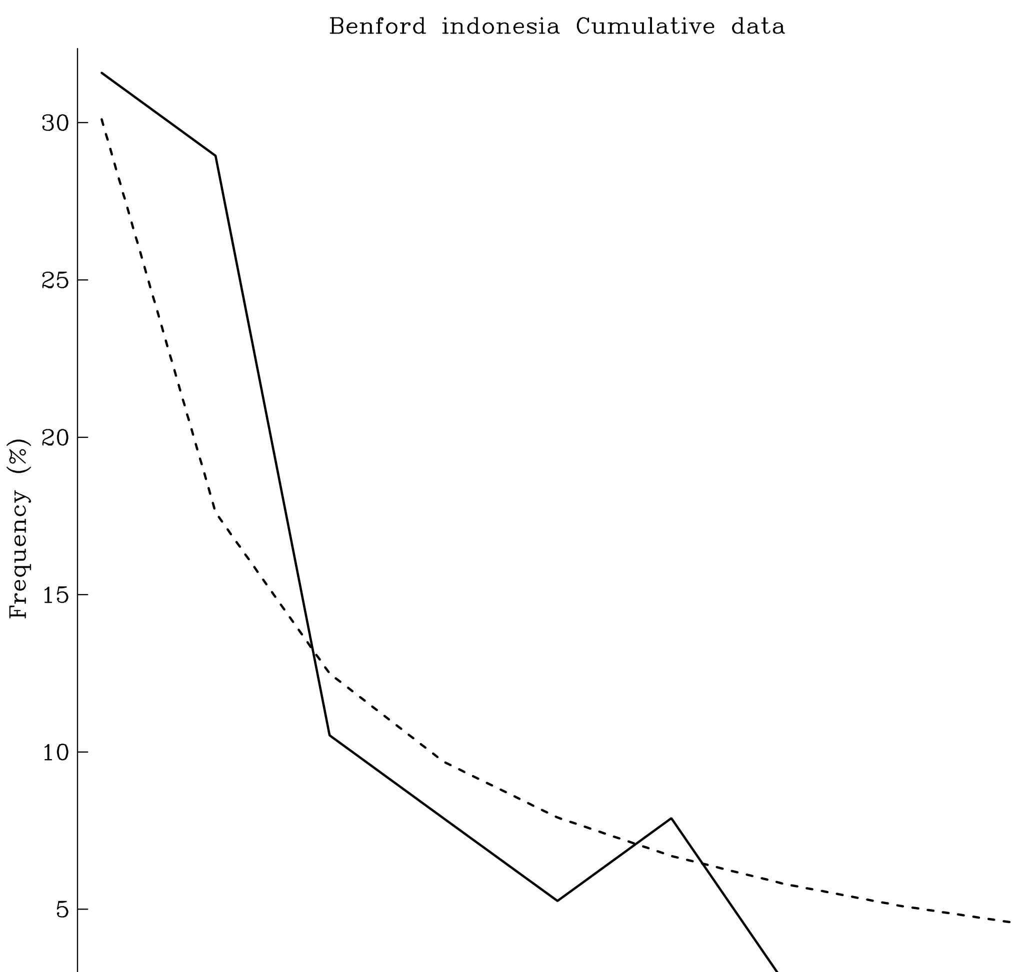
Benford greece Cumulative data
Frequency (%)
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit


Benford hungary Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

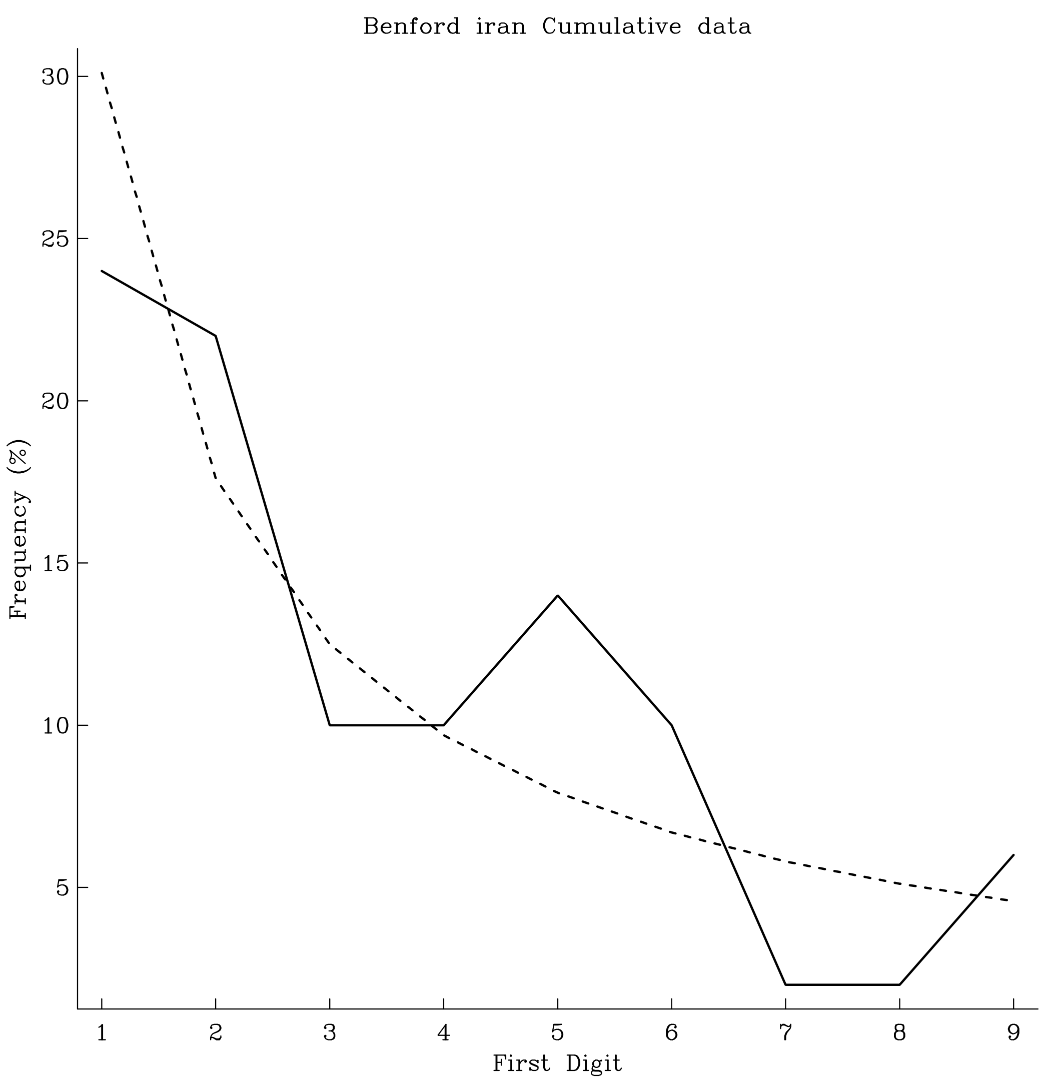


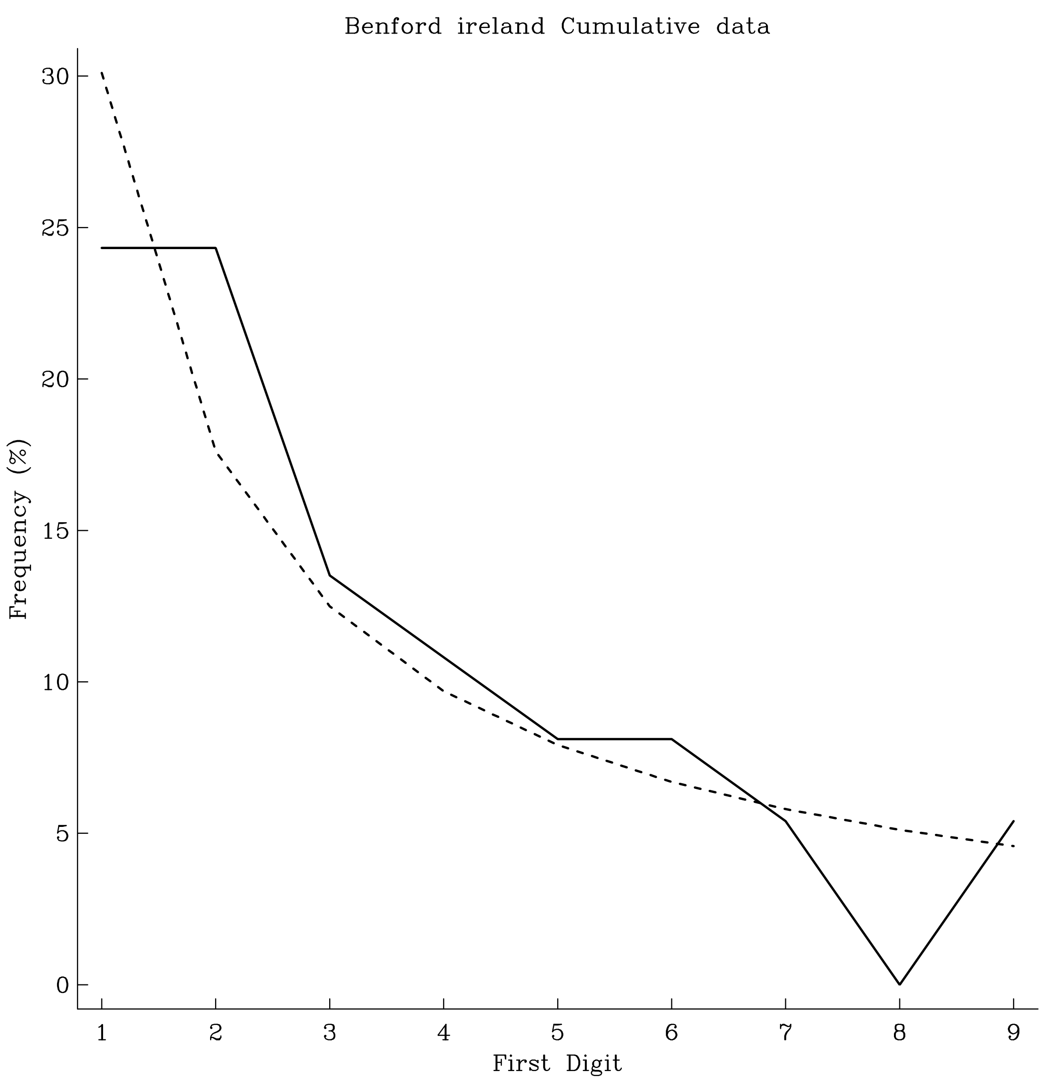
Benford iceland Cumulative data
Frequency (%)
40
30
20
10
1
2
3
4
5
6
7
8
9
First Digit


Benford india Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

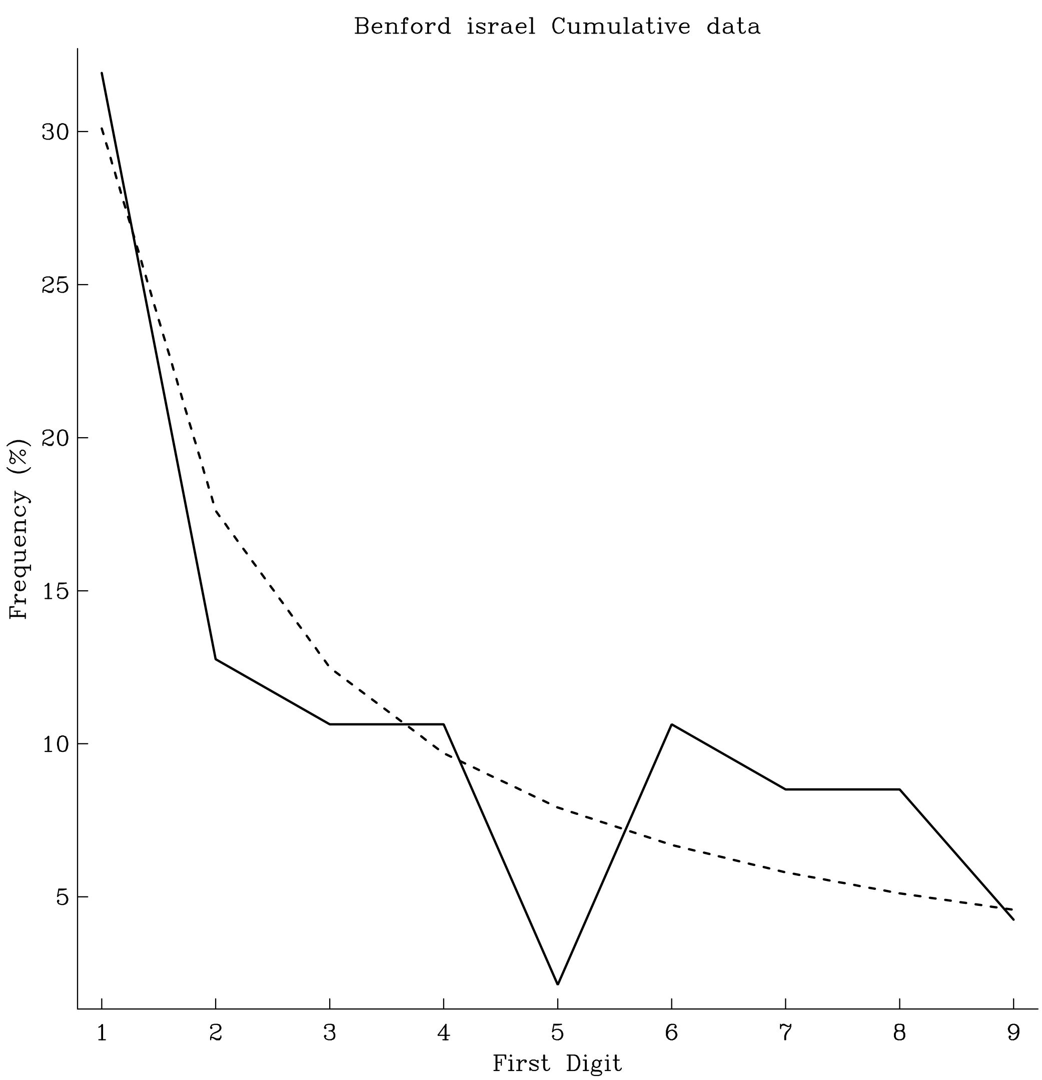


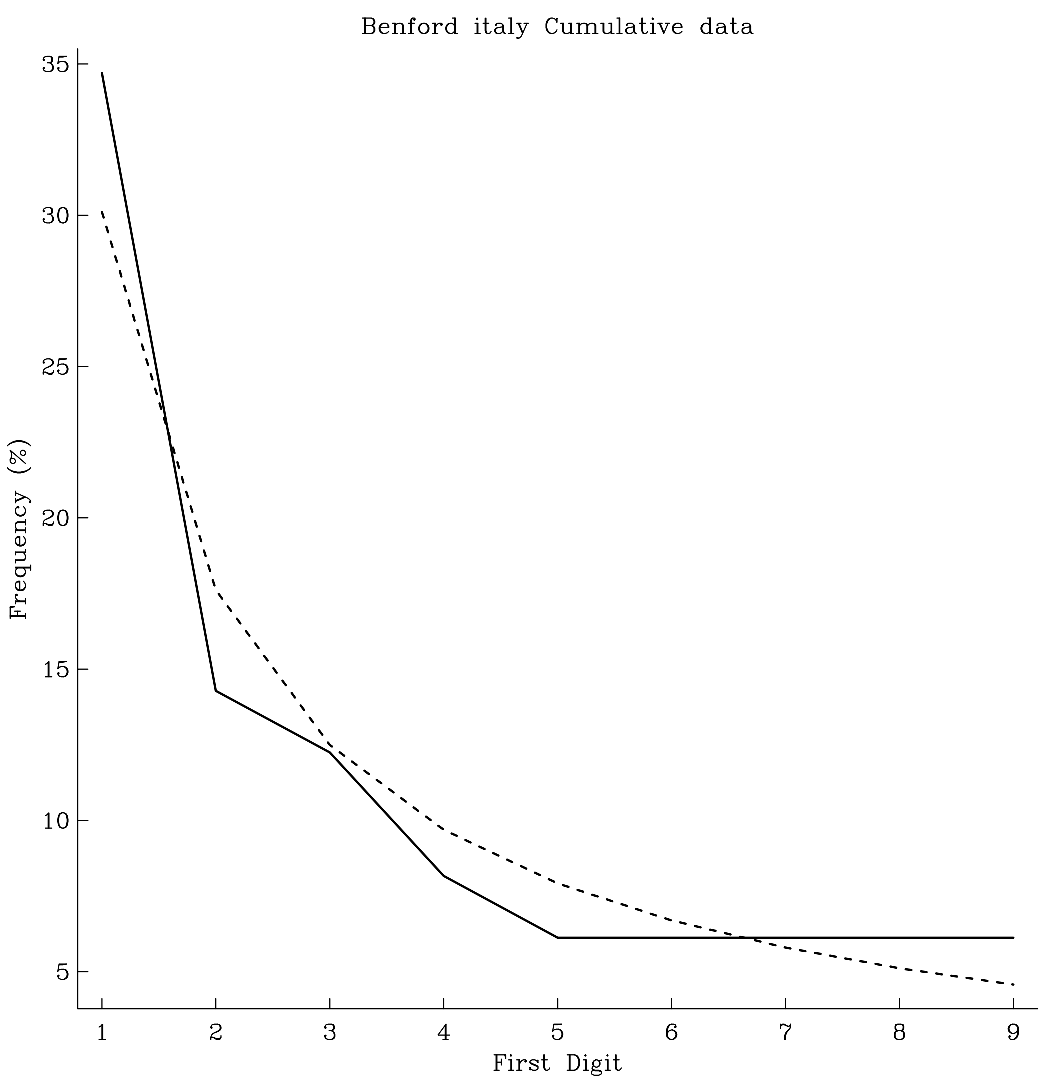
Benford indonesia Cumulative data
Frequency (%)
First Digit
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

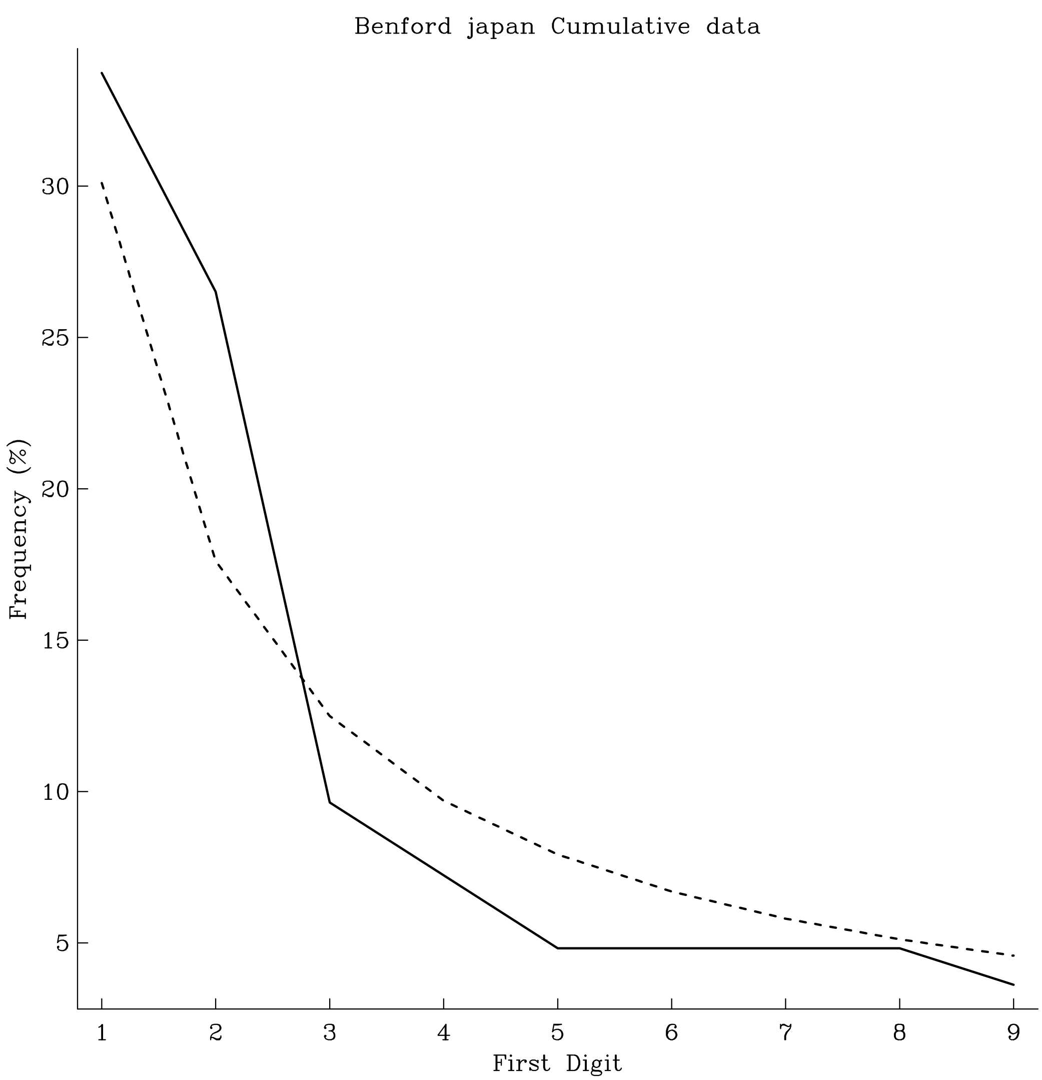

Benford iran Cumulative data
Frequency (%)
First Digit
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9



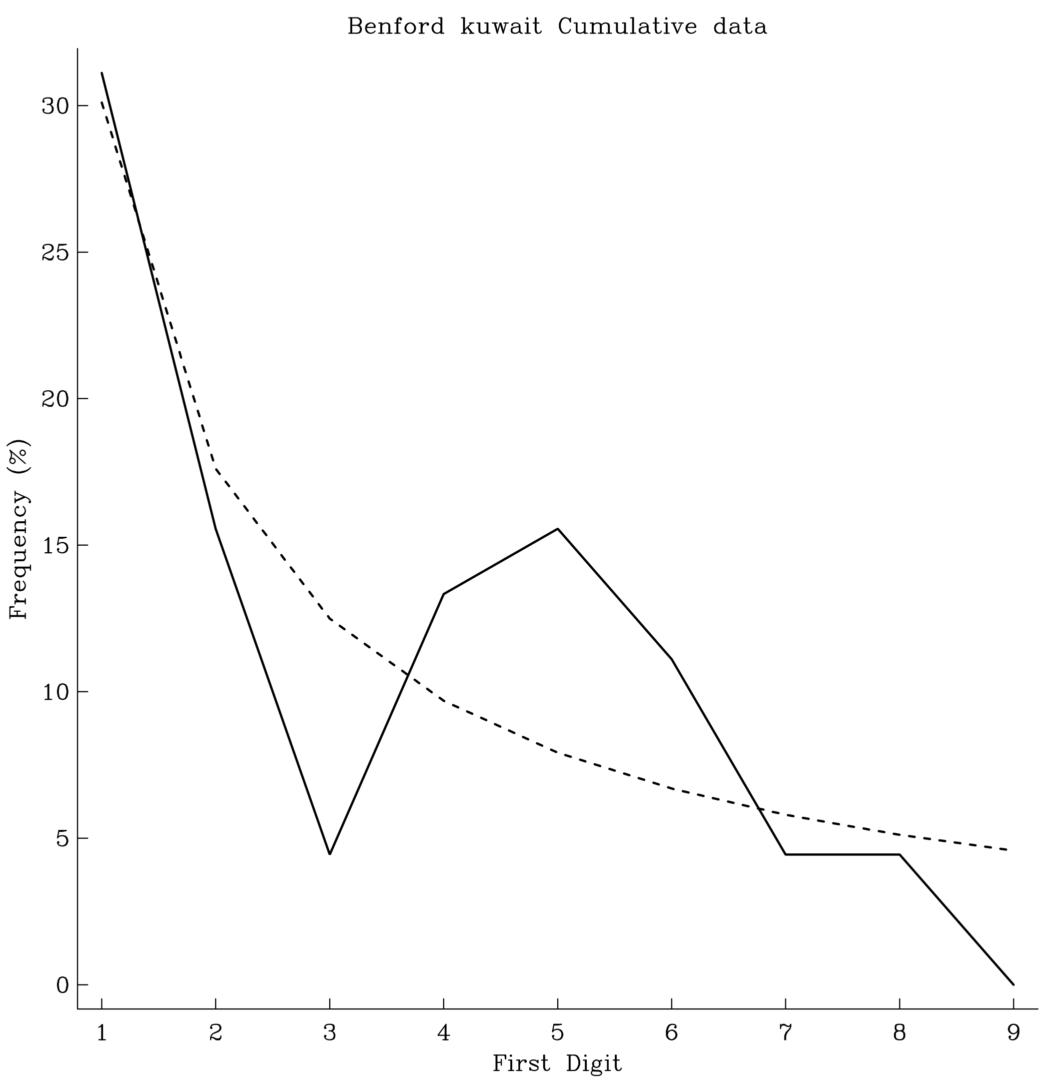
Benford ireland Cumulative data
Frequency (%)
First Digit
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9


Benford israel Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

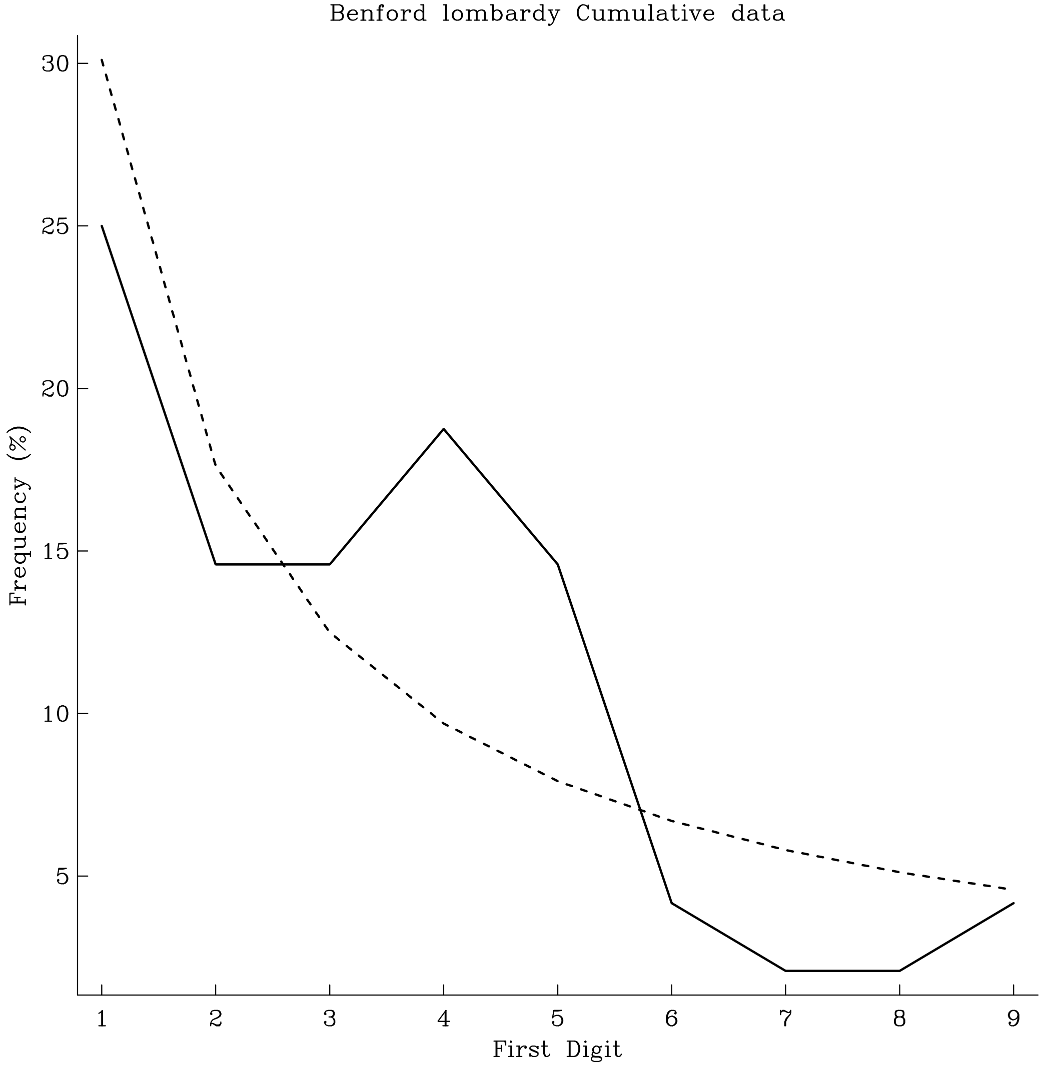

Benford italy Cumulative data
Frequency (%)
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

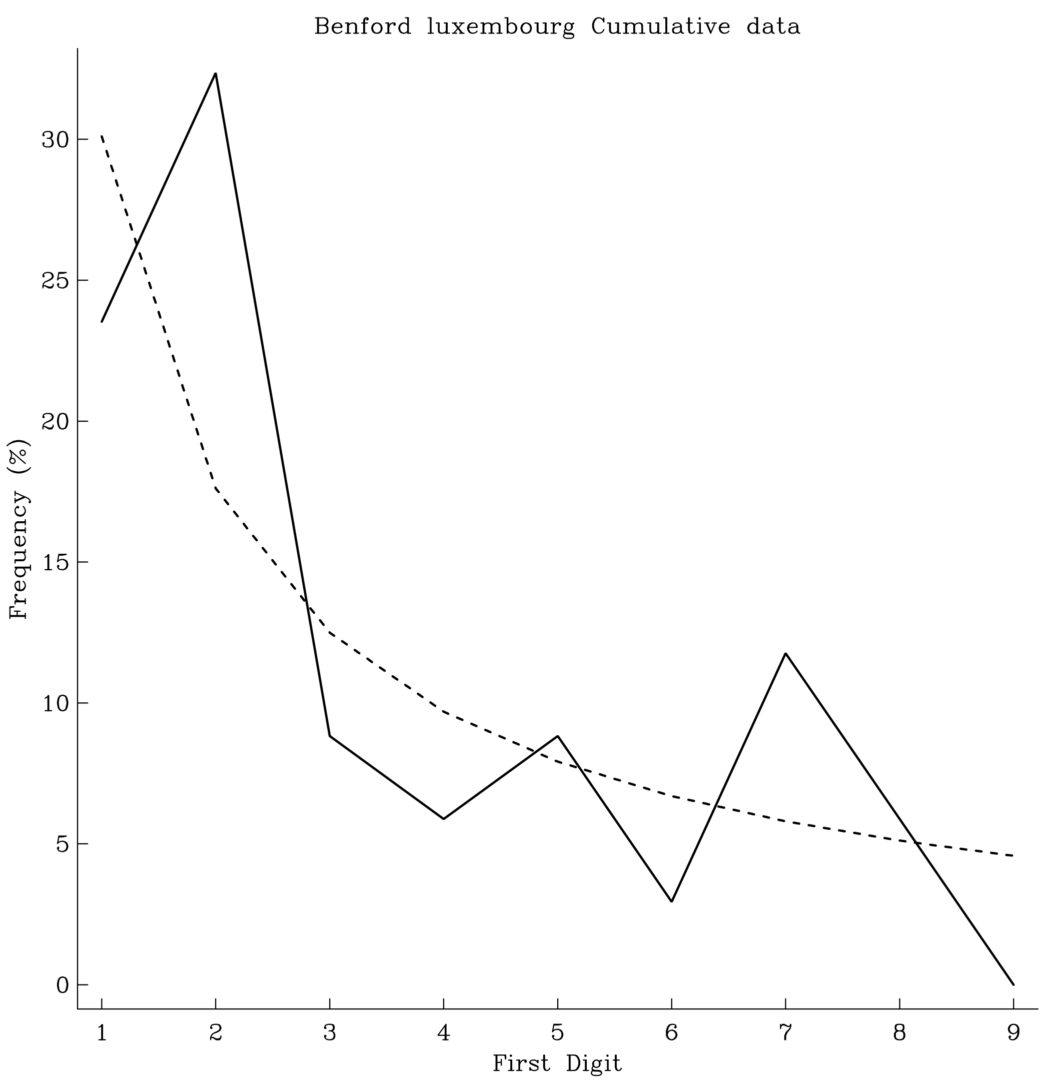


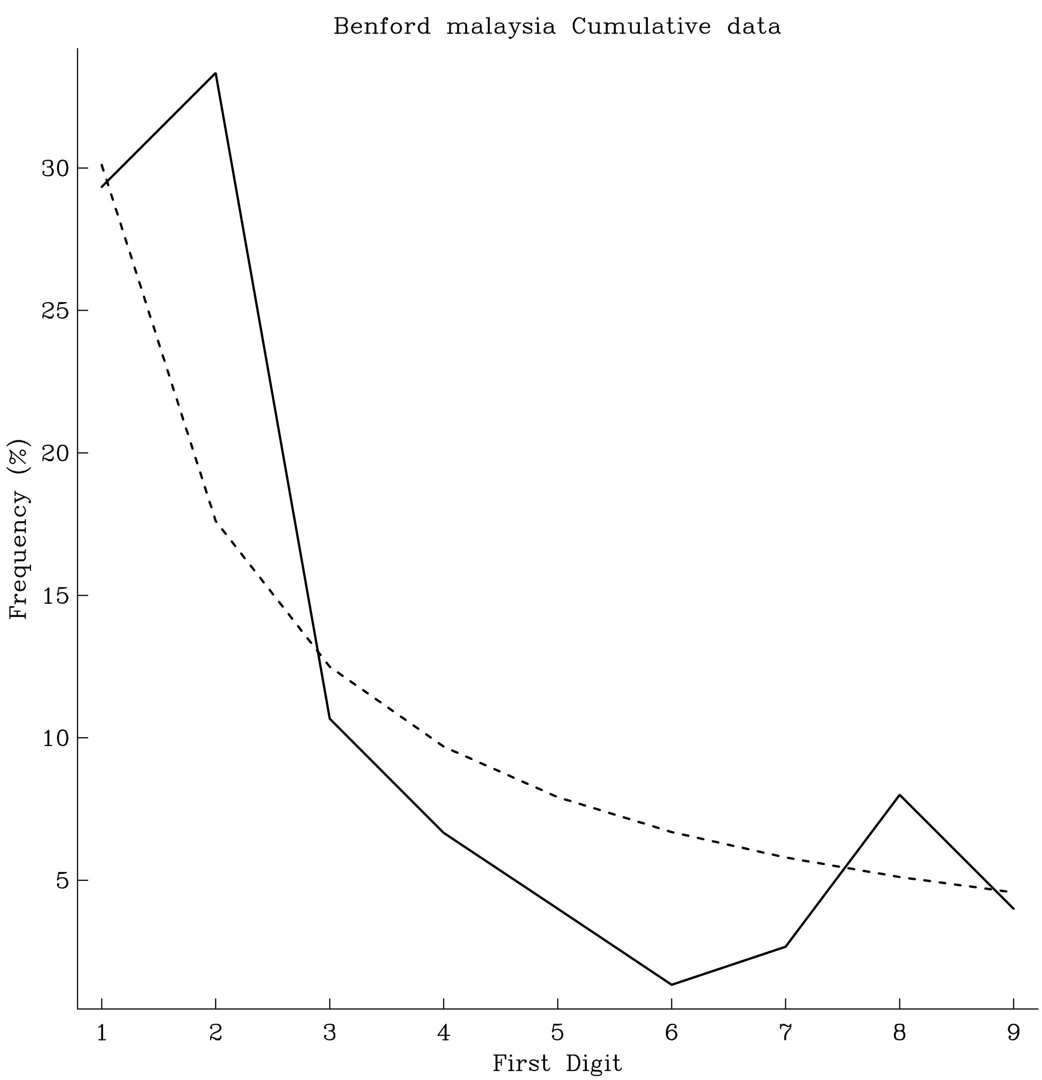
Benford japan Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit


Benford kuwait Cumulative data
Frequency (%)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First Digit

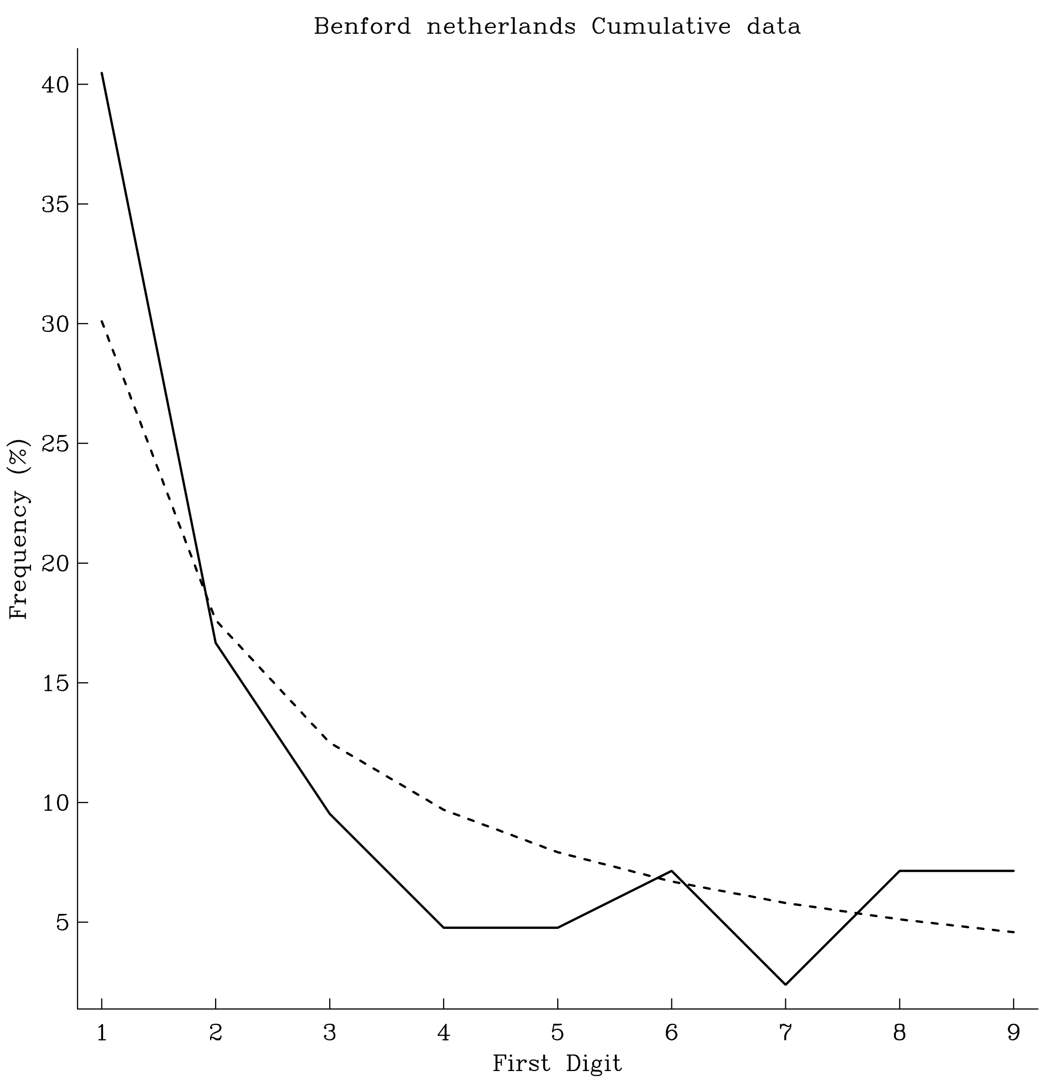
Benford lombardy Cumulative data



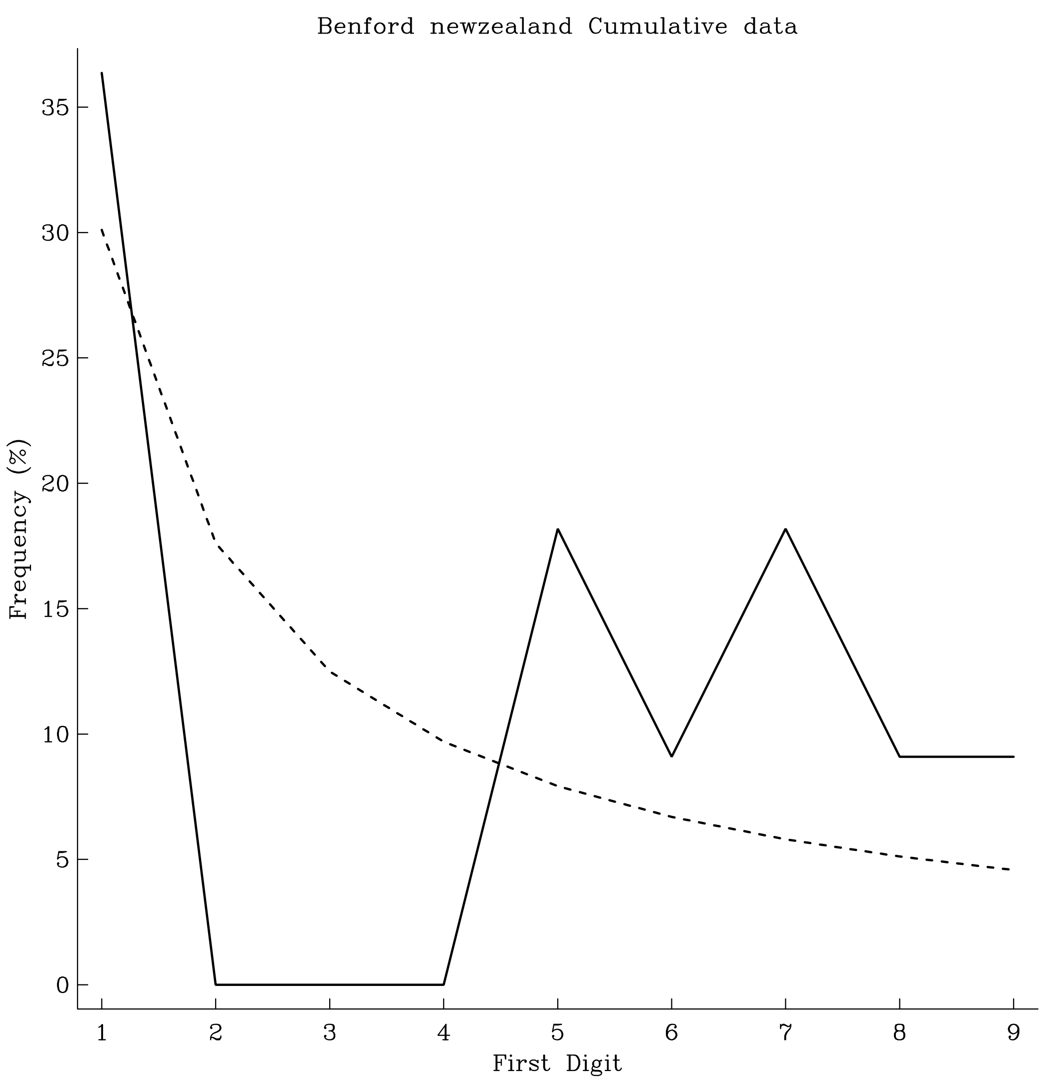

Benford luxembourg Cumulative data
Frequency (%)
First Digit
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9



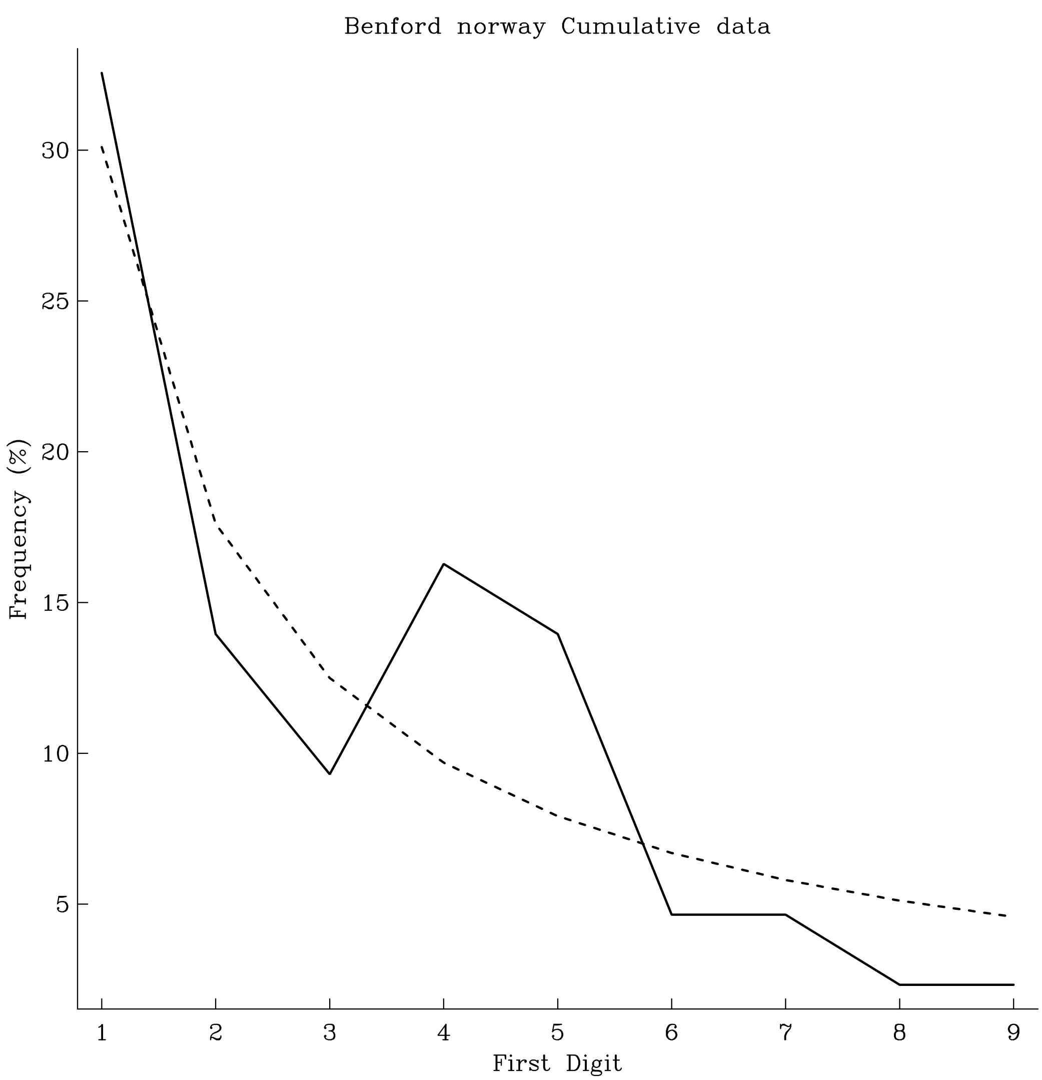
Benford malaysia Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

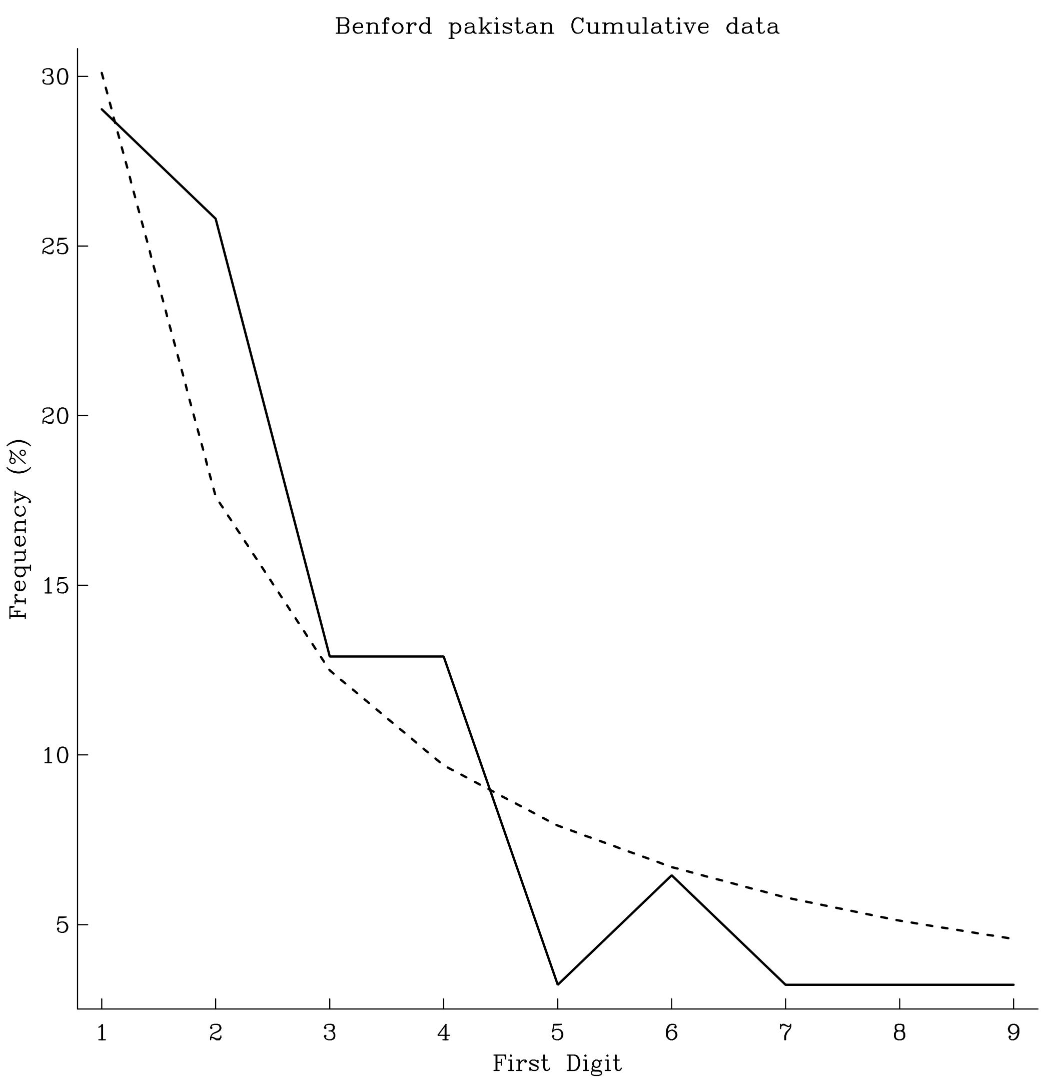


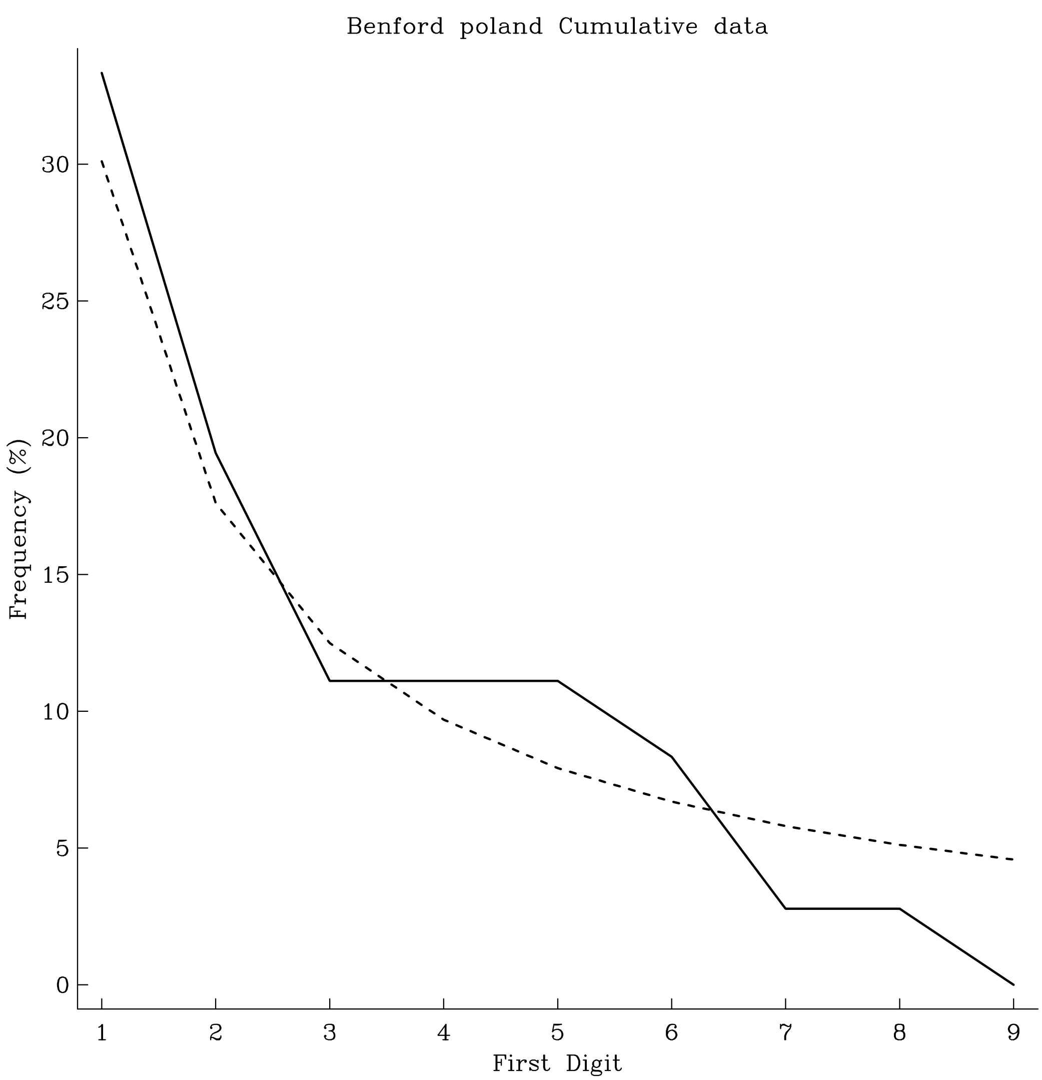
Benford netherlands Cumulative data
Frequency (%)
40
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

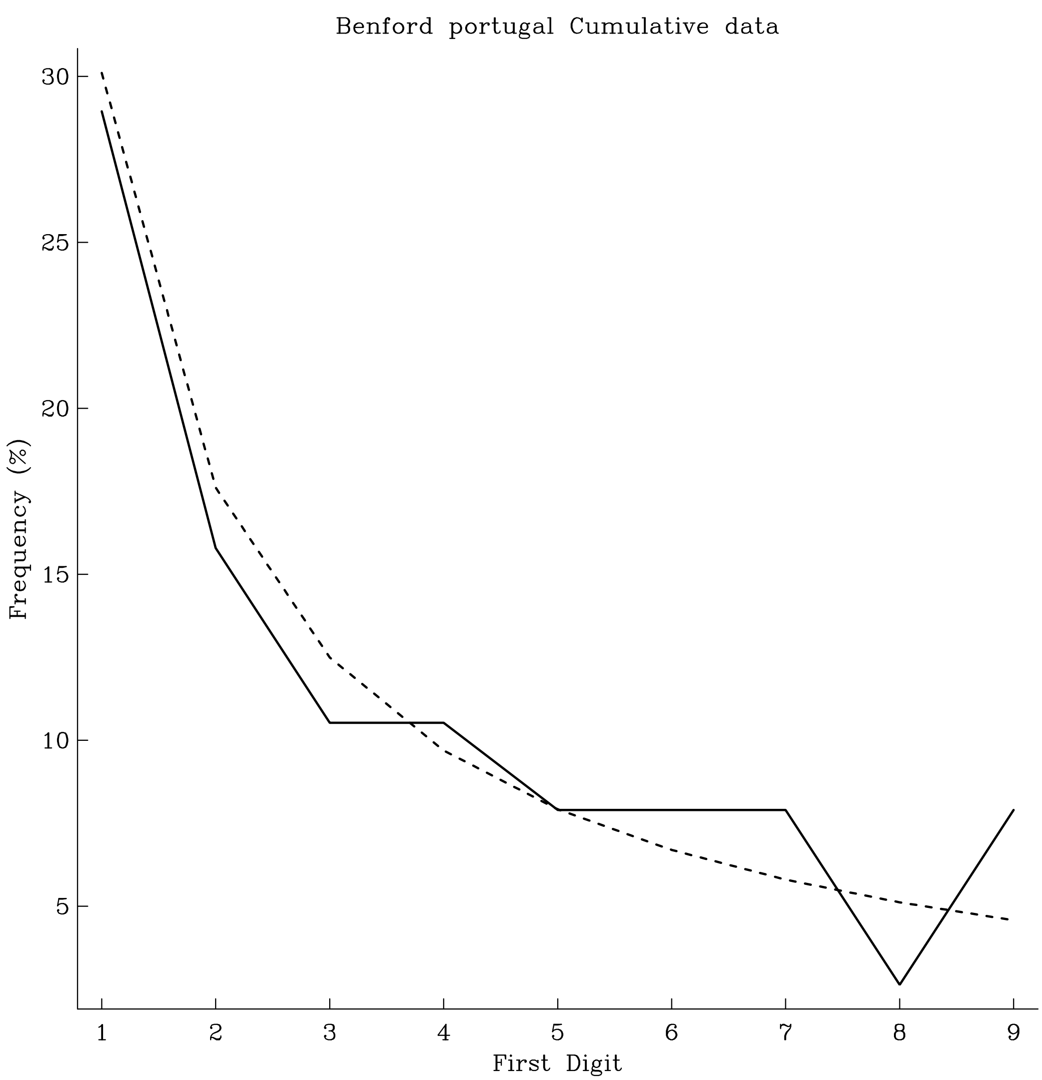


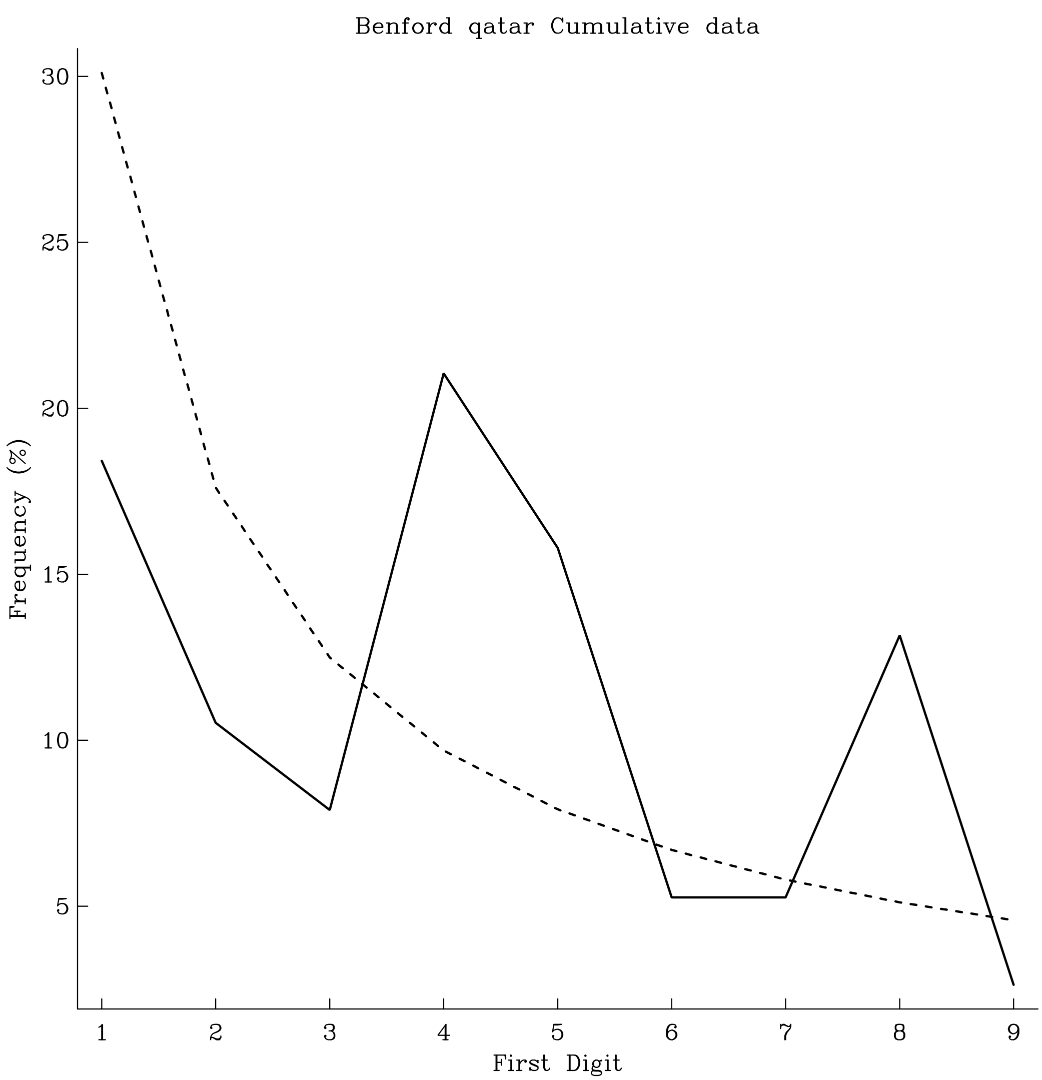
Benford newzealand Cumulative data
Frequency (%)
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First Digit

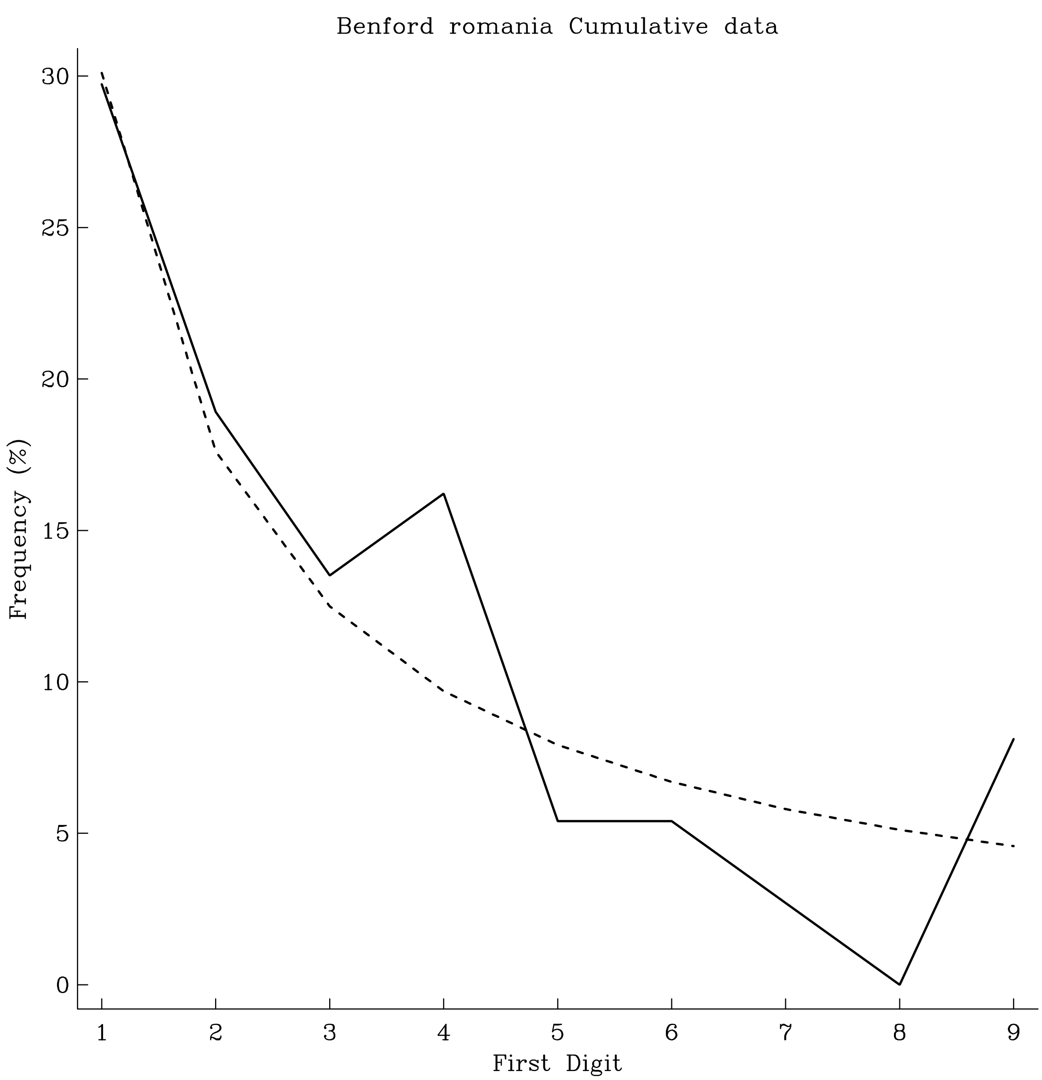

Benford norway Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit


Benford pakistan Cumulative data
30
25
20
15
10
5
Frequency (%)
1
2
3
4
5
6
7
8
9
First Digit

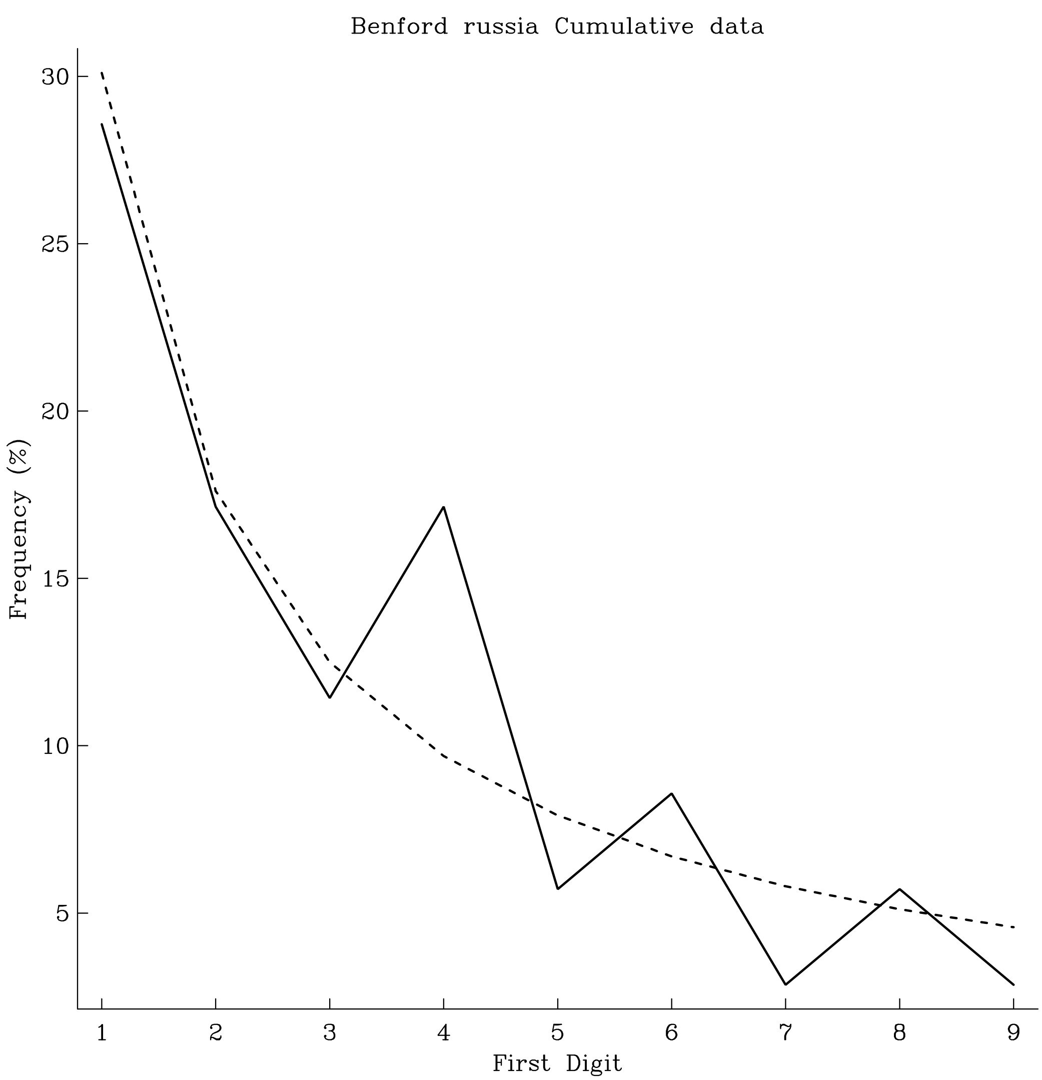

Benford poland Cumulative data
Frequency (%)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First Digit


Benford portugal Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

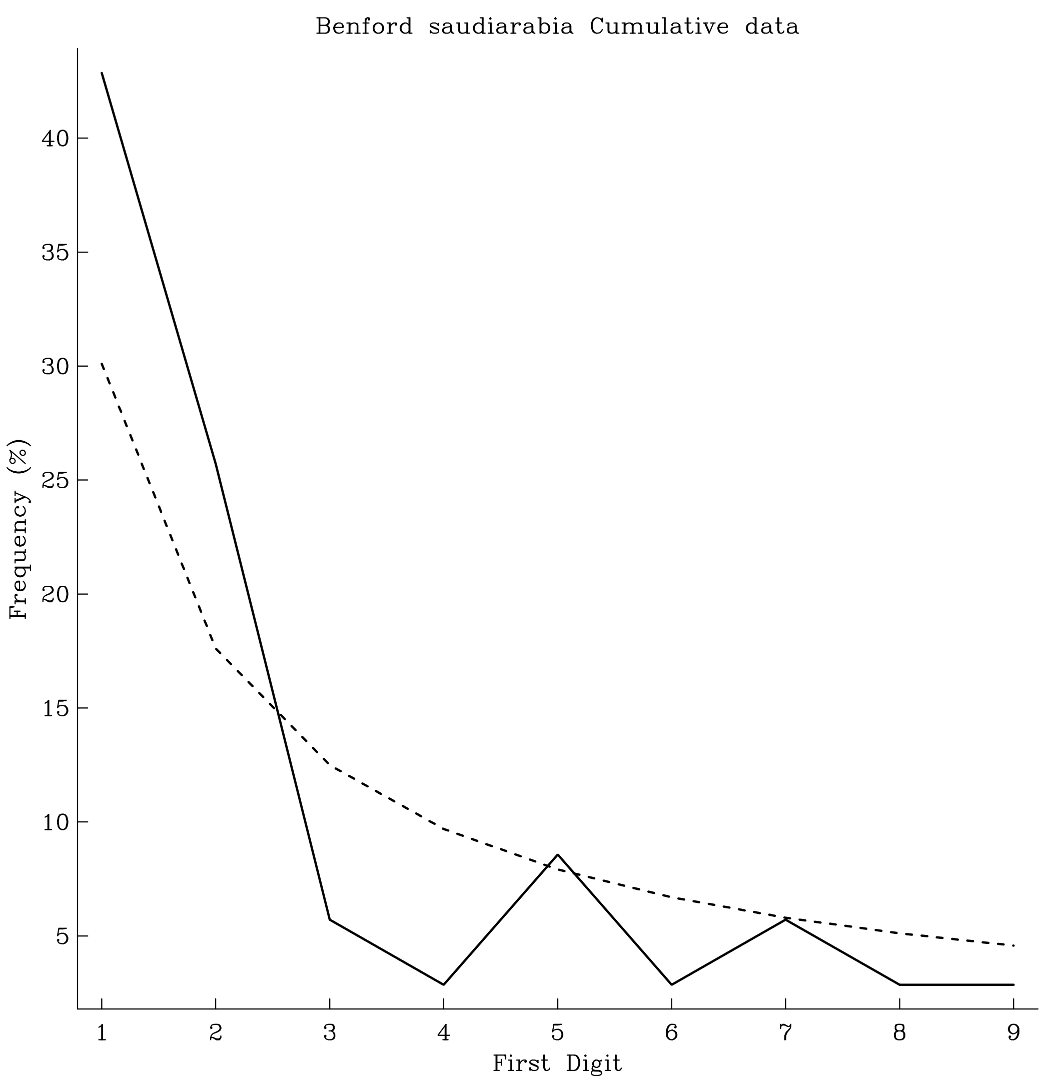

Benford qatar Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

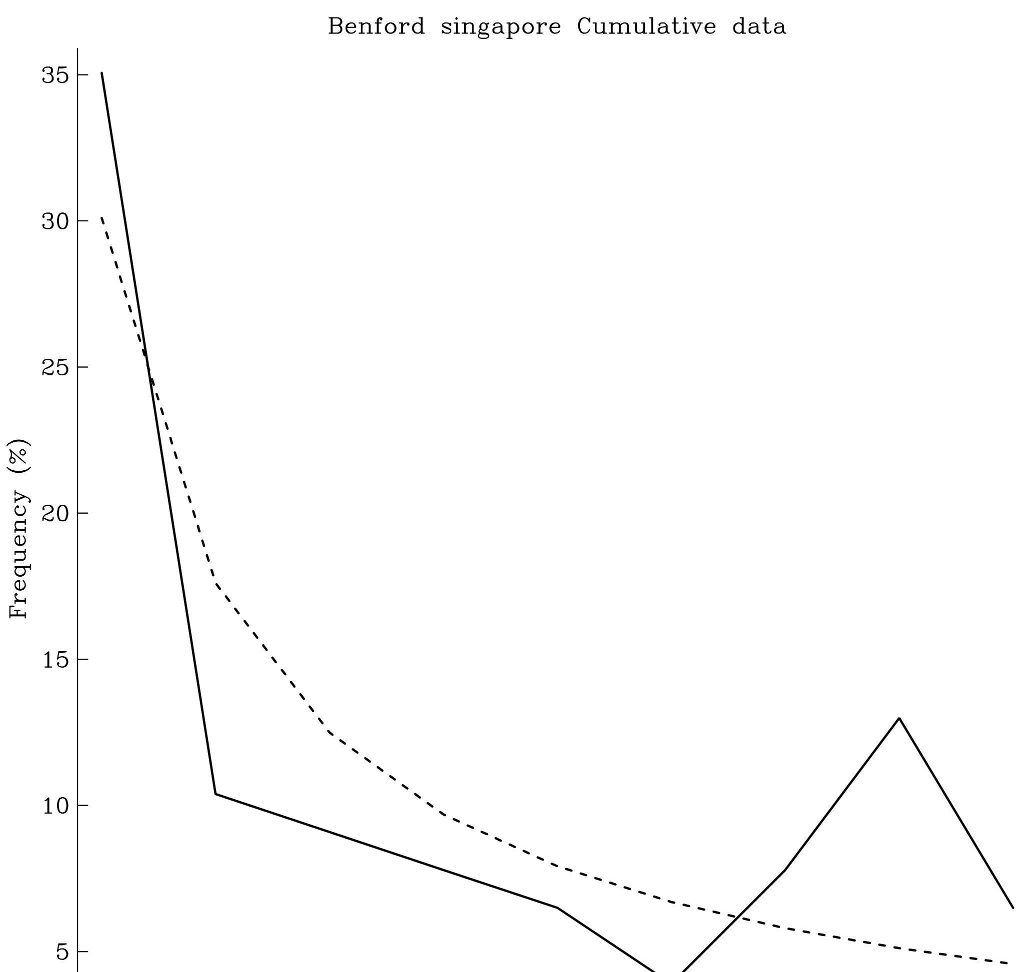


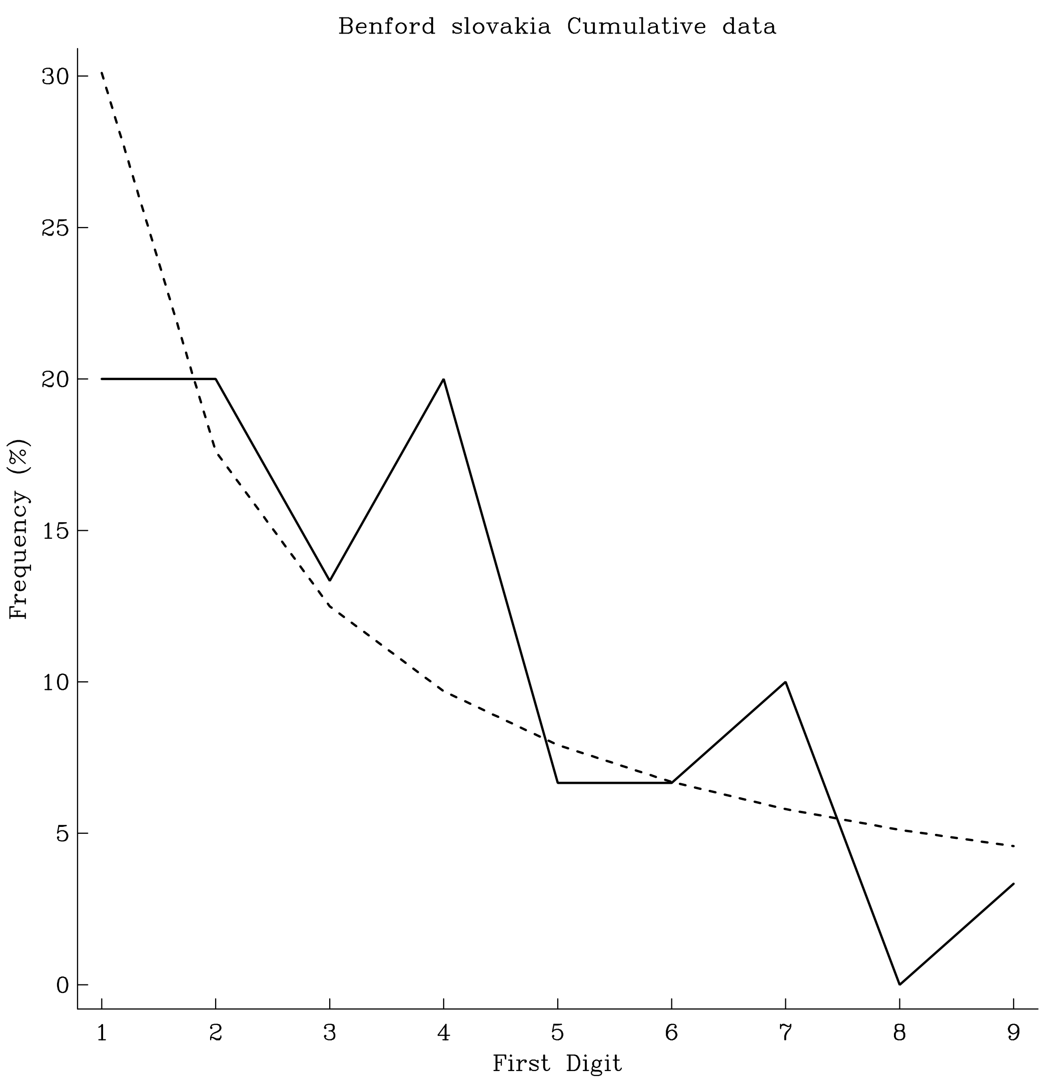
Benford romania Cumulative data
Frequency (%)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First Digit



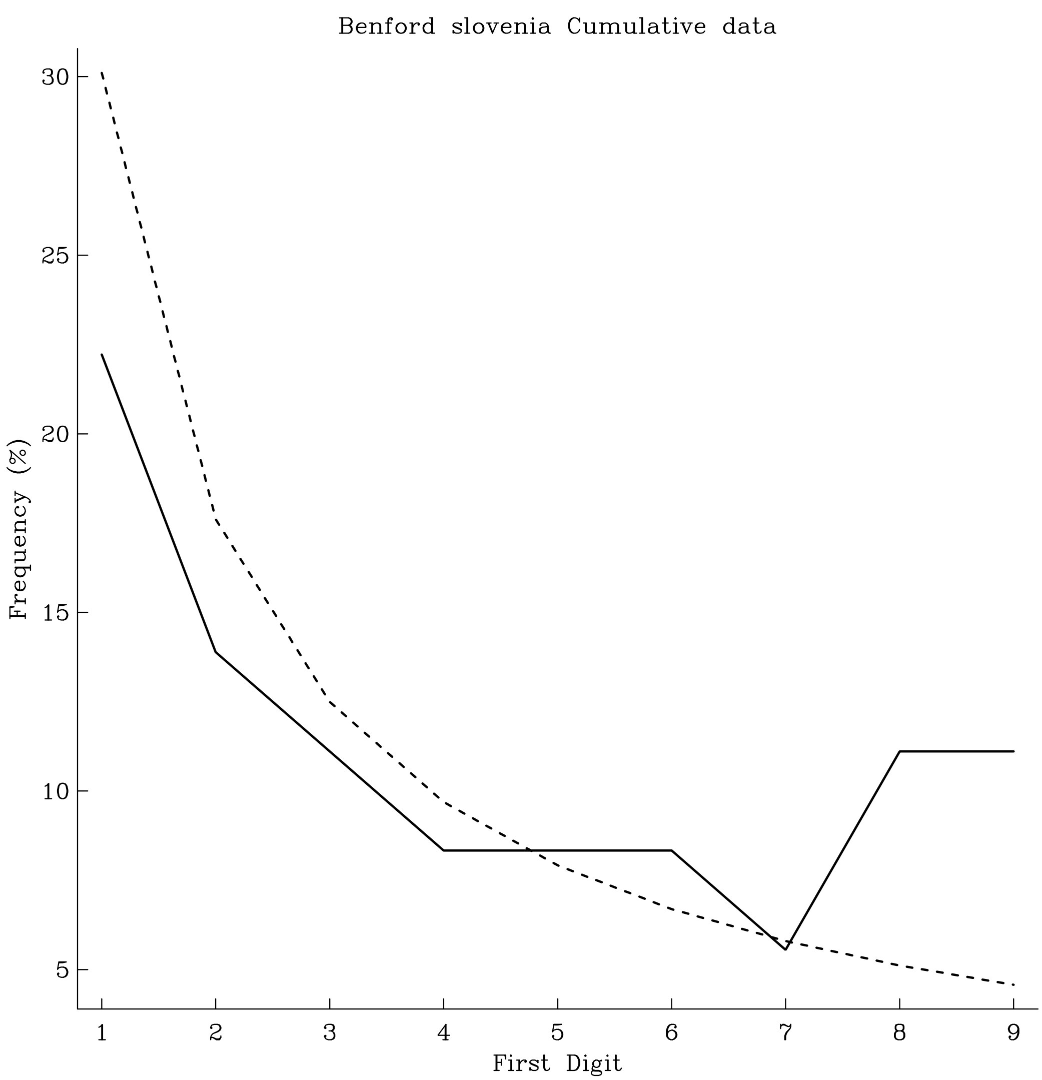
Benford russia Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit


Benford saudiarabia Cumulative data
40
35
30
25
20
15
10
5
Frequency (%)
1
2
3
4
5
6
7
8
9
First Digit

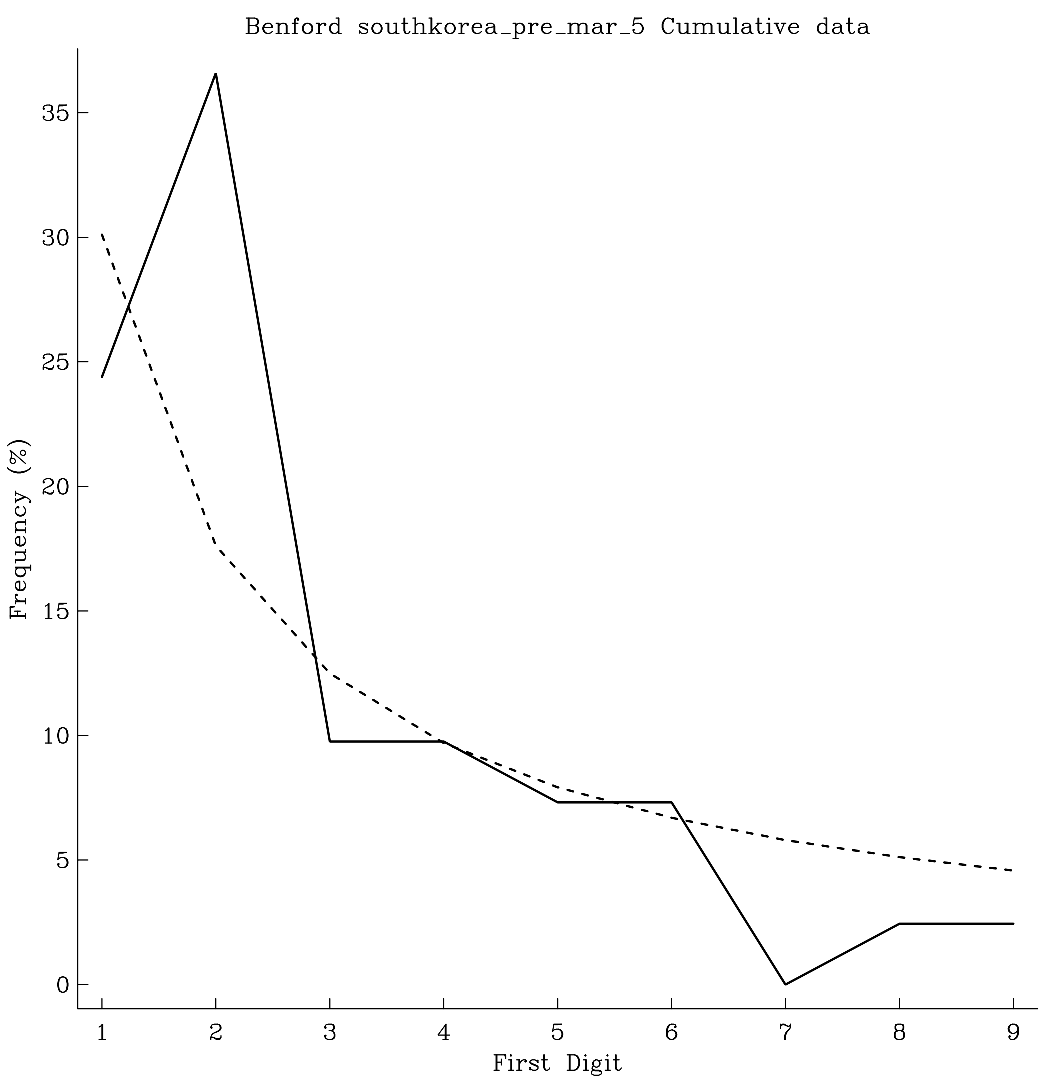

Benford singapore Cumulative data
Frequency (%)
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

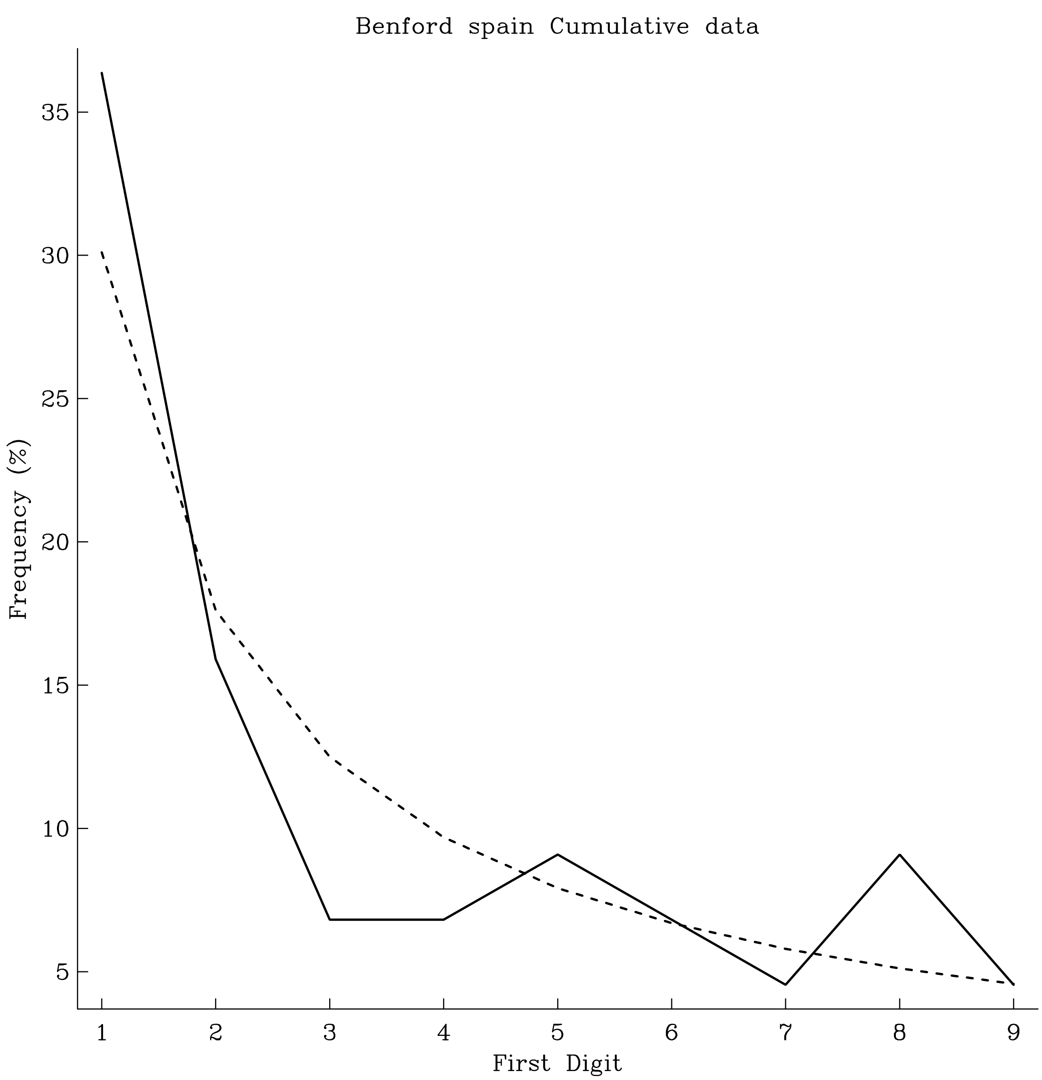


Benford slovakia Cumulative data
Frequency (%)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First Digit

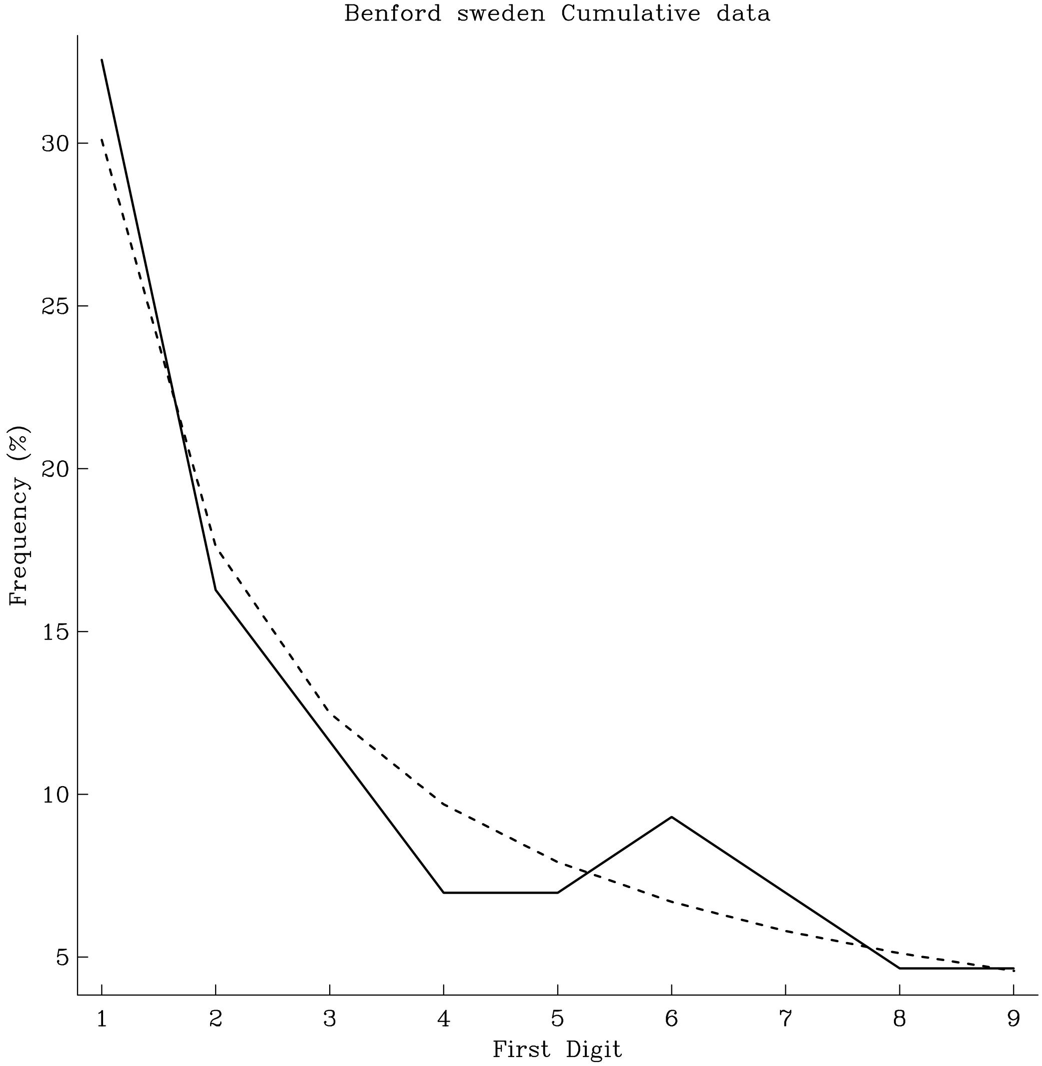


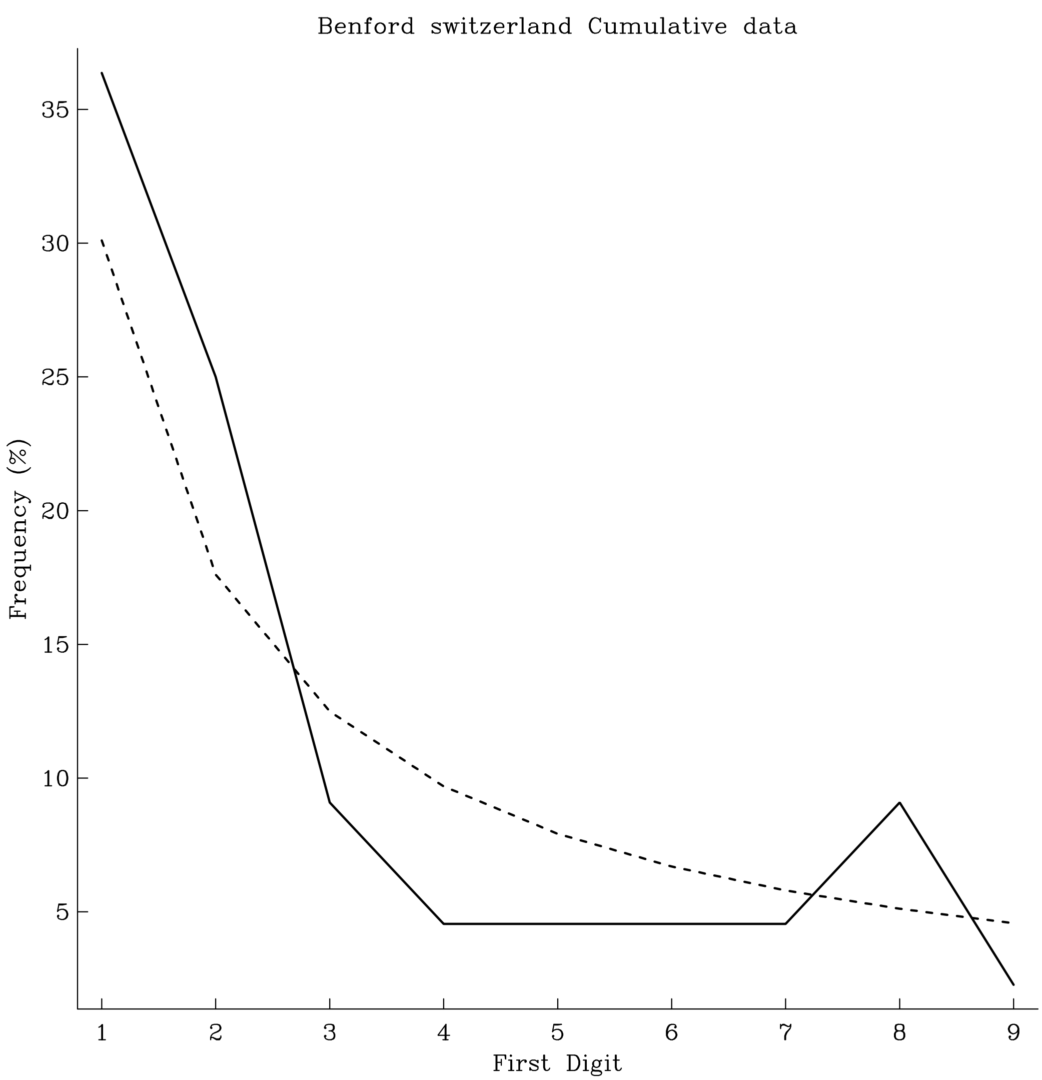
Benford slovenia Cumulative data
Frequency (%)
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

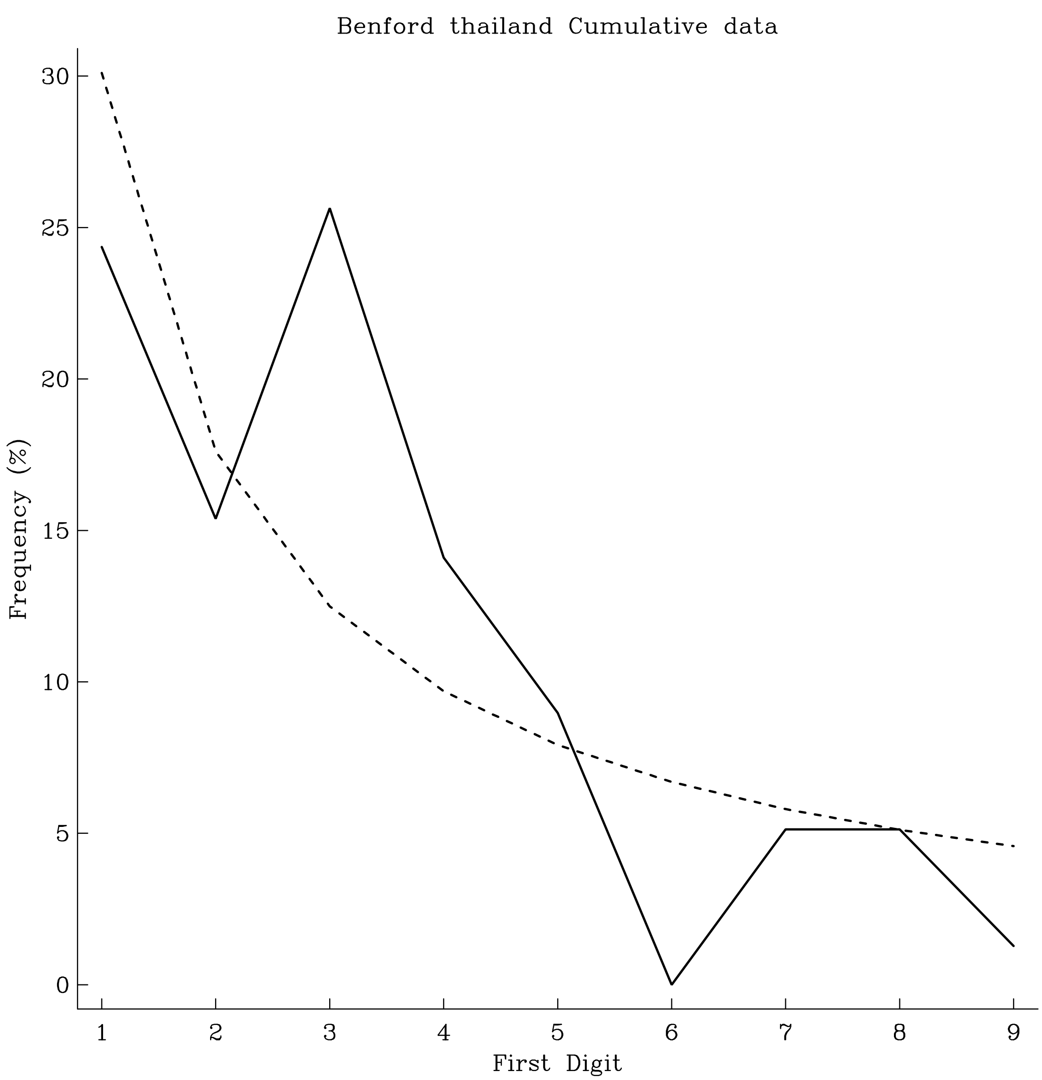

Benford southkorea_pre_mar_5 Cumulative data
Frequency (%)
First Digit
0
5
10
15
20
25
30
35
1
2
3
4
5
6
7
8
9

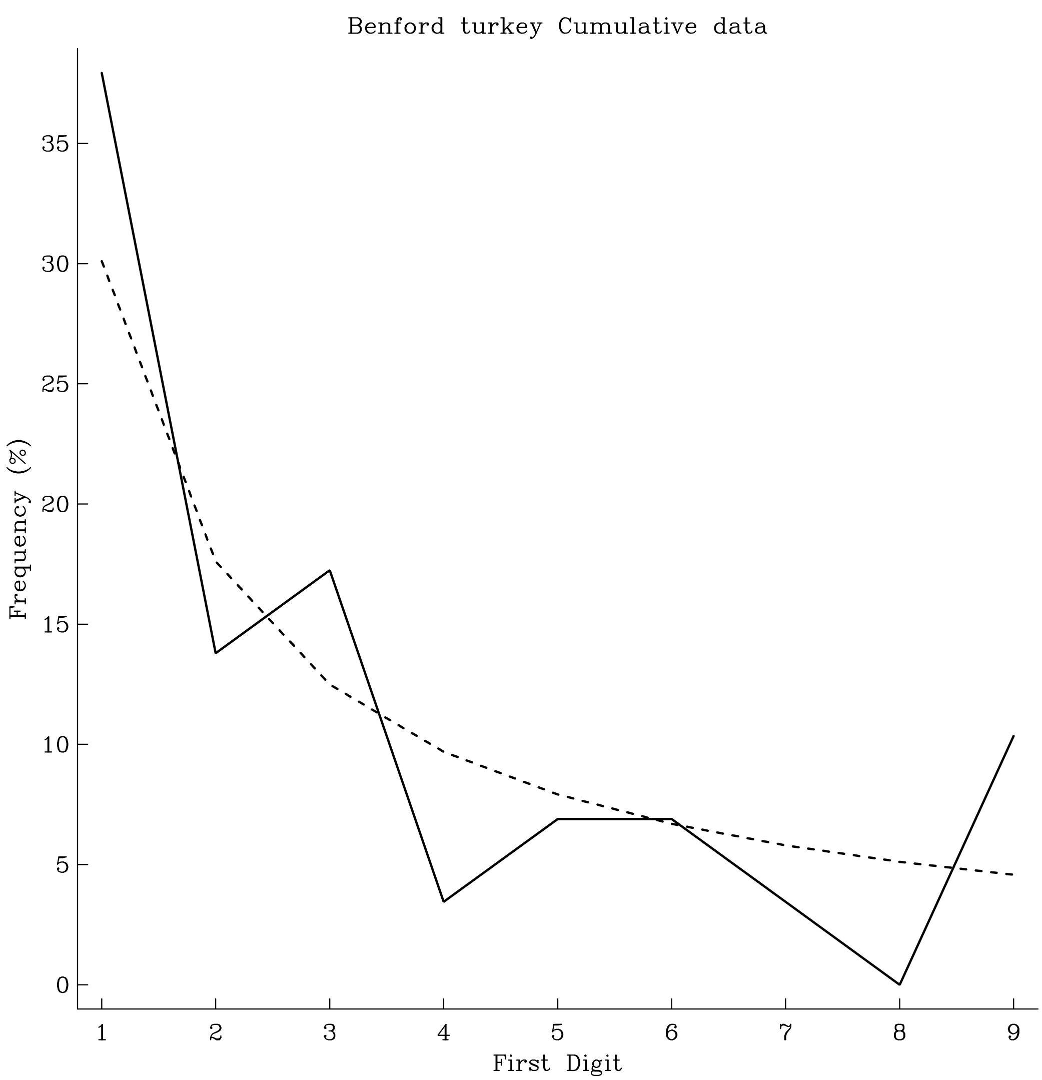

Benford spain Cumulative data
Frequency (%)
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

Benford sweden Cumulative data

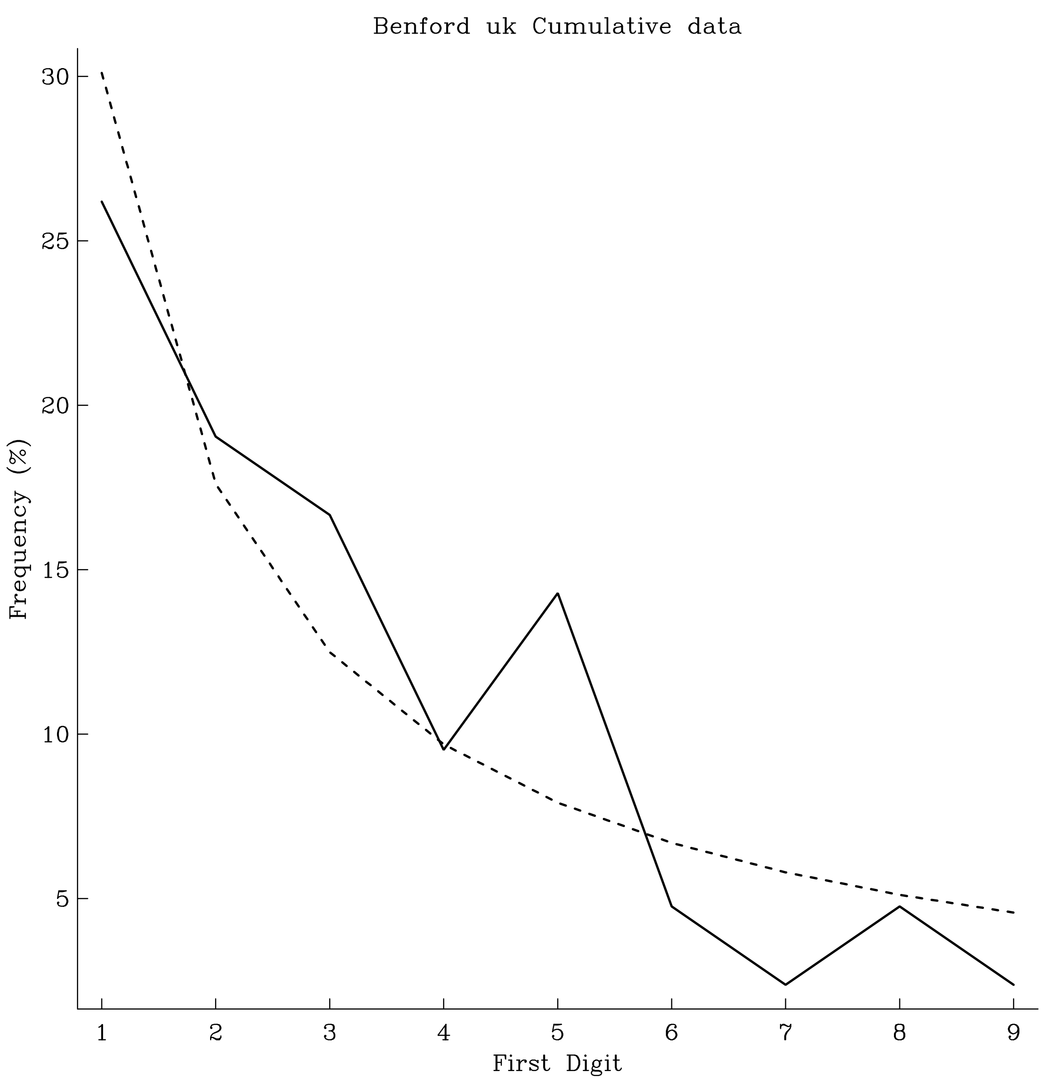



Benford switzerland Cumulative data
Frequency (%)
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
First Digit

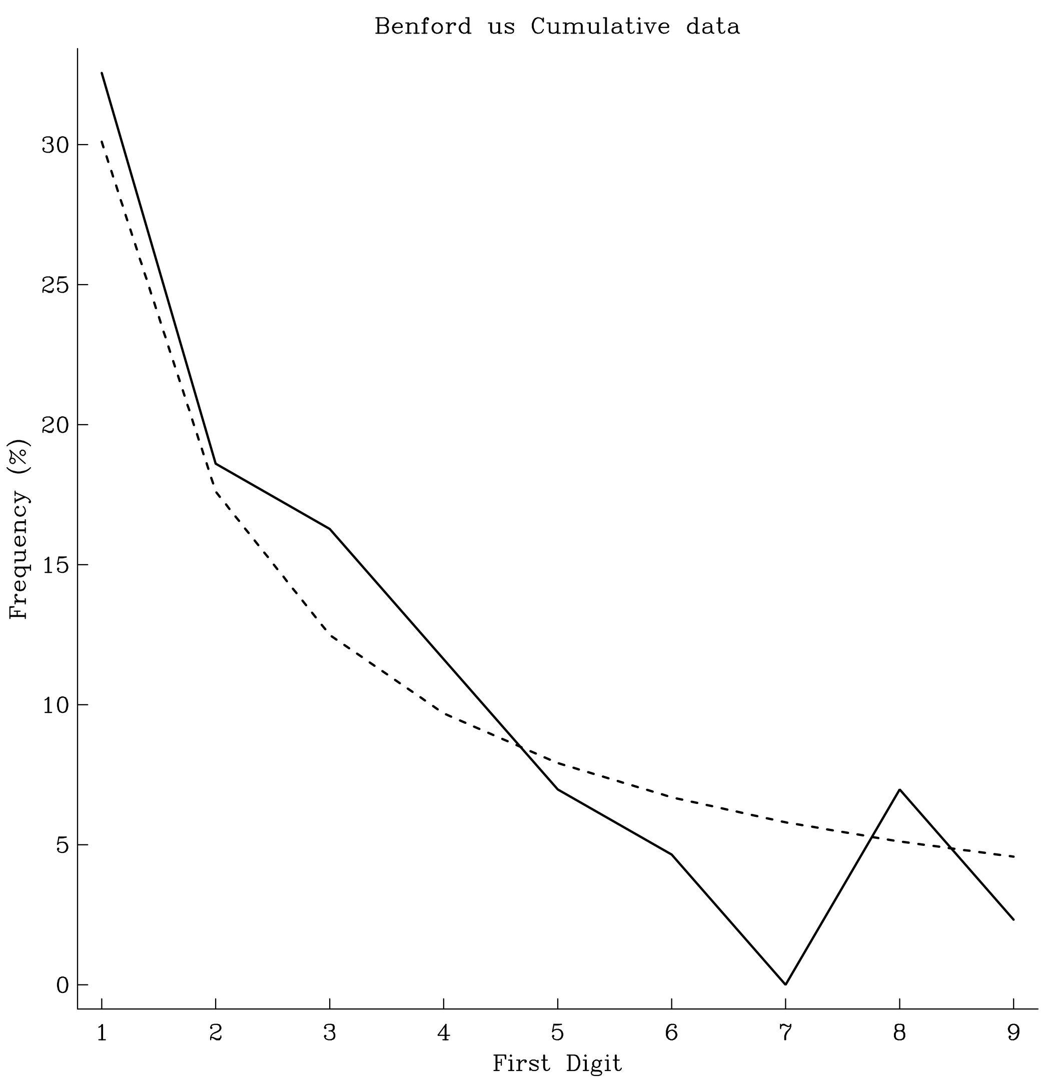

Benford thailand Cumulative data
Frequency (%)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
First Digit



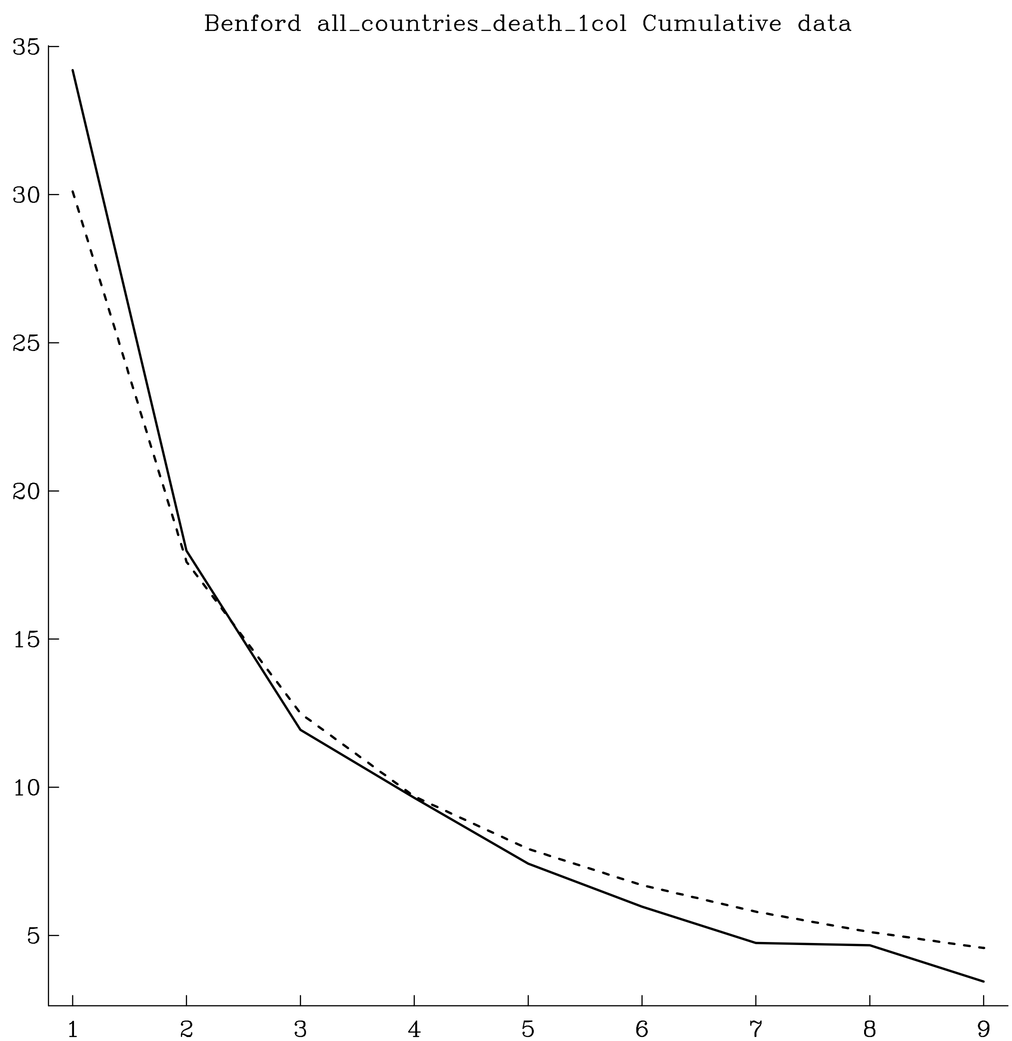
Benford turkey Cumulative data
Frequency (%)
0
5
10
15
20
25
30
35
1
2
3
4
5
6
7
8
9
First Digit

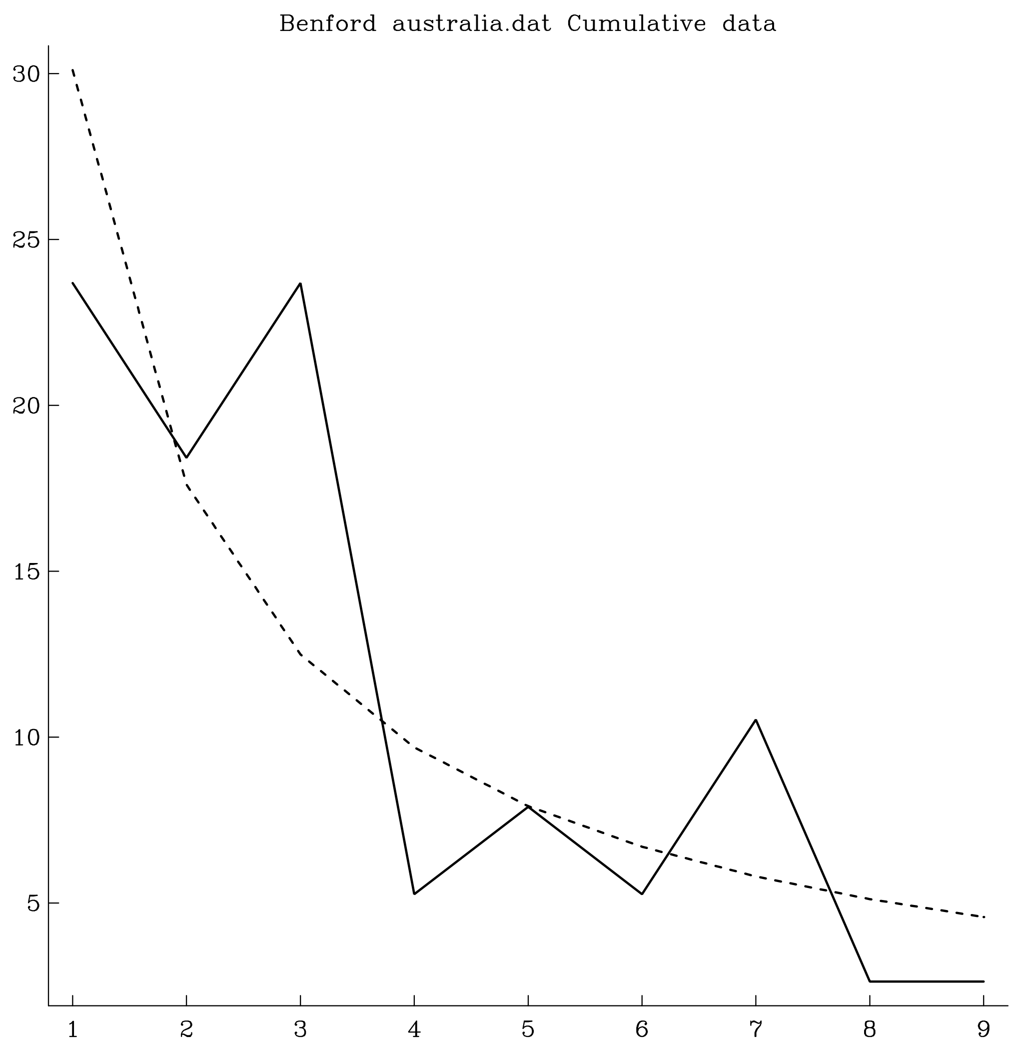


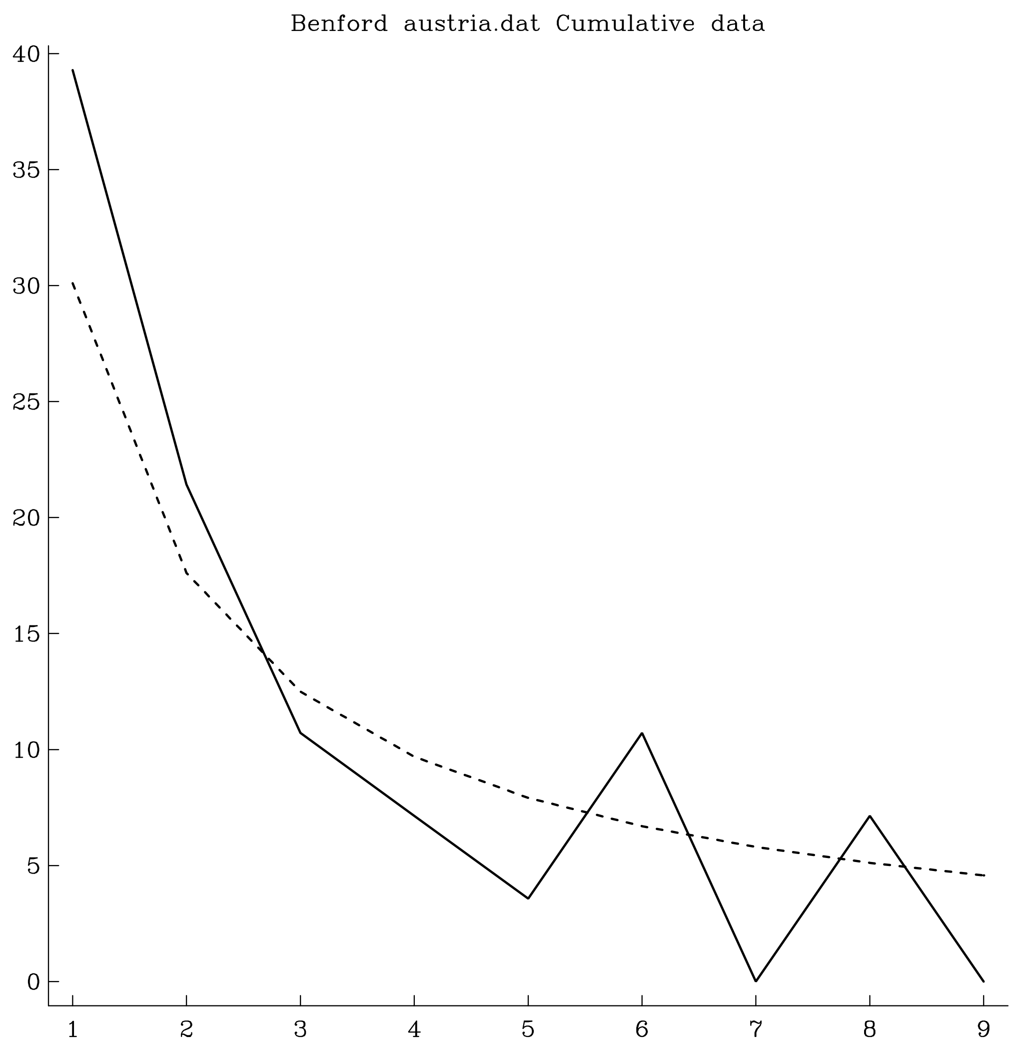
Benford uk Cumulative data
30
25
20
15
10
5
Frequency (%)
1
2
3
4
5
6
7
8
9
First Digit

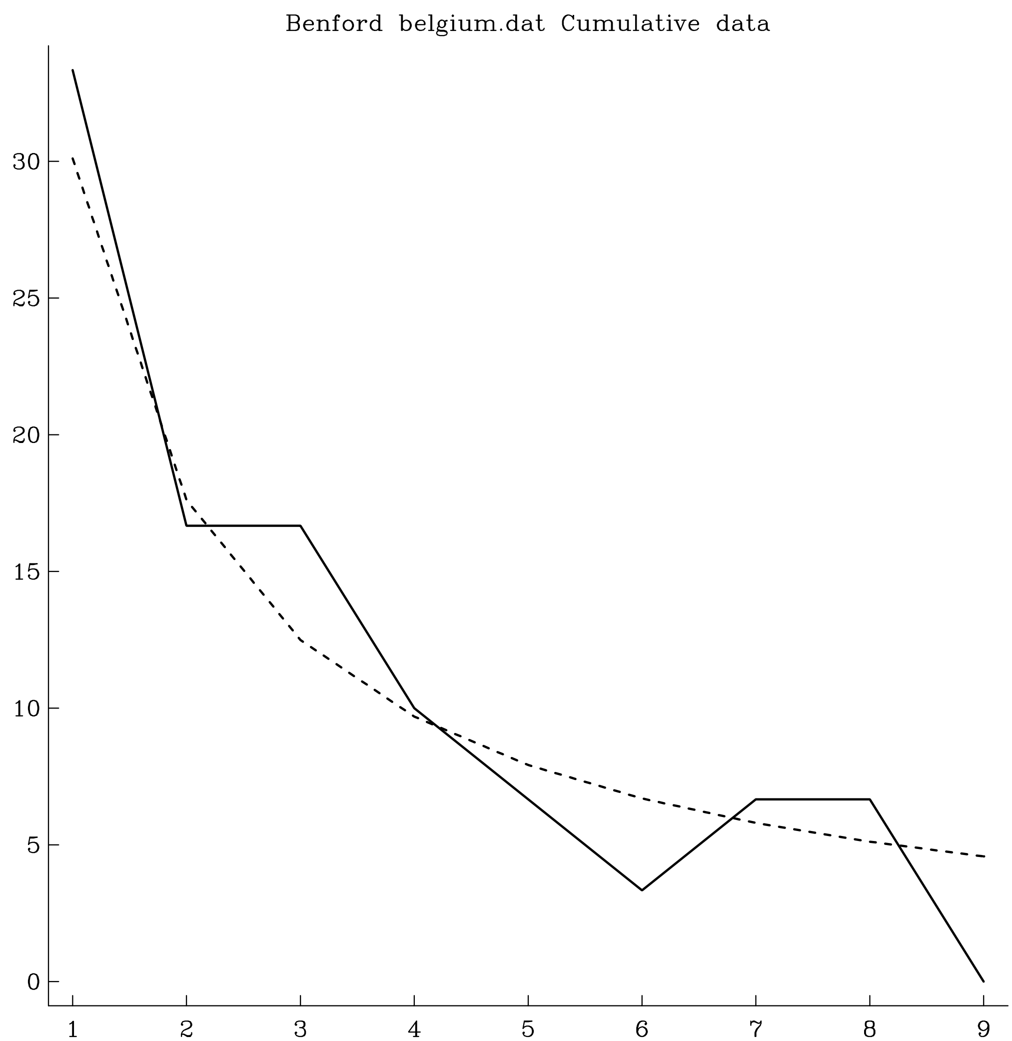

Benford us Cumulative data
Frequency (%)
First Digit
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9

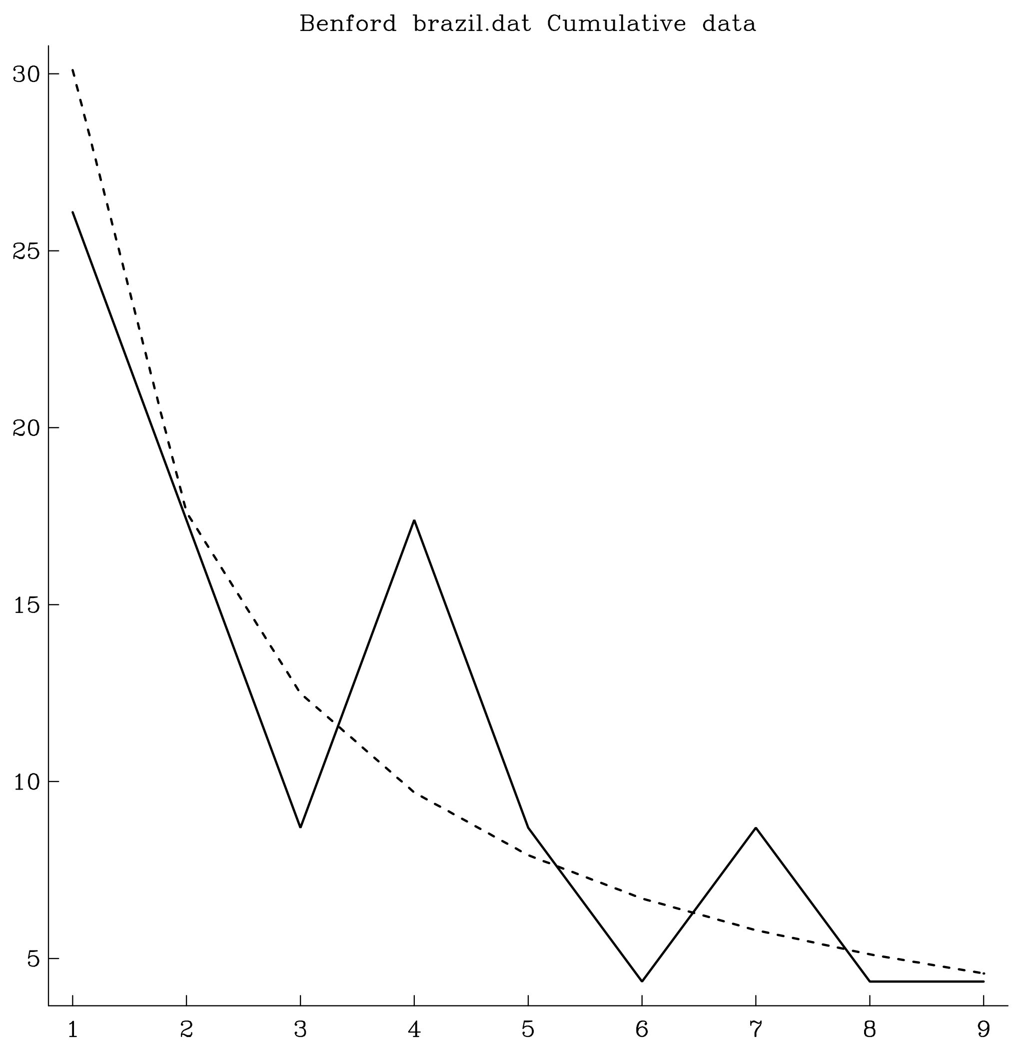
Deaths by country


Benford all_countries_death_1col Cumulative data
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

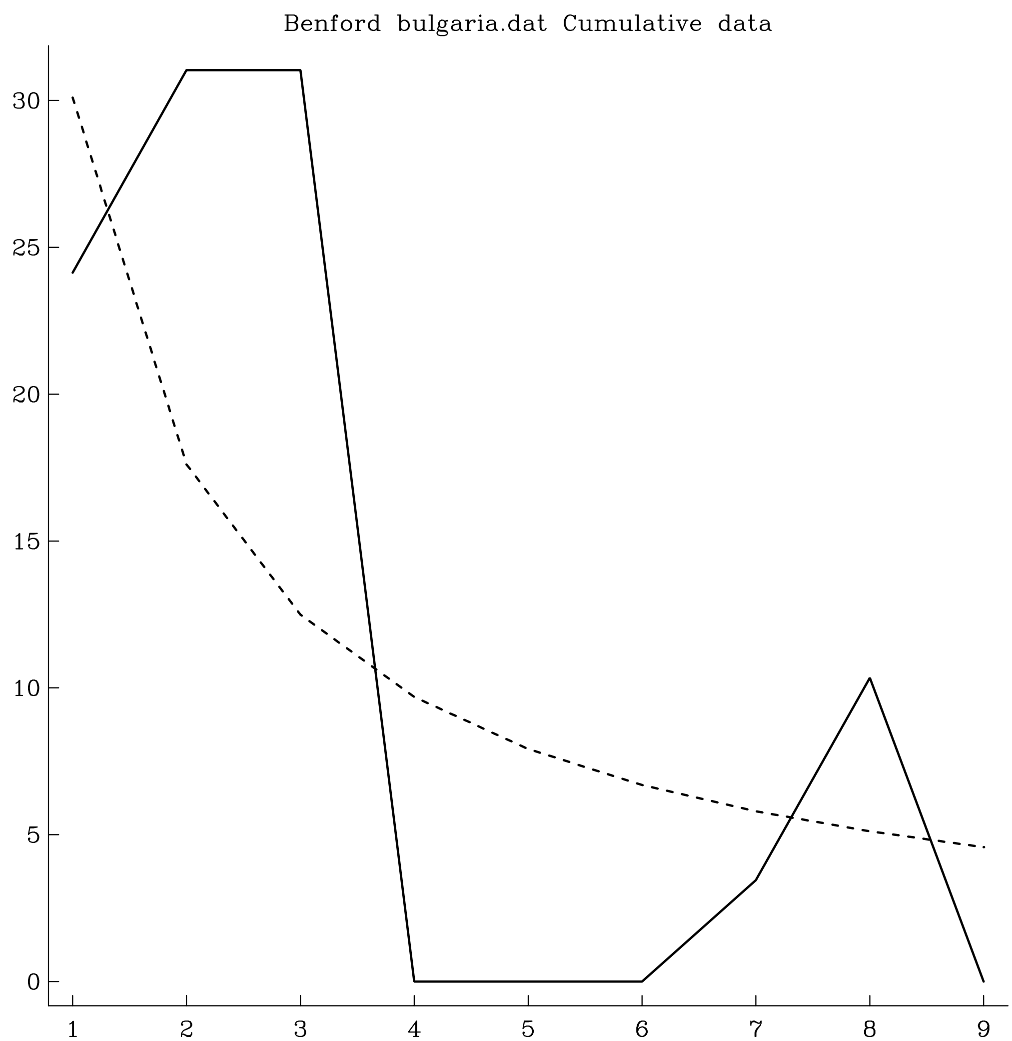

Benford australia.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9


Benford austria.dat Cumulative data
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

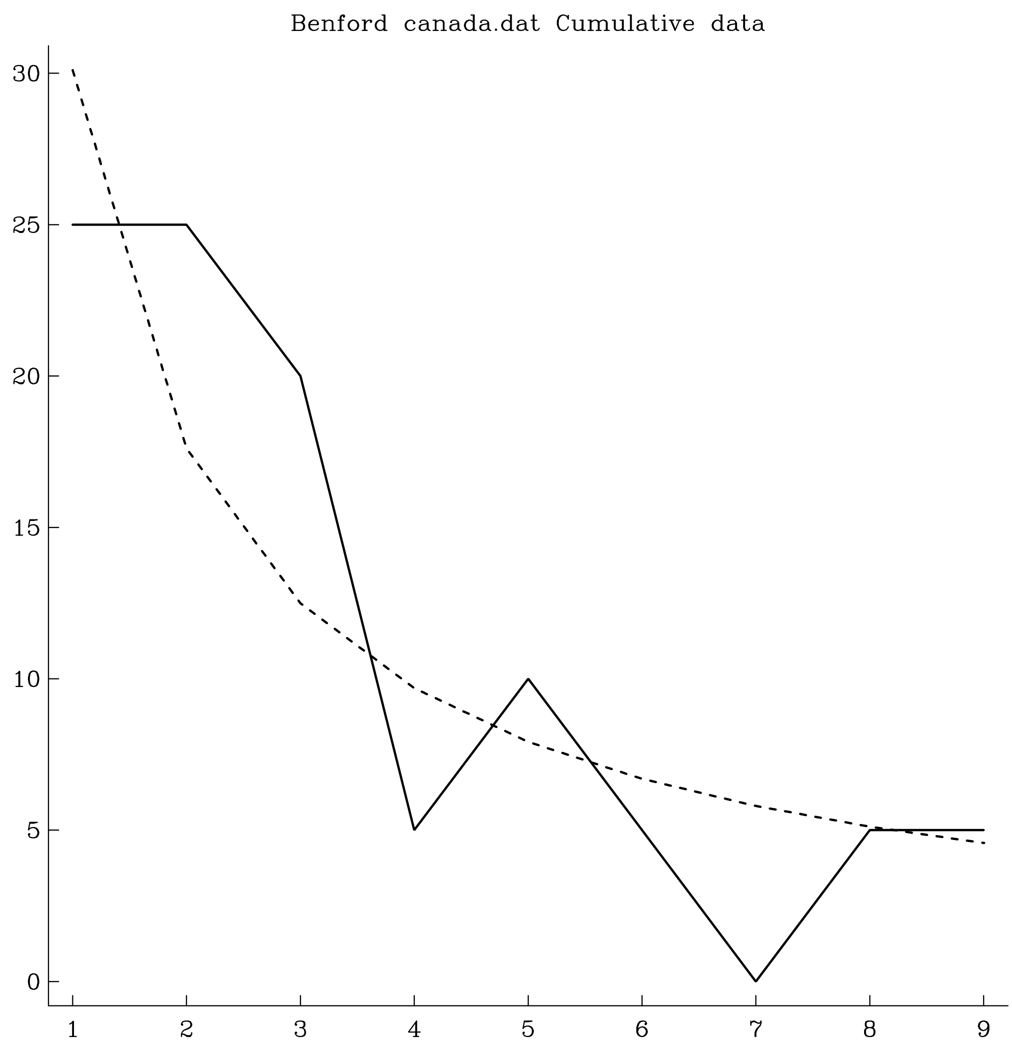

Benford belgium.dat Cumulative data
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9

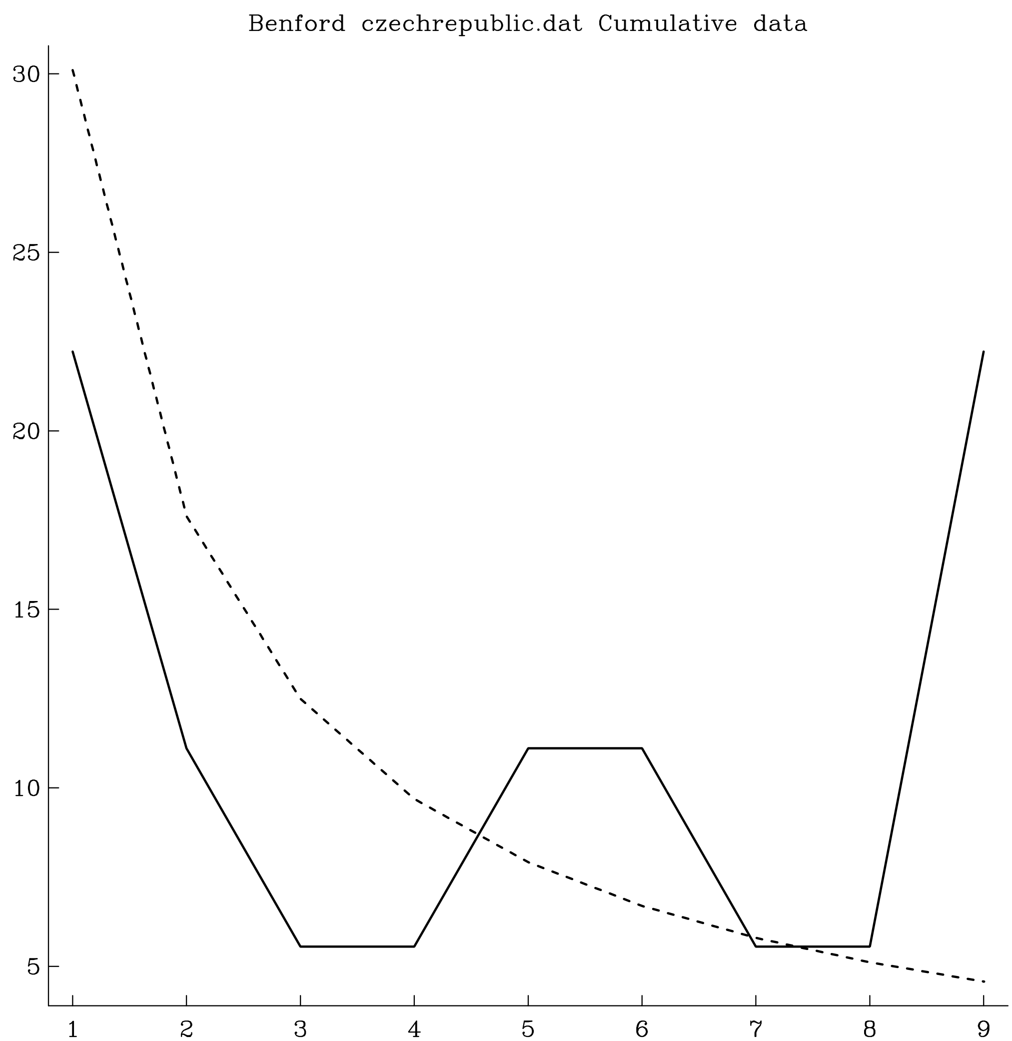


Benford brazil.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

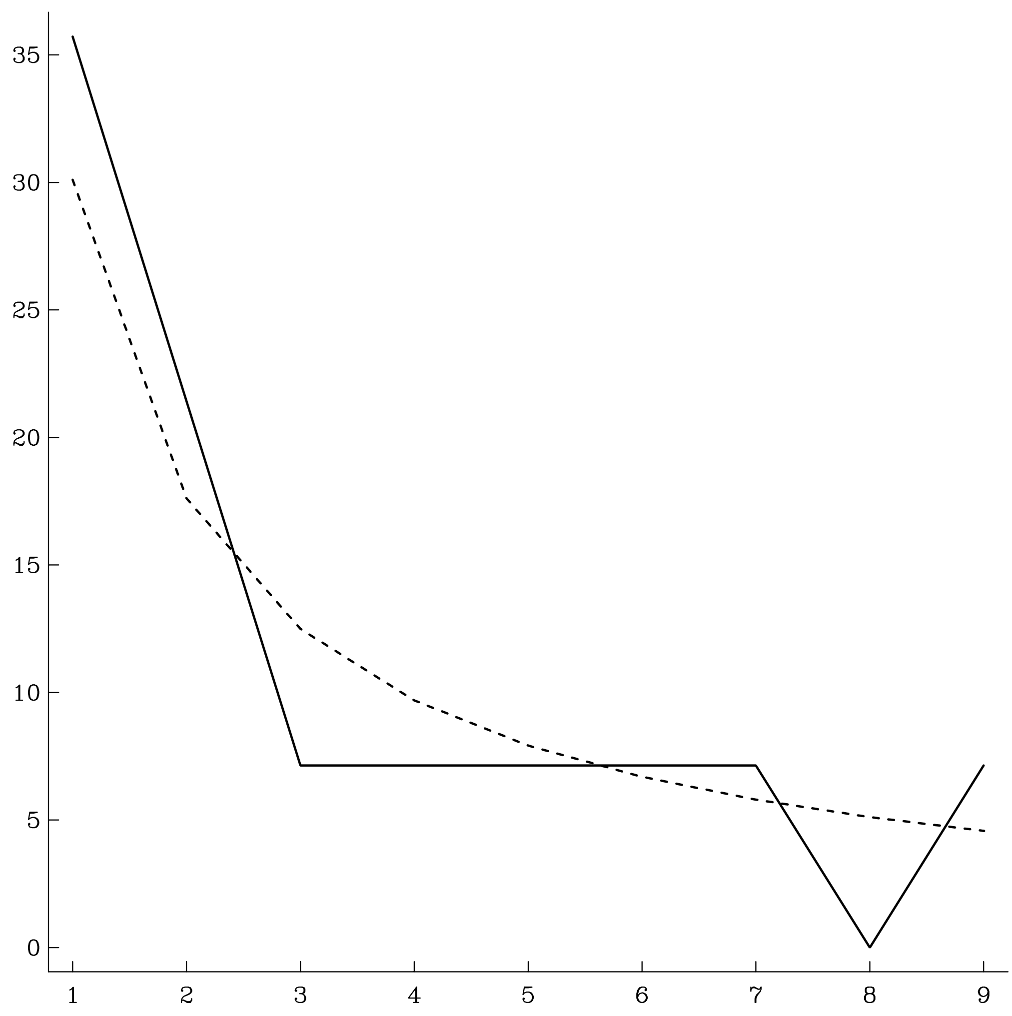


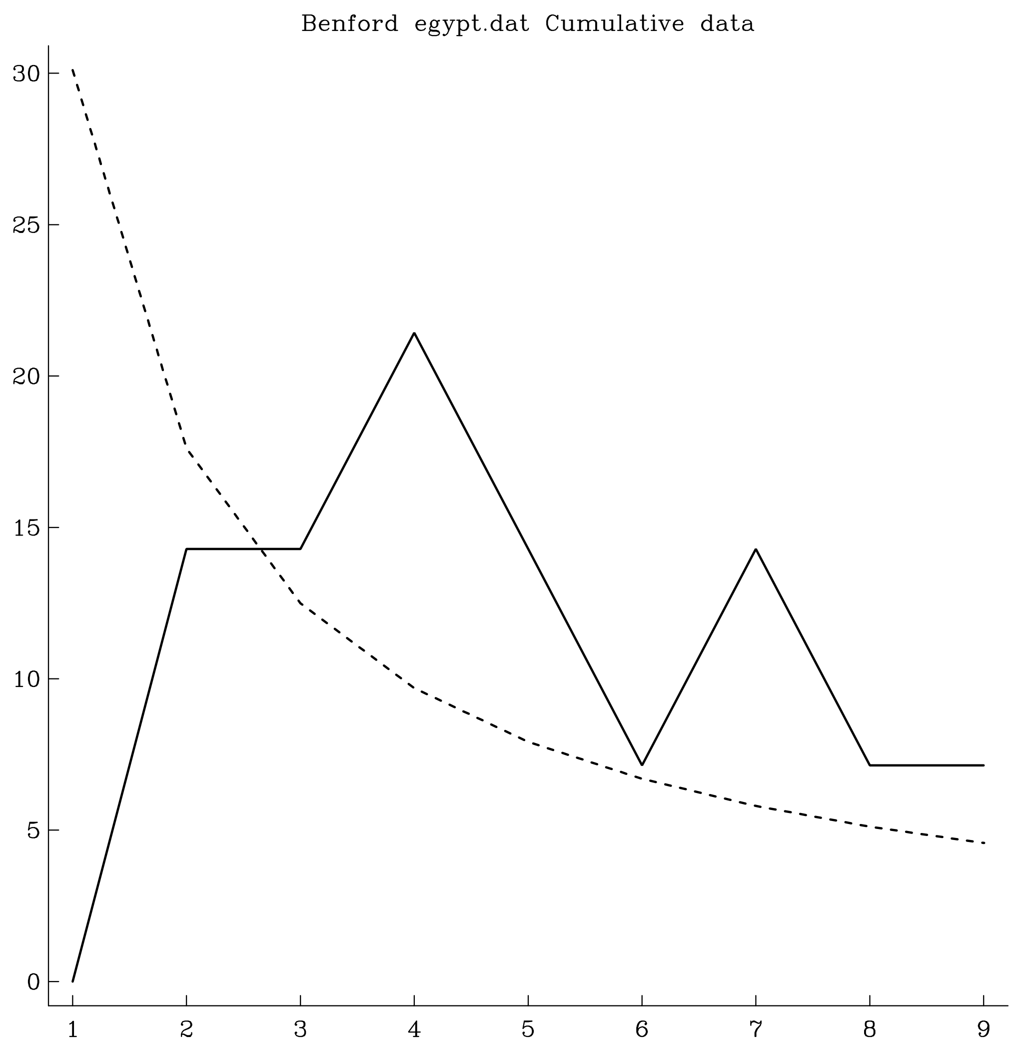
Benford bulgaria.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

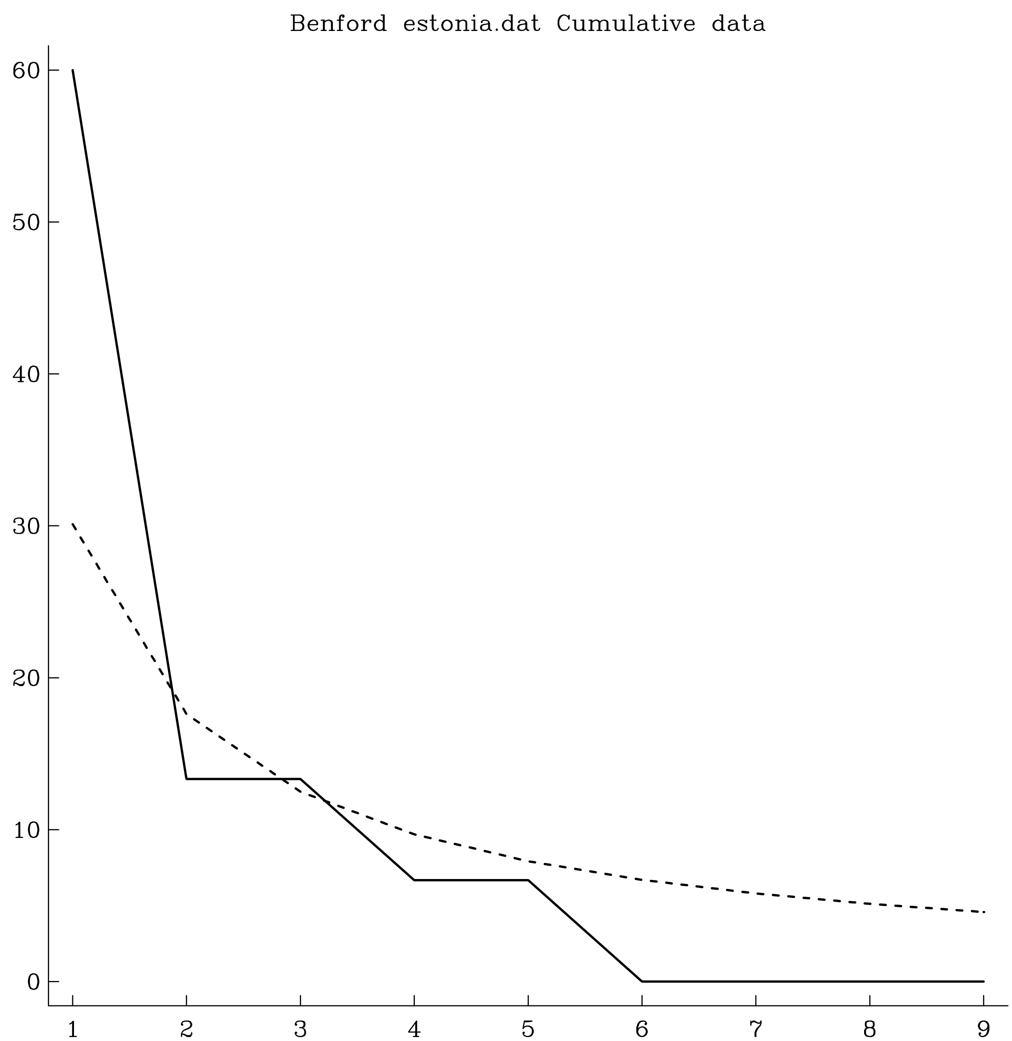

Benford canada.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9


Benford czechrepublic.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

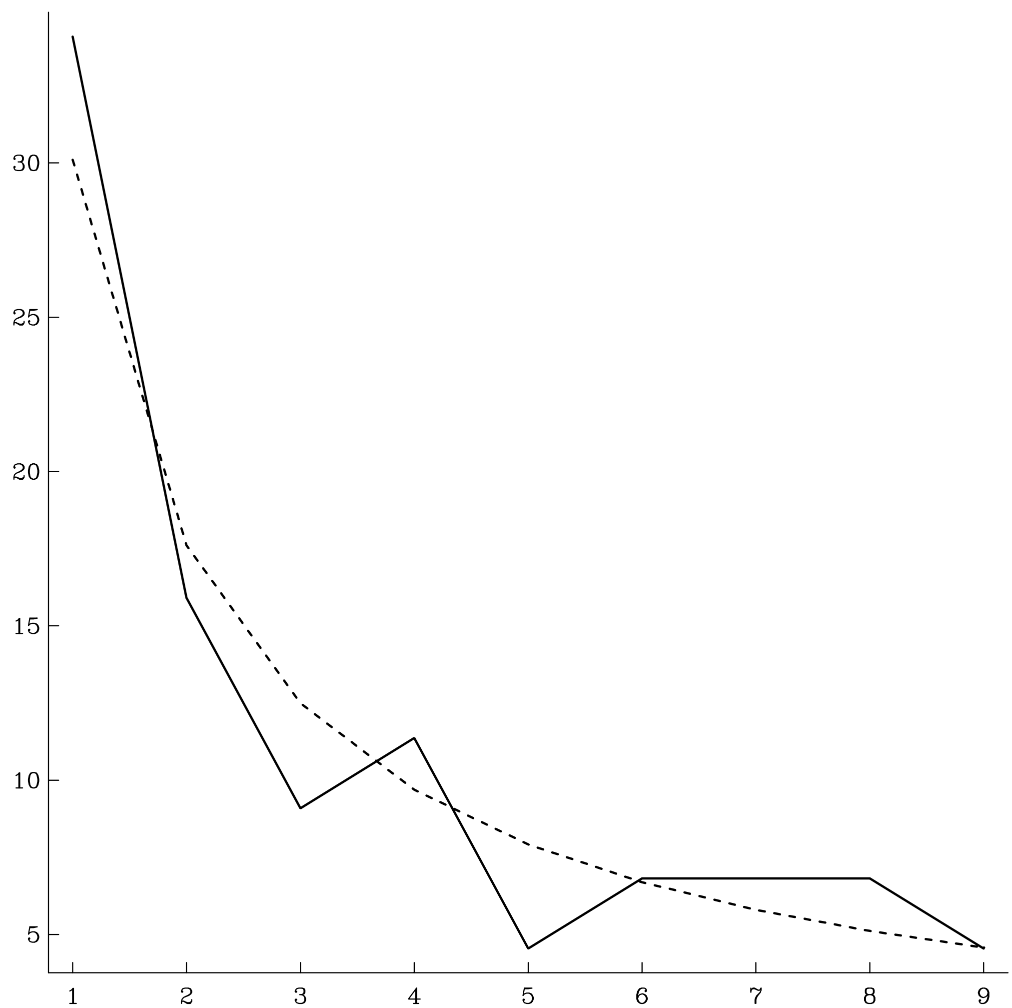
Benford denmark.dat Cumulative data



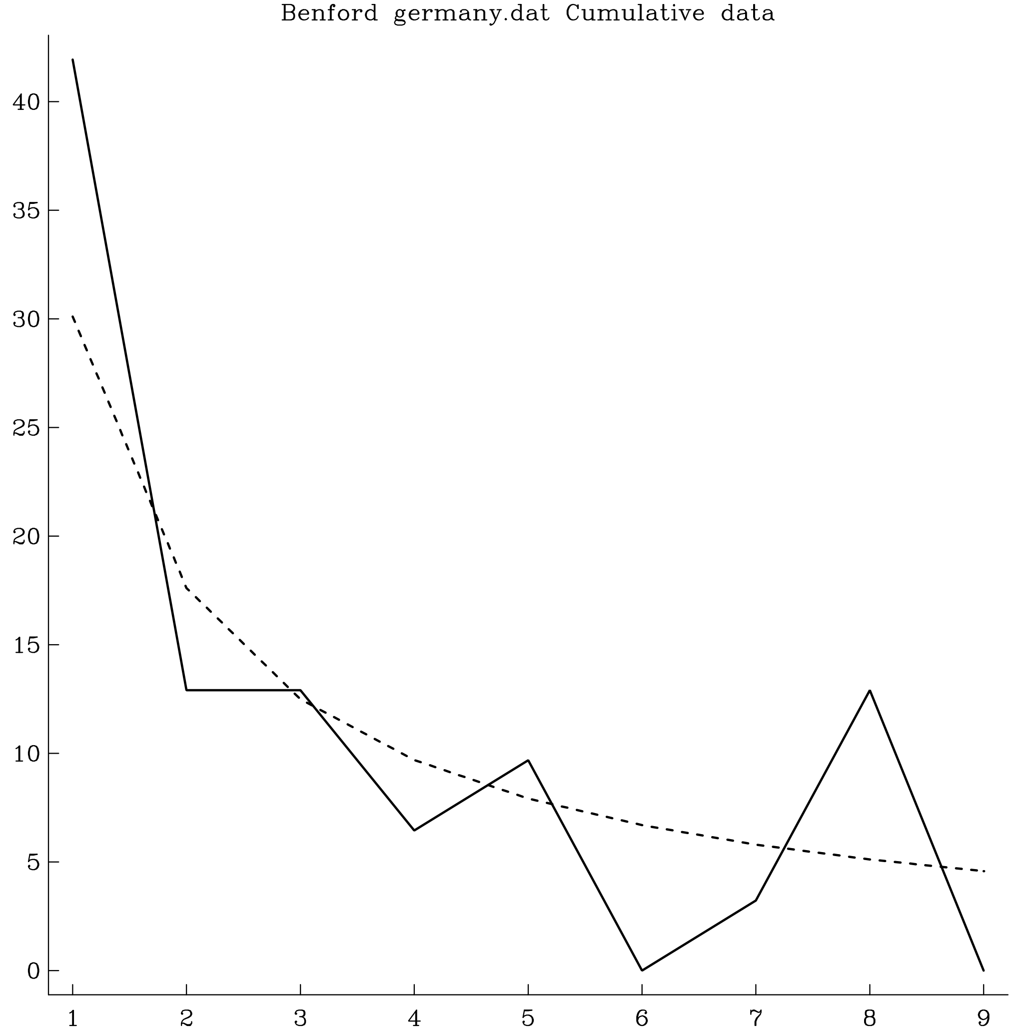

Benford egypt.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

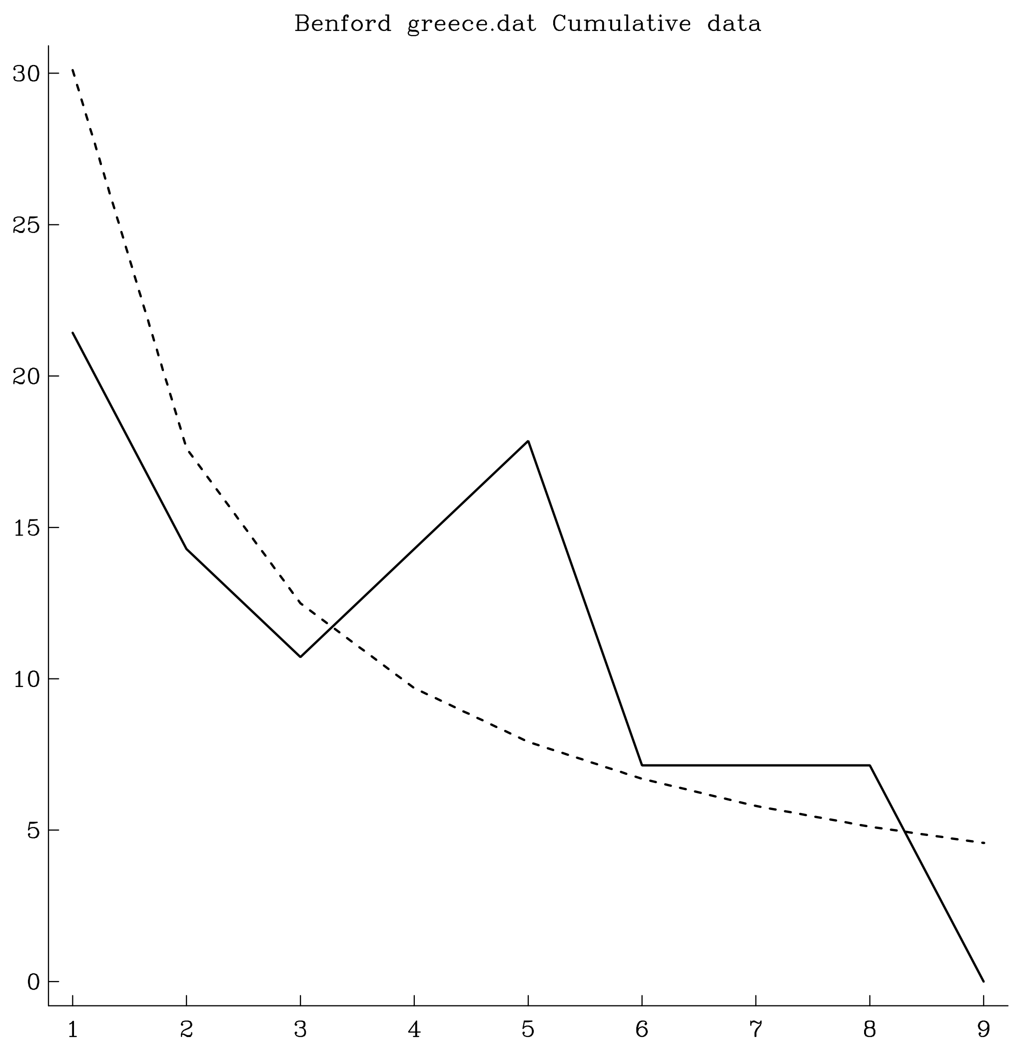

Benford estonia.dat Cumulative data
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9

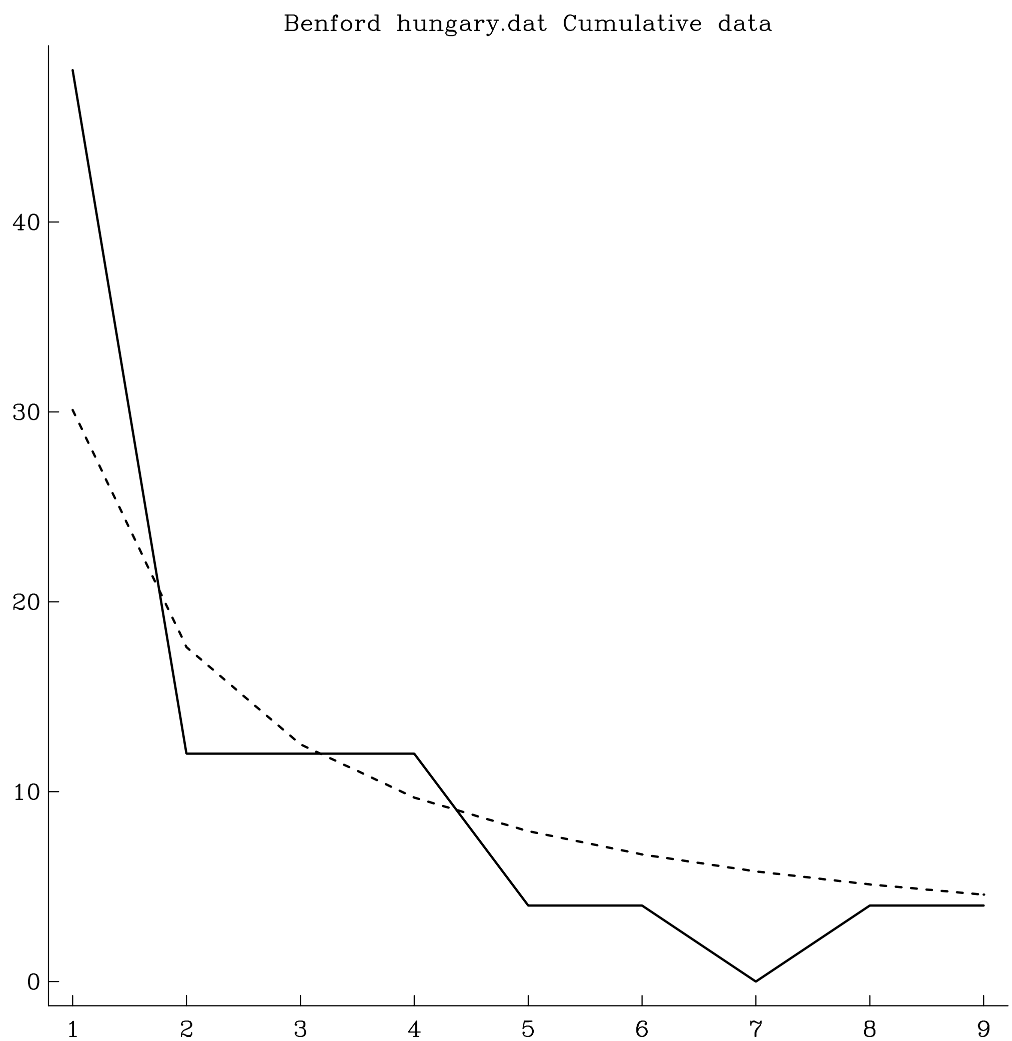
Benford france.dat Cumulative data

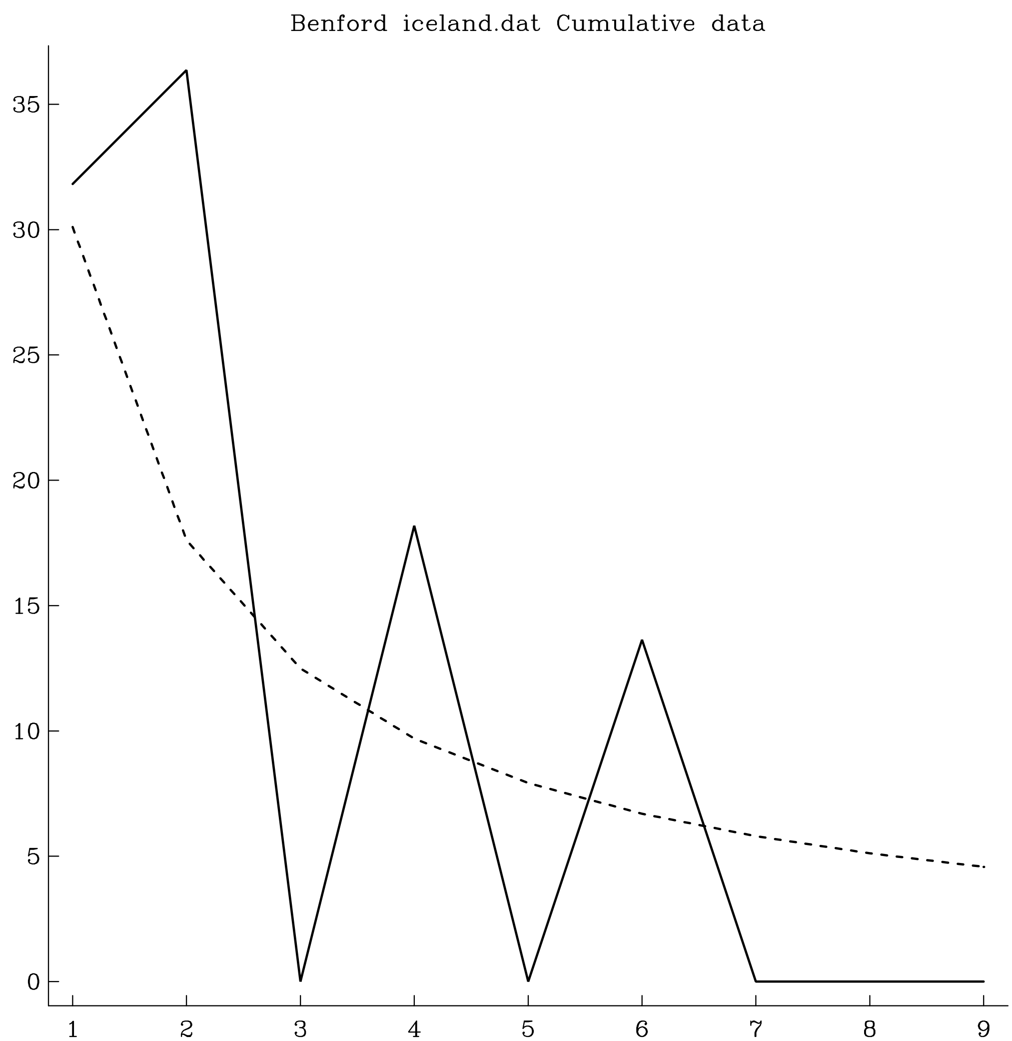


Benford germany.dat Cumulative data



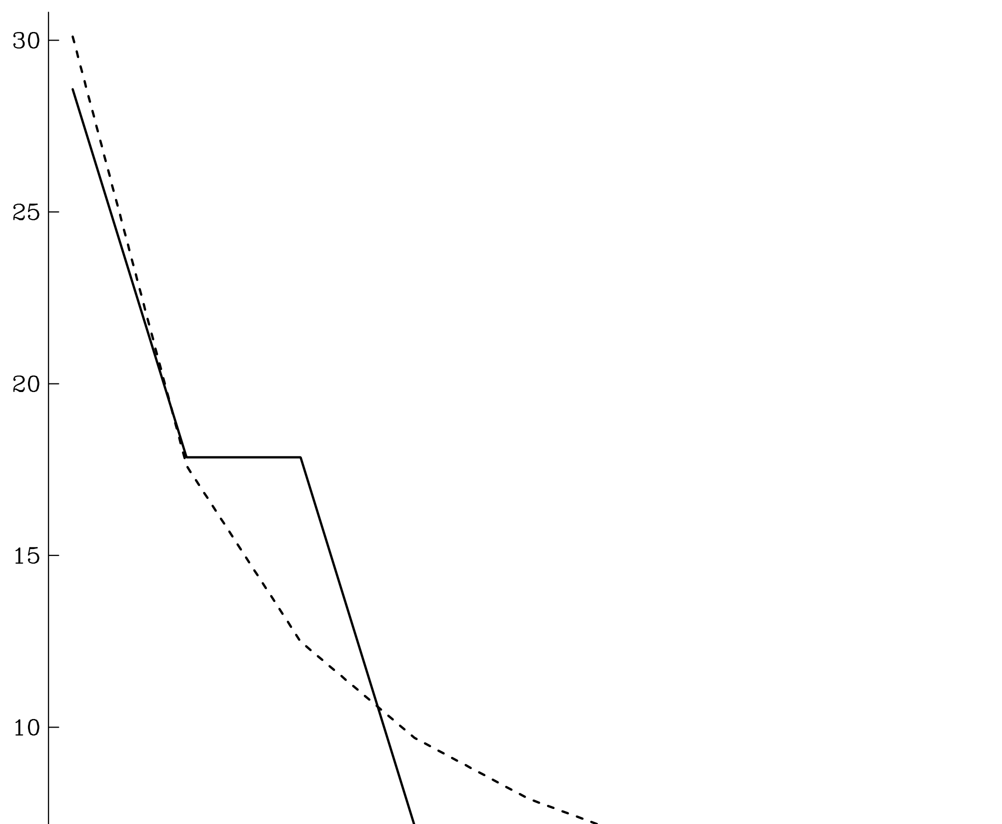

Benford greece.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

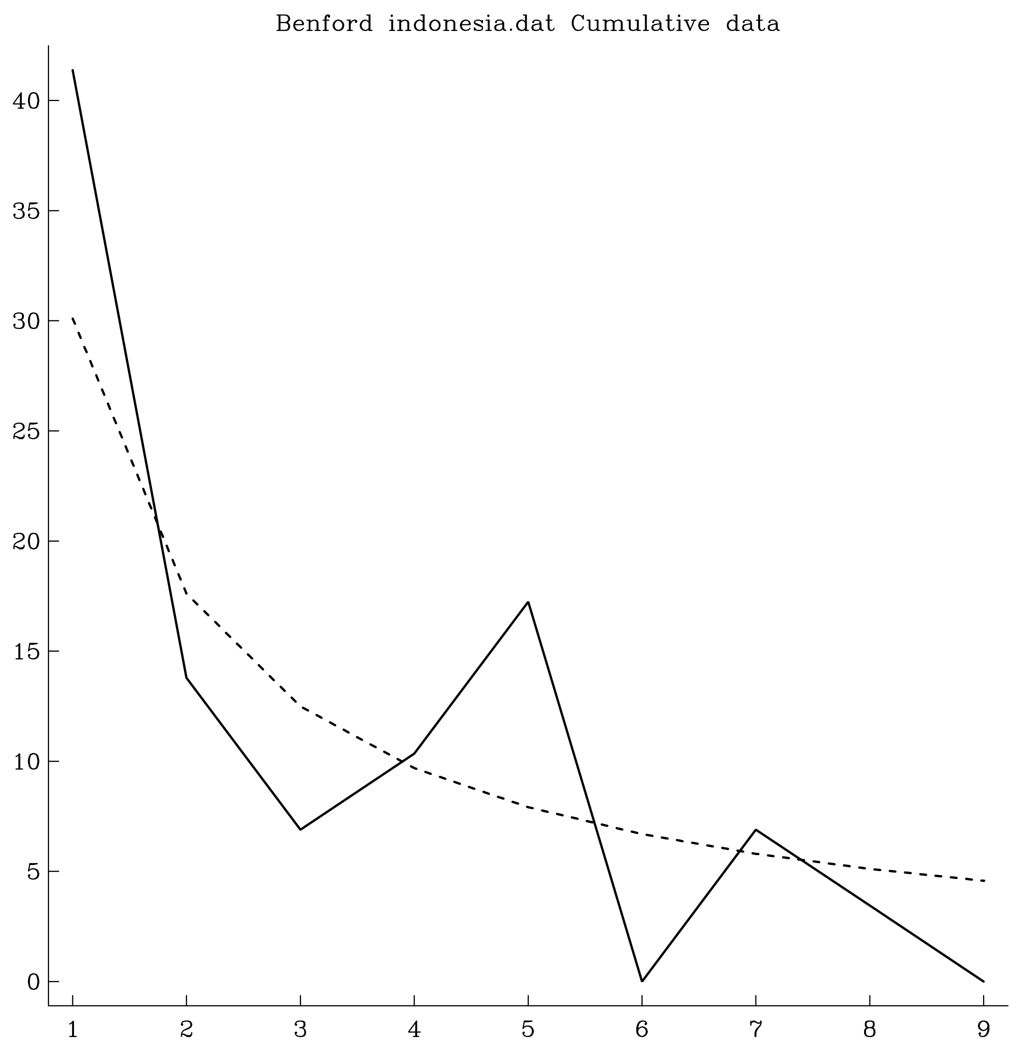

Benford hungary.dat Cumulative data
40
30
20
10
0
1
2
3
4
5
6
7
8
9

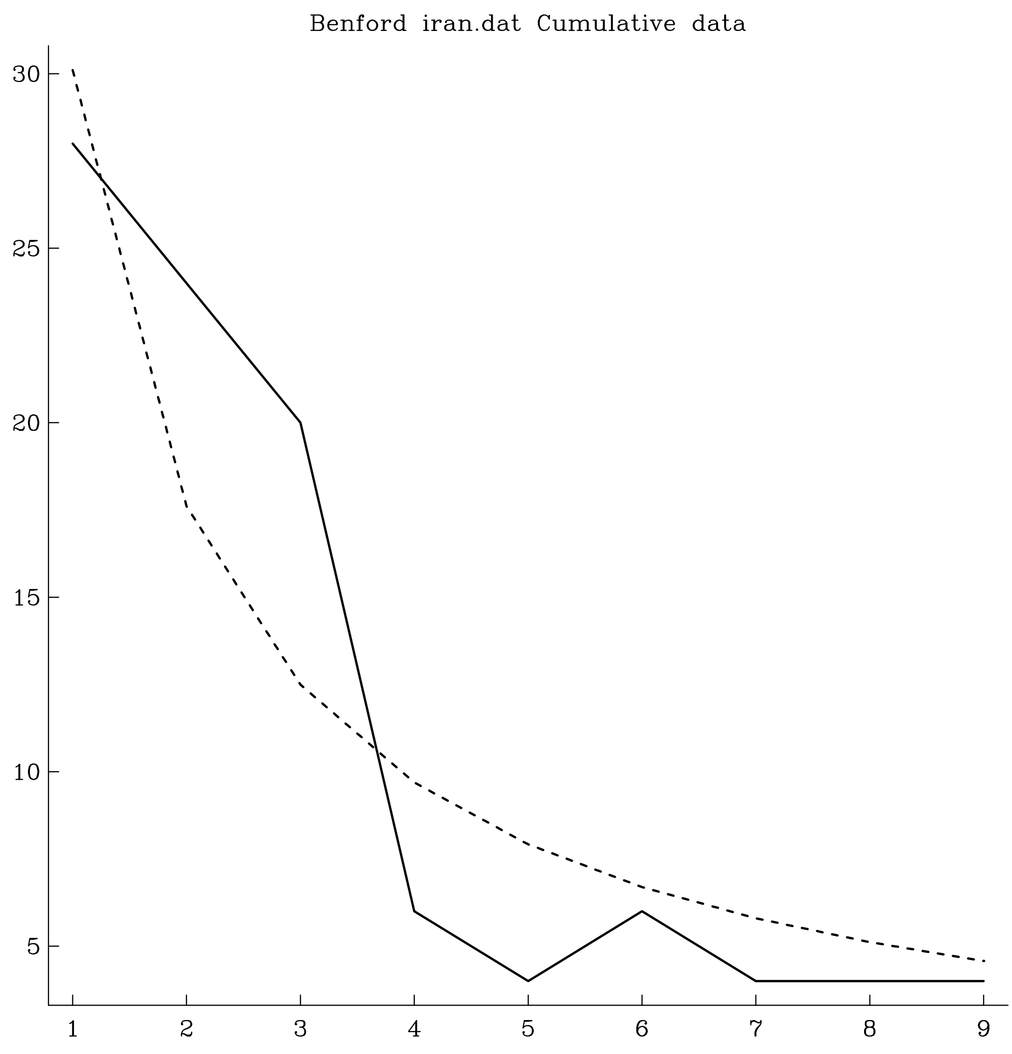


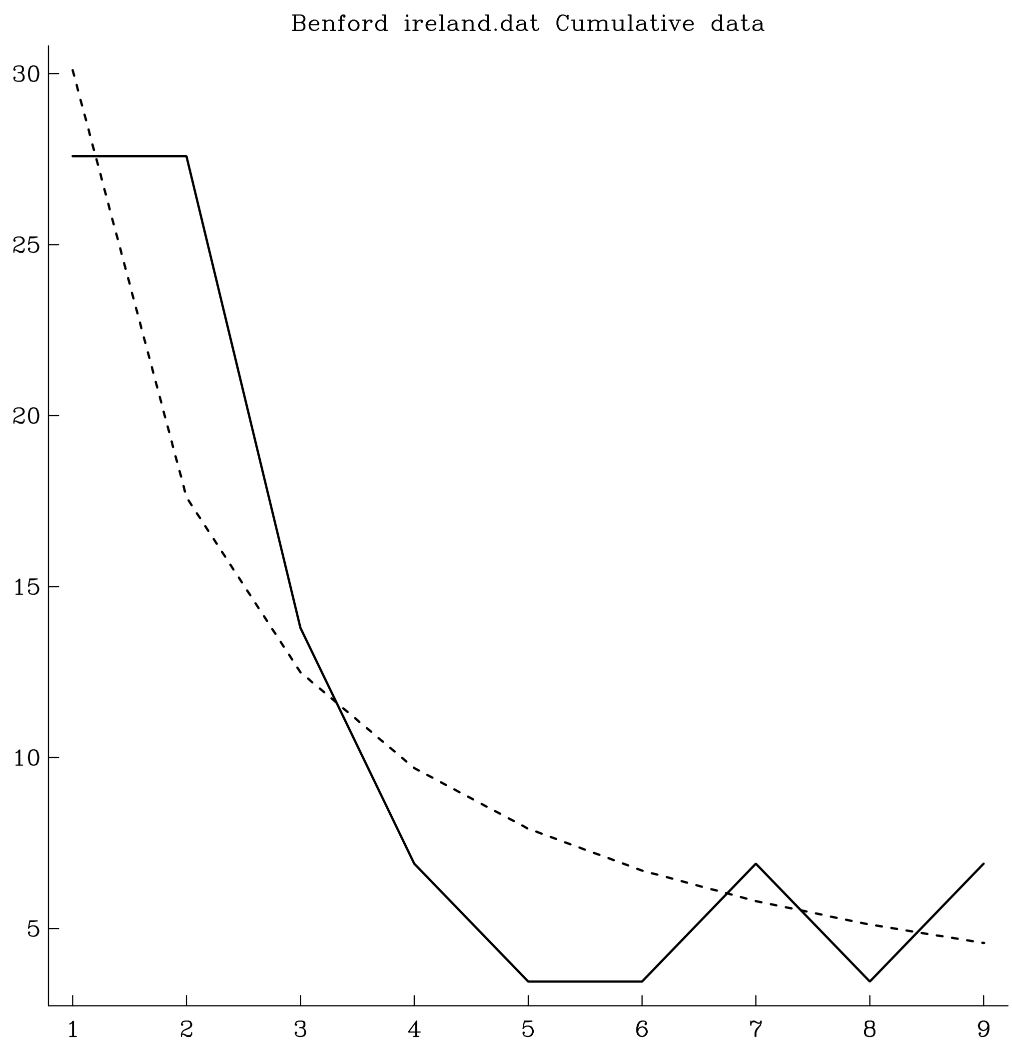
Benford iceland.dat Cumulative data
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

Benford india.dat Cumulative data

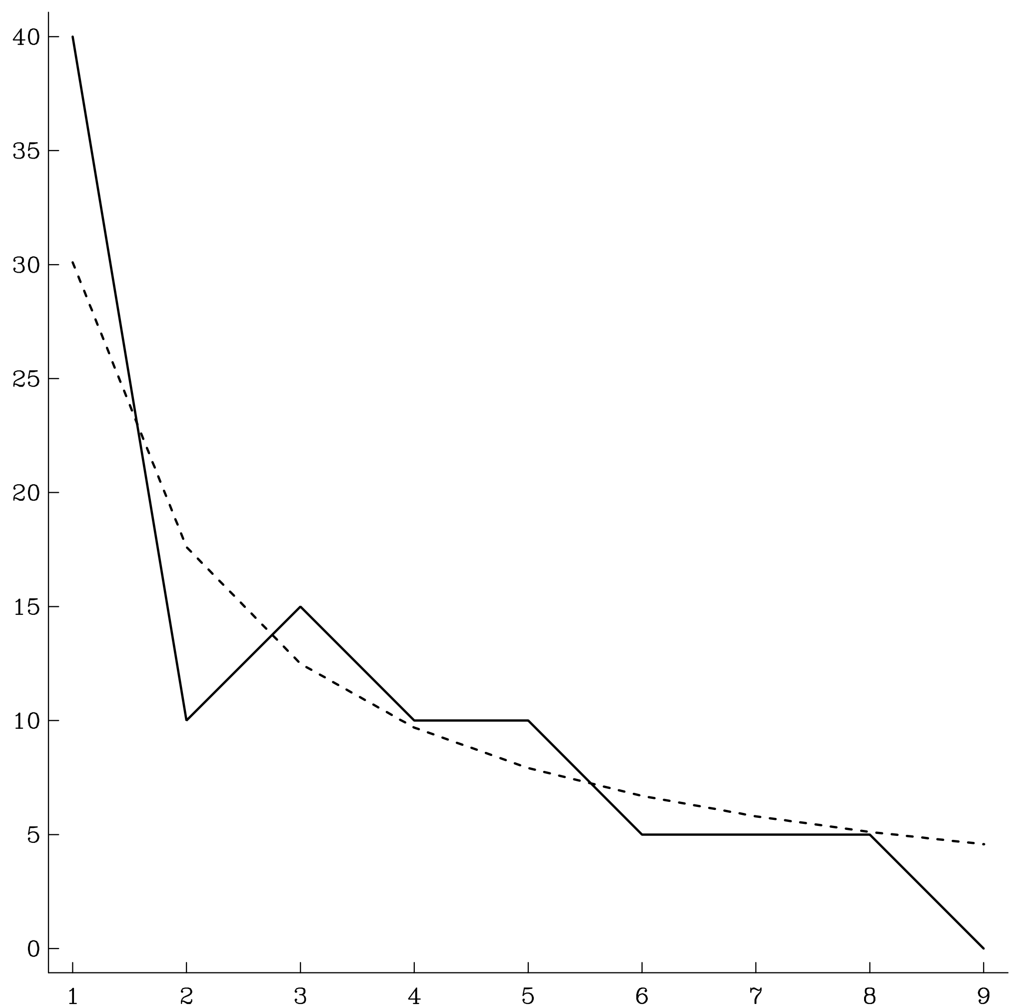



Benford indonesia.dat Cumulative data
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

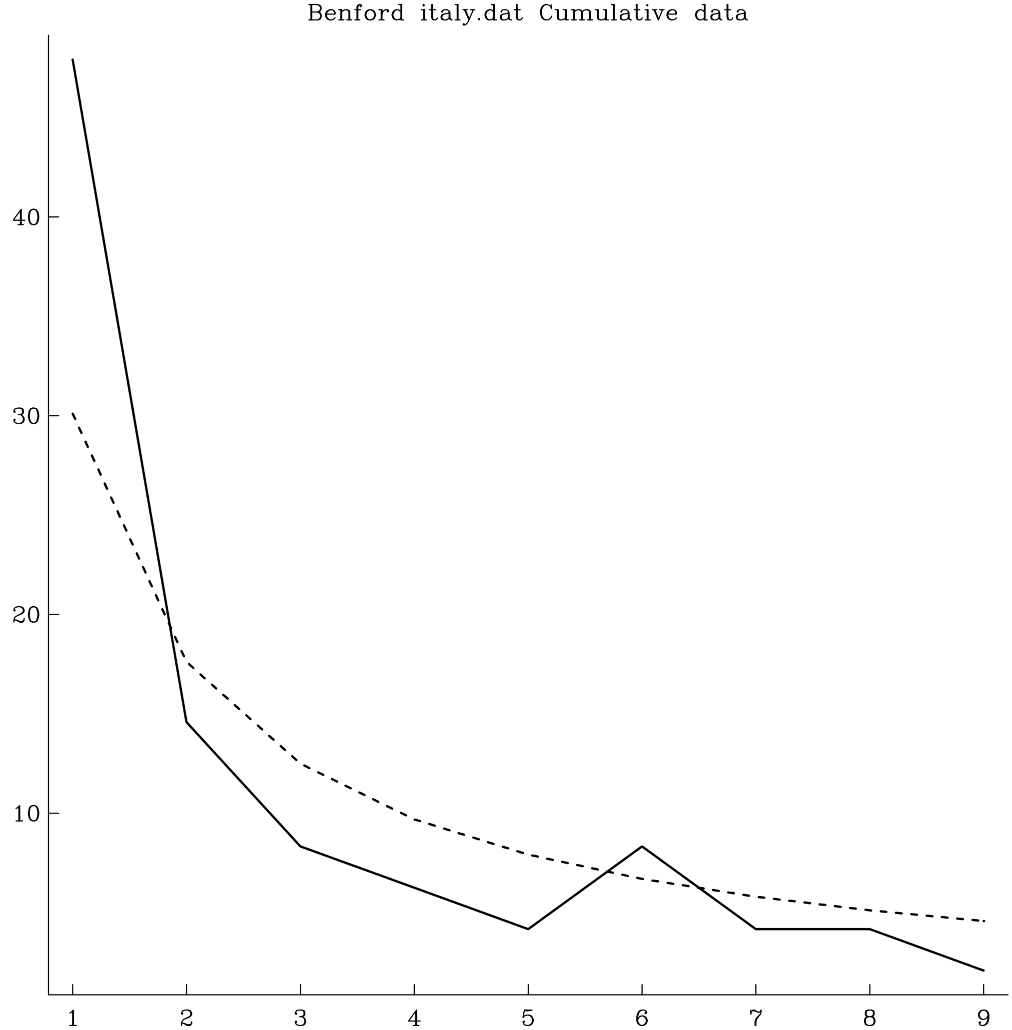

Benford iran.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

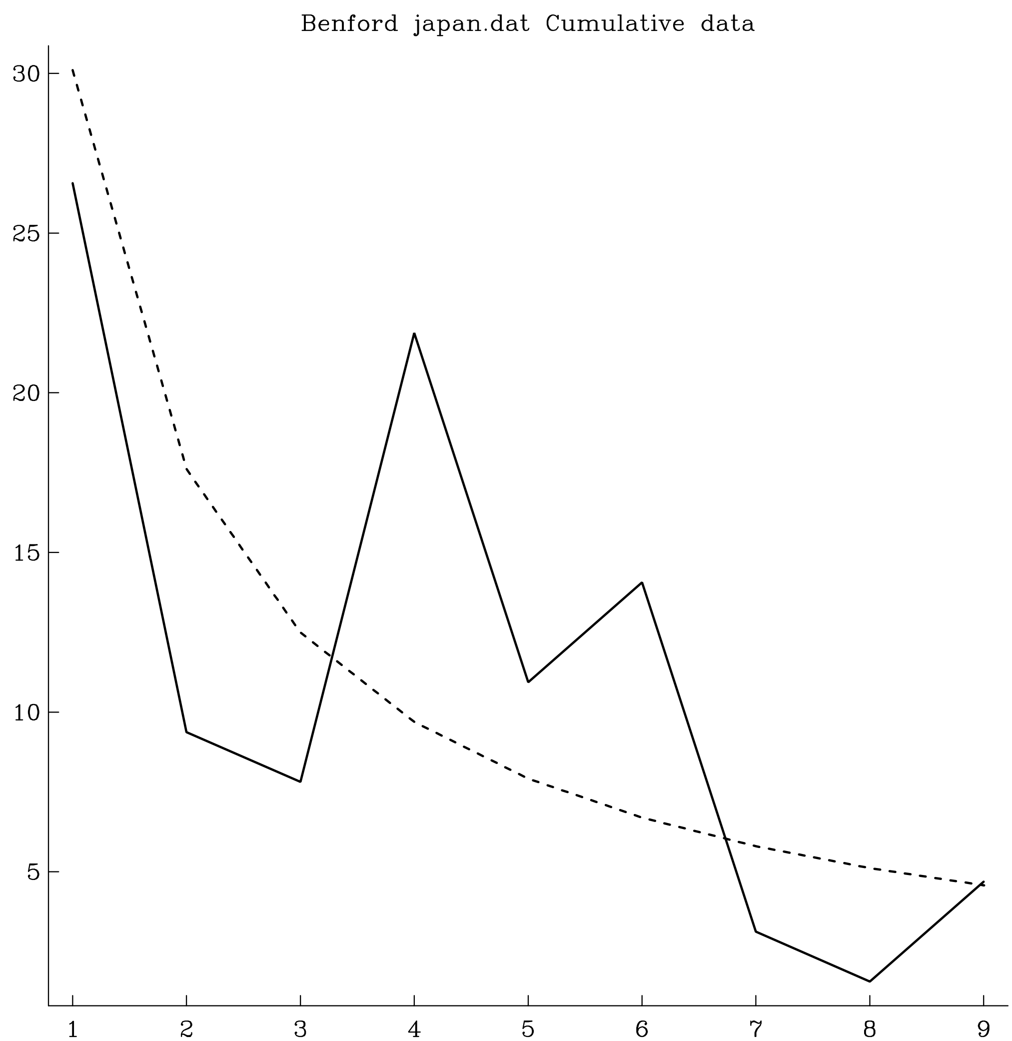

Benford ireland.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

Benford israel.dat Cumulative data



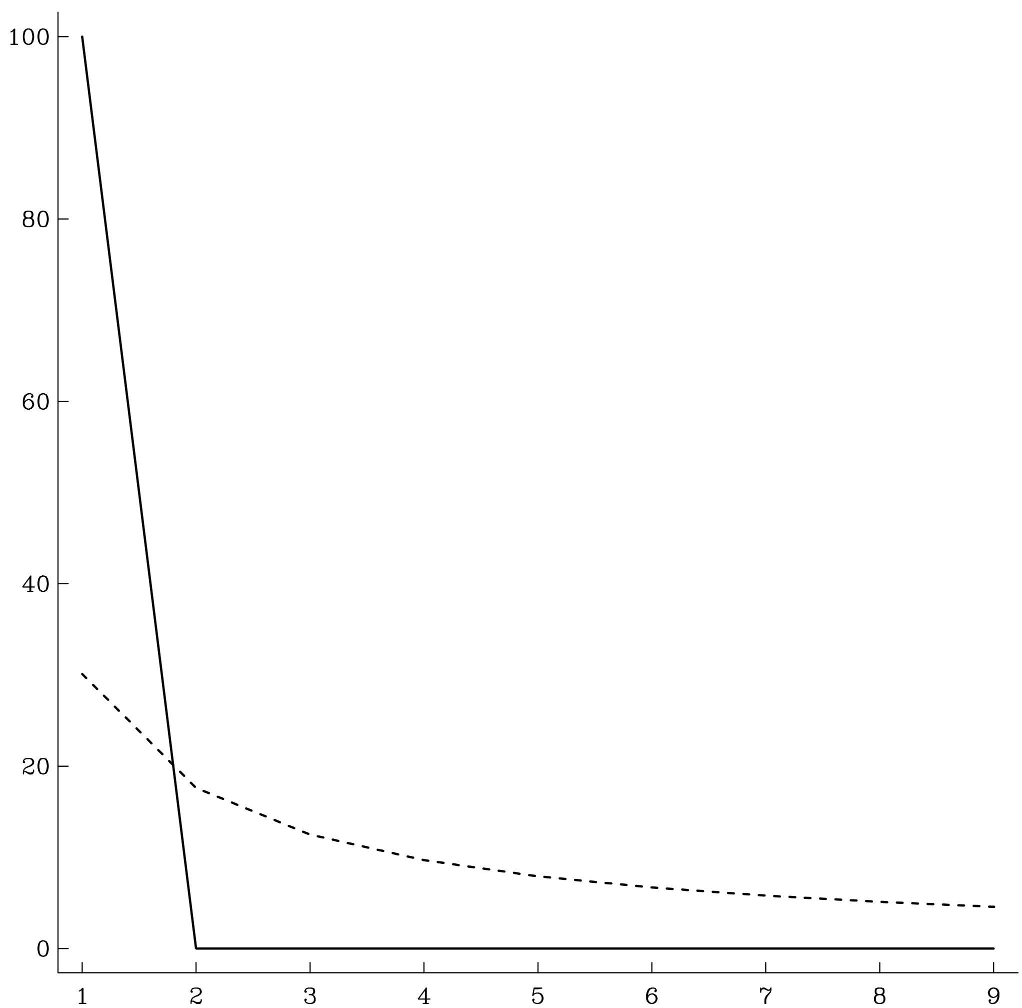
Benford italy.dat Cumulative data

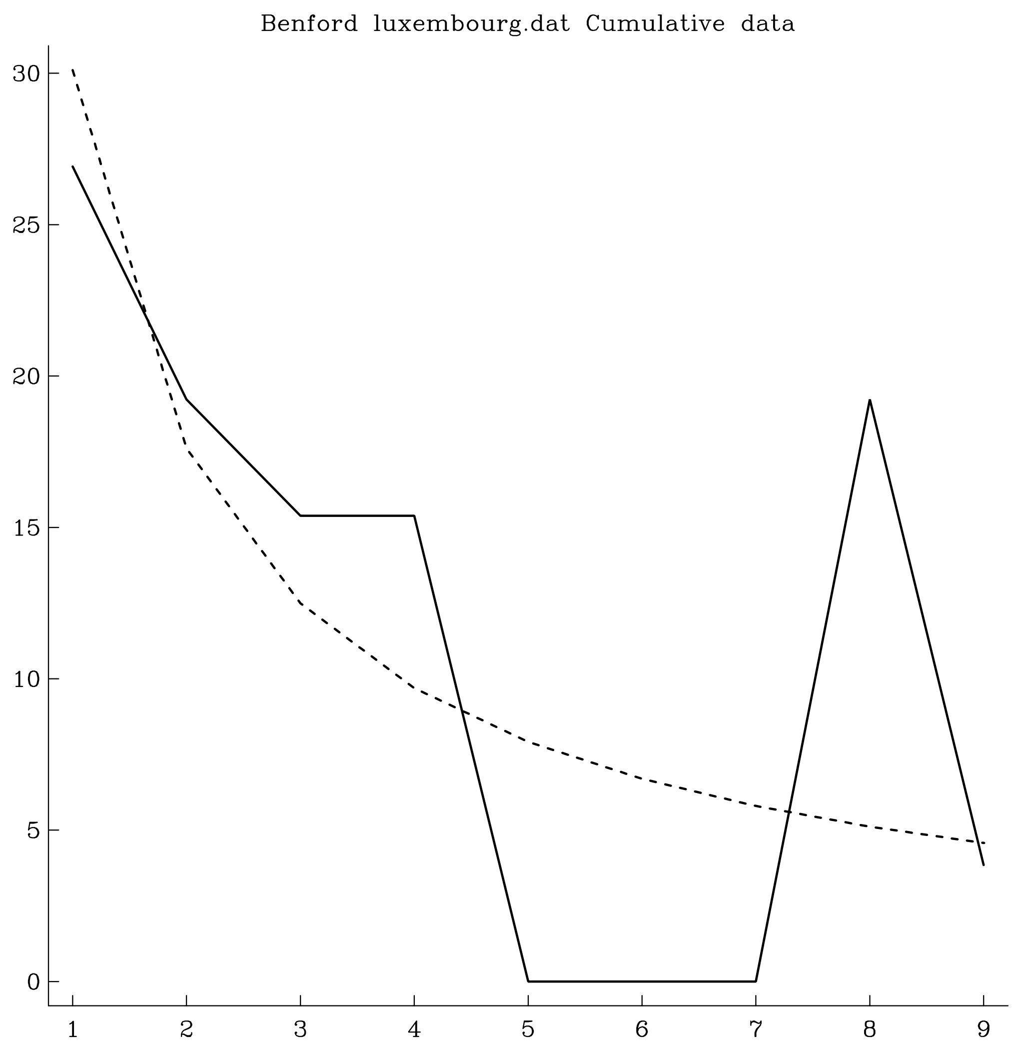




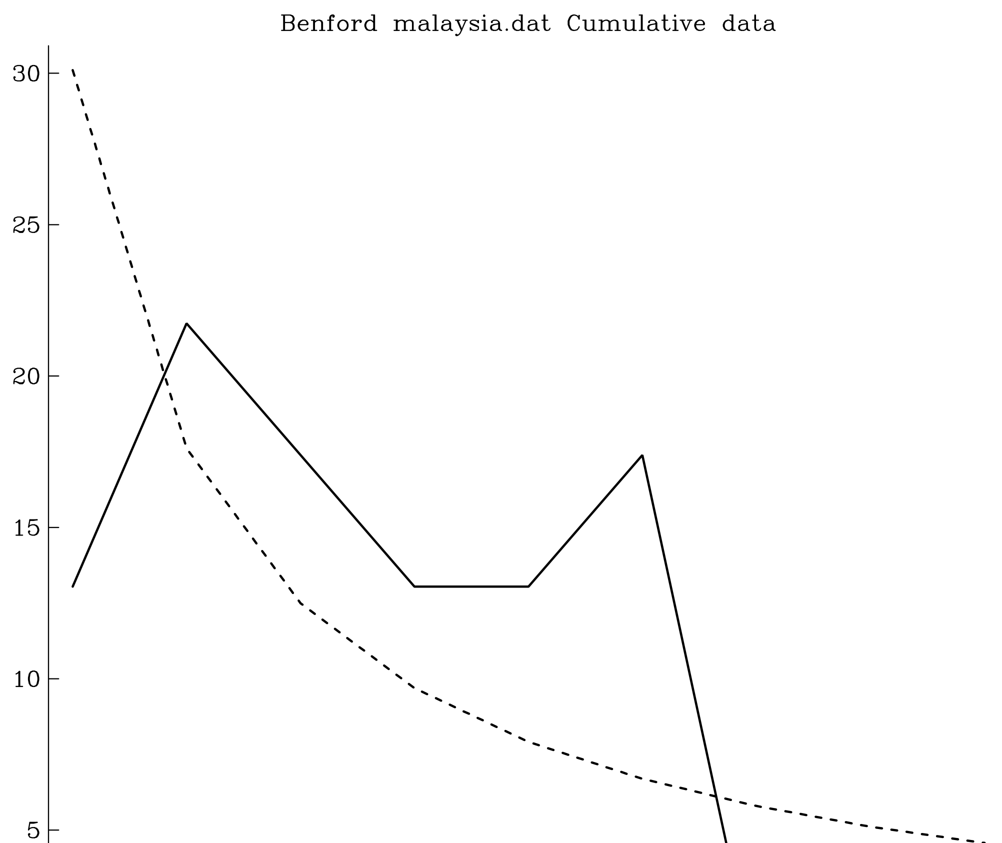
Benford japan.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

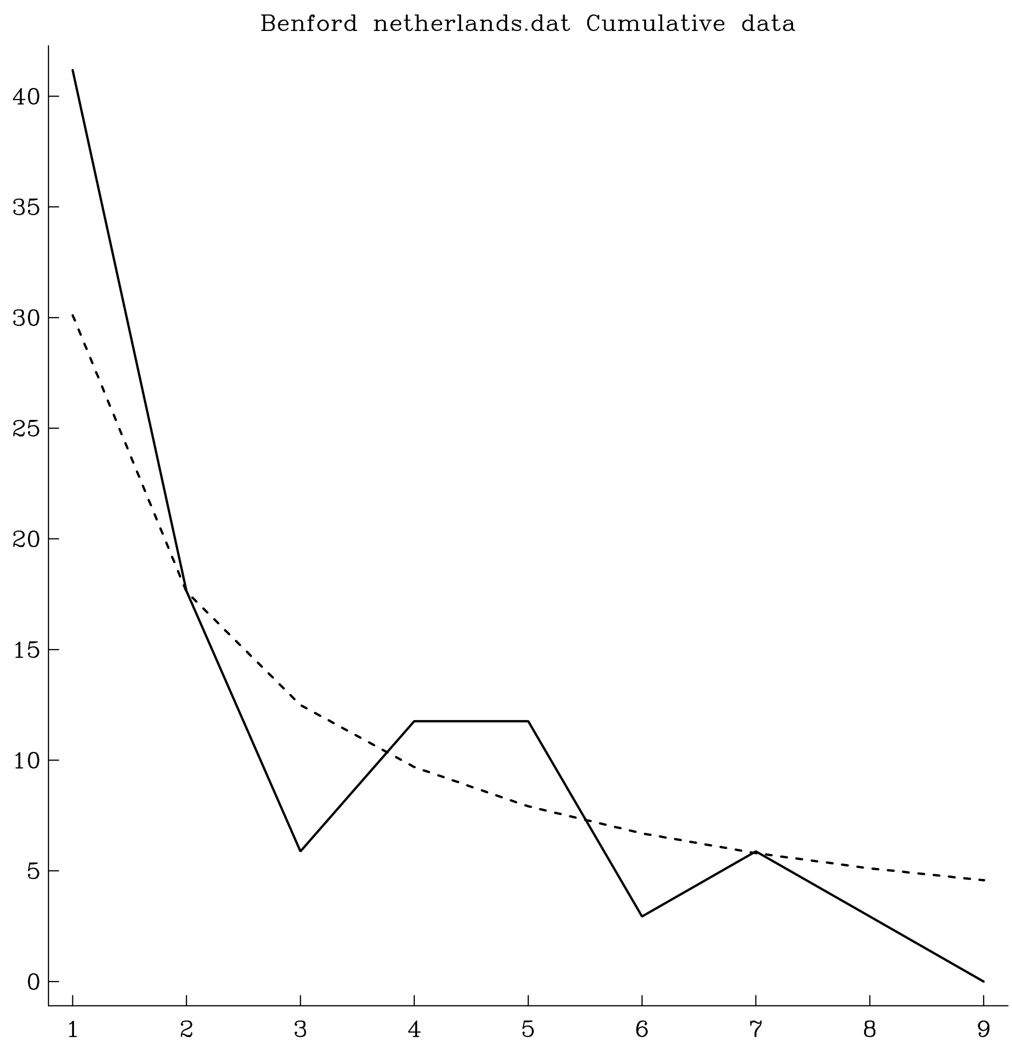
Benford kuwait.dat Cumulative data

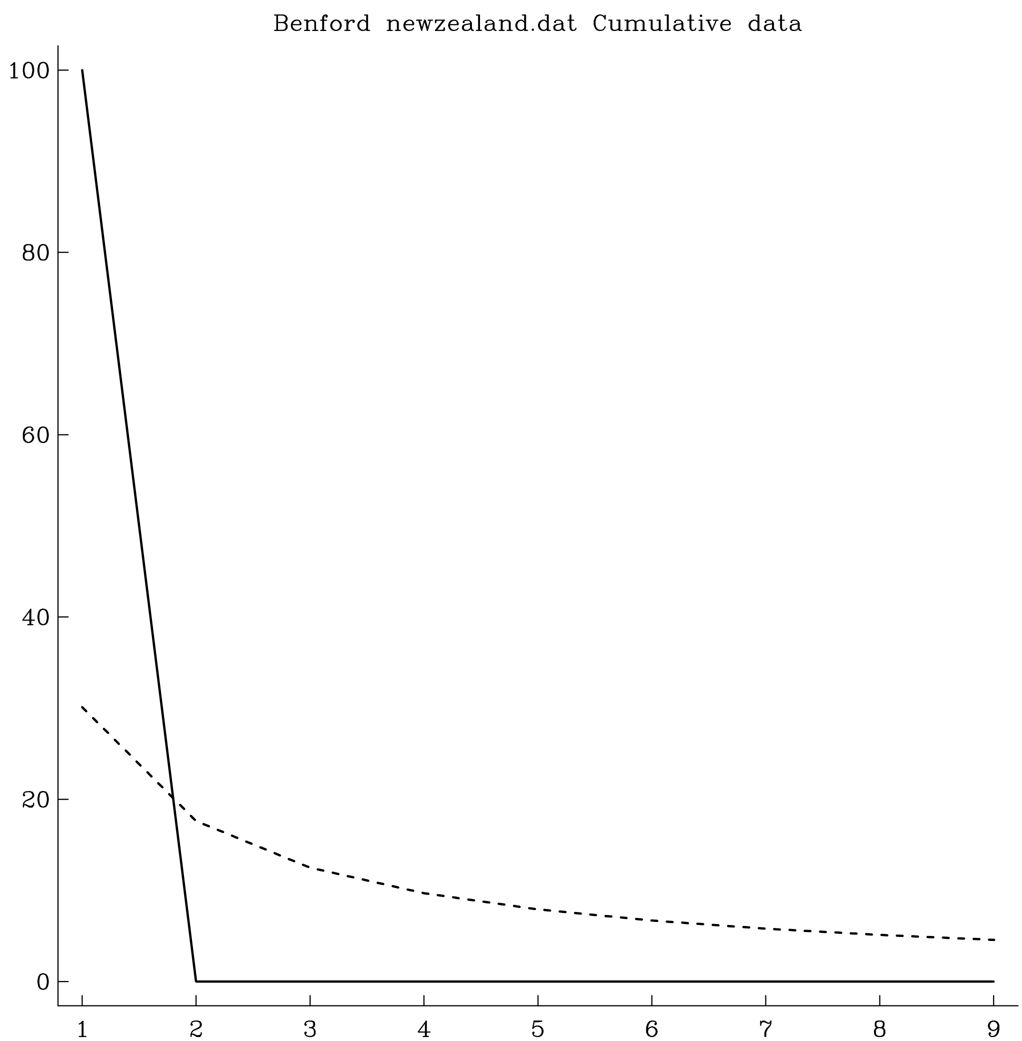


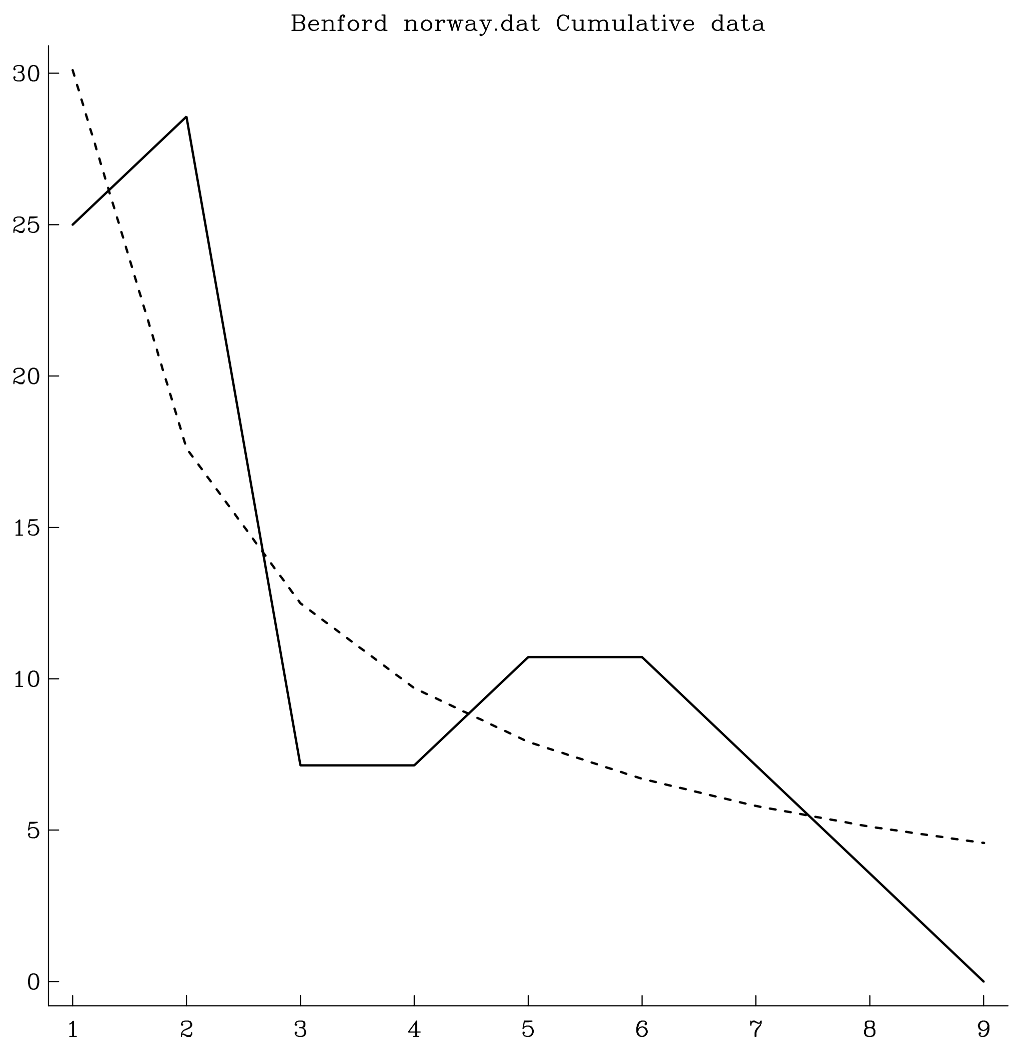


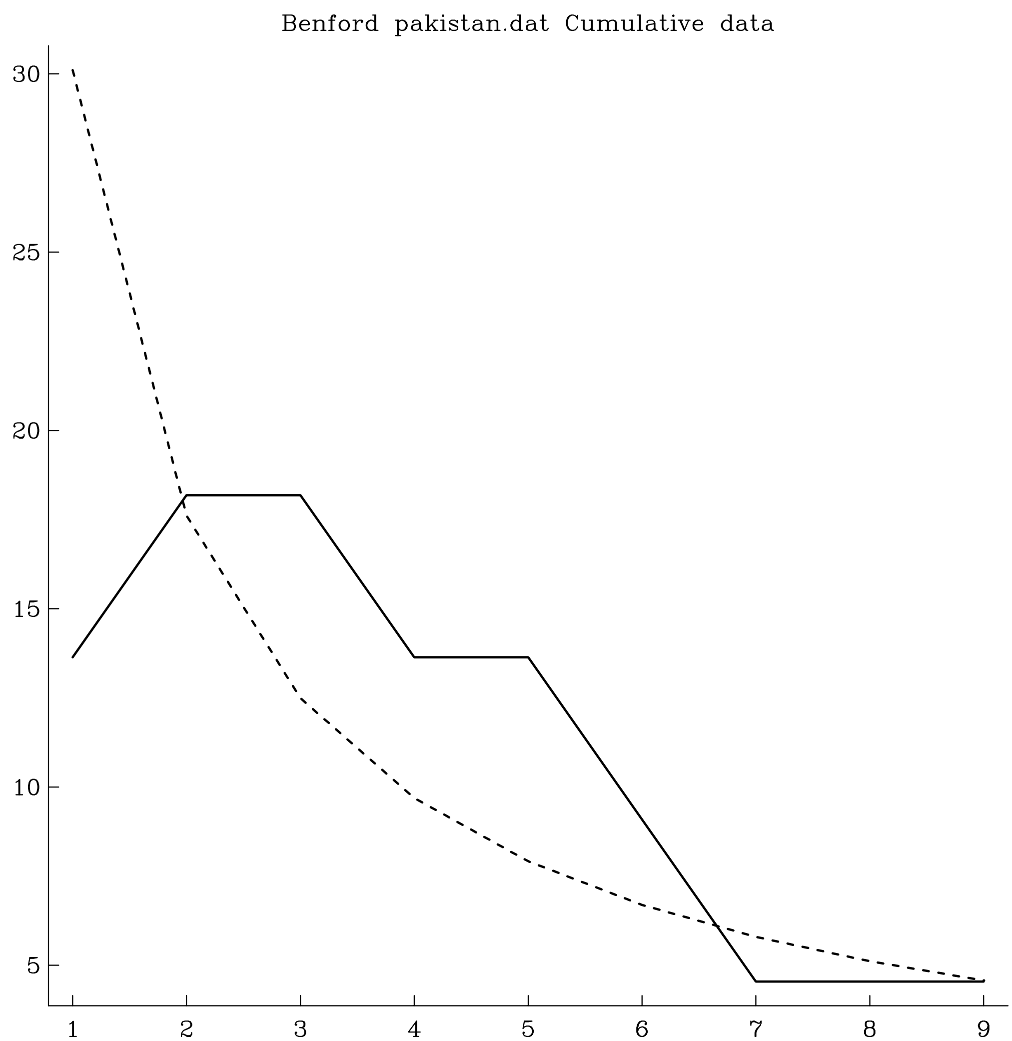
Benford luxembourg.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

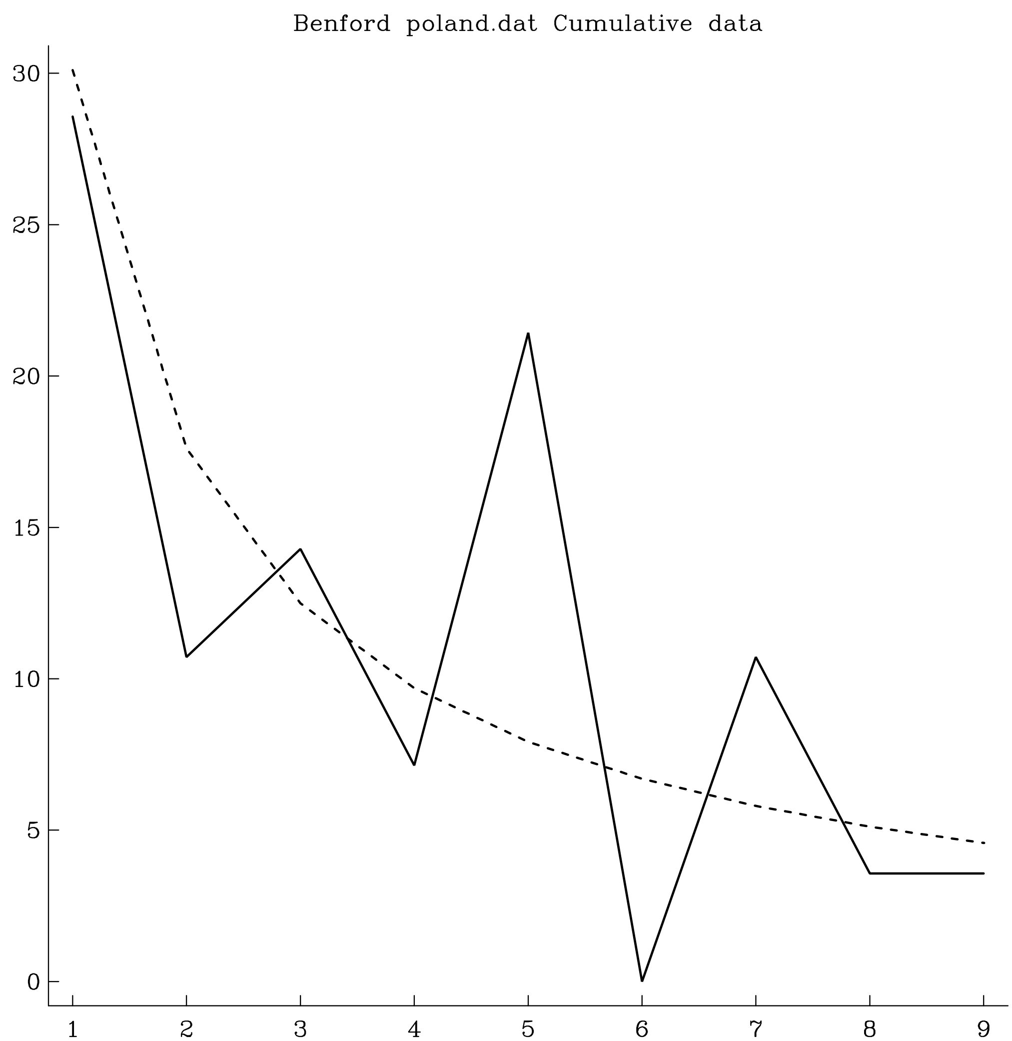

Benford malaysia.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9


Benford netherlands.dat Cumulative data
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

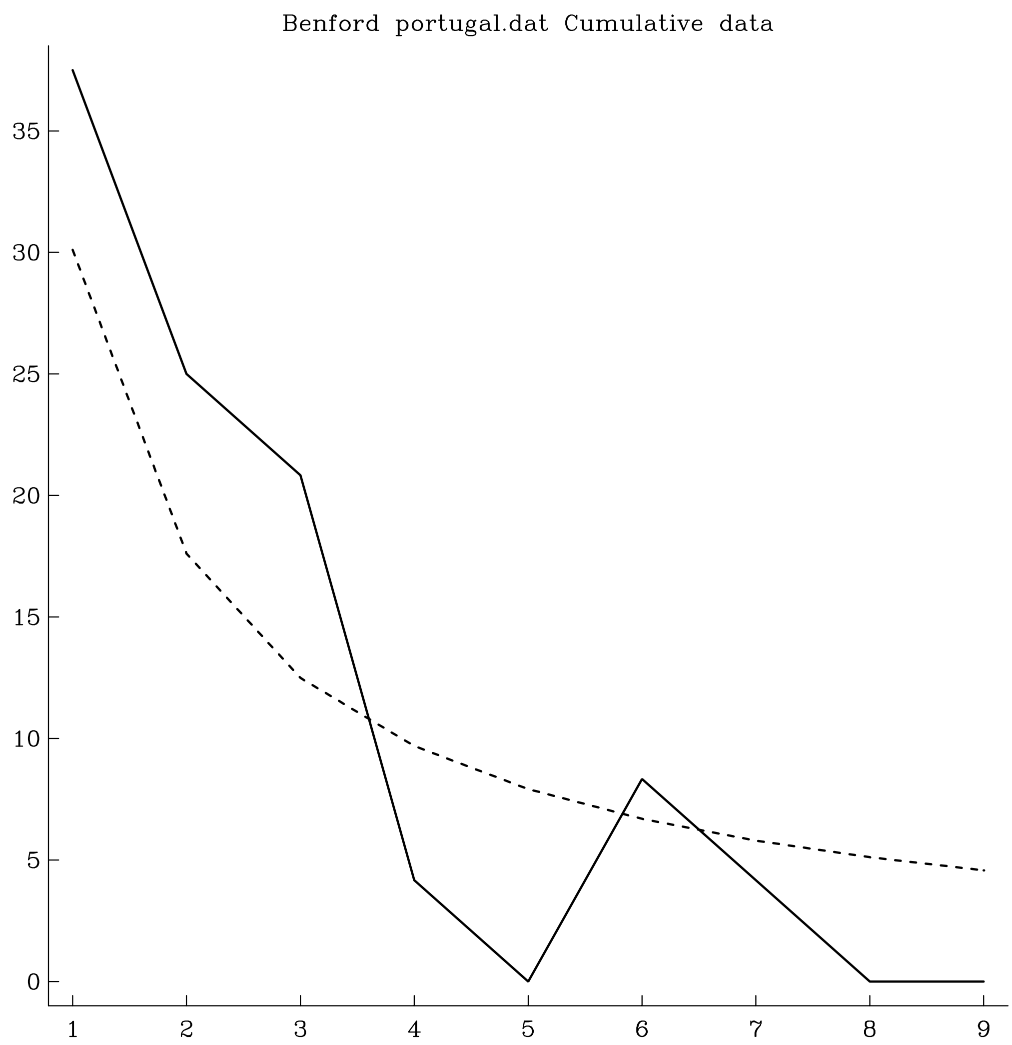

Benford newzealand.dat Cumulative data
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9

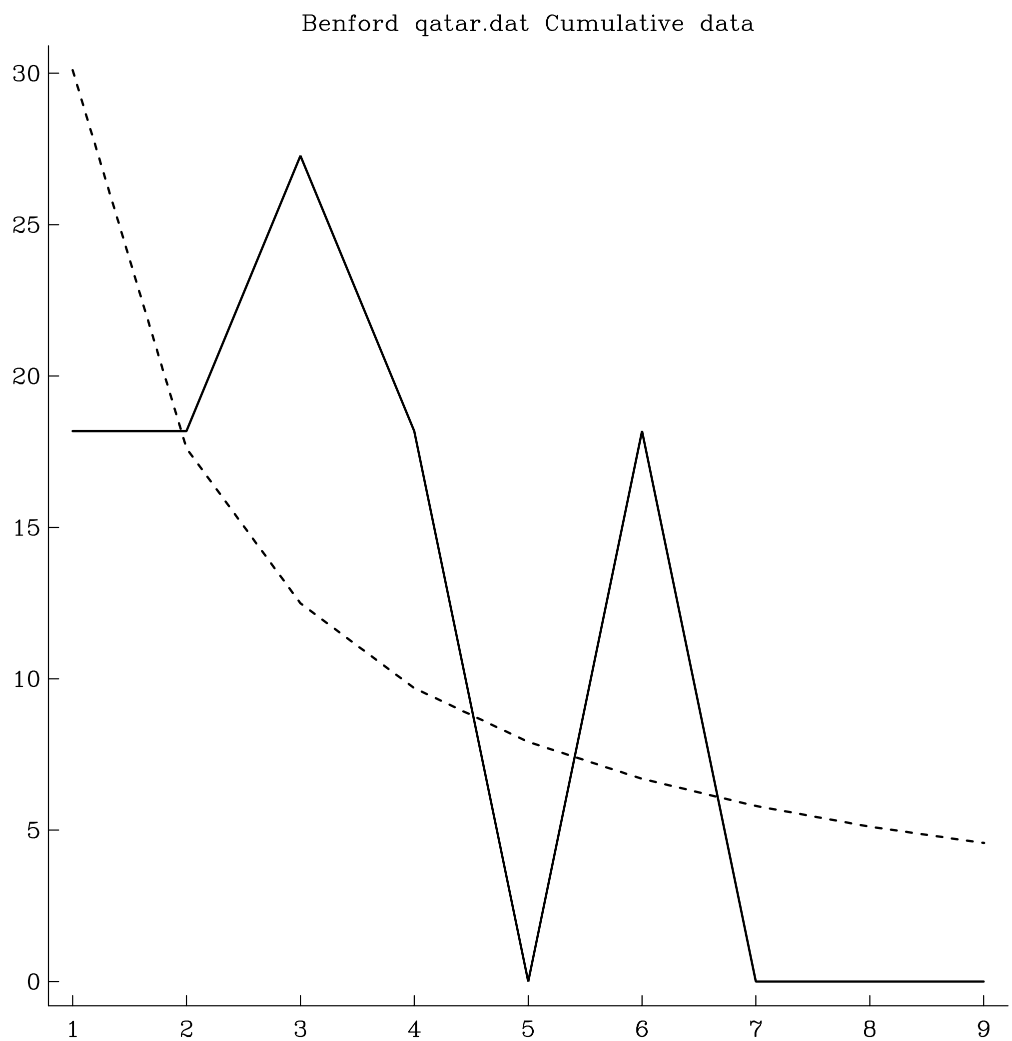

Benford norway.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9


Benford pakistan.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9

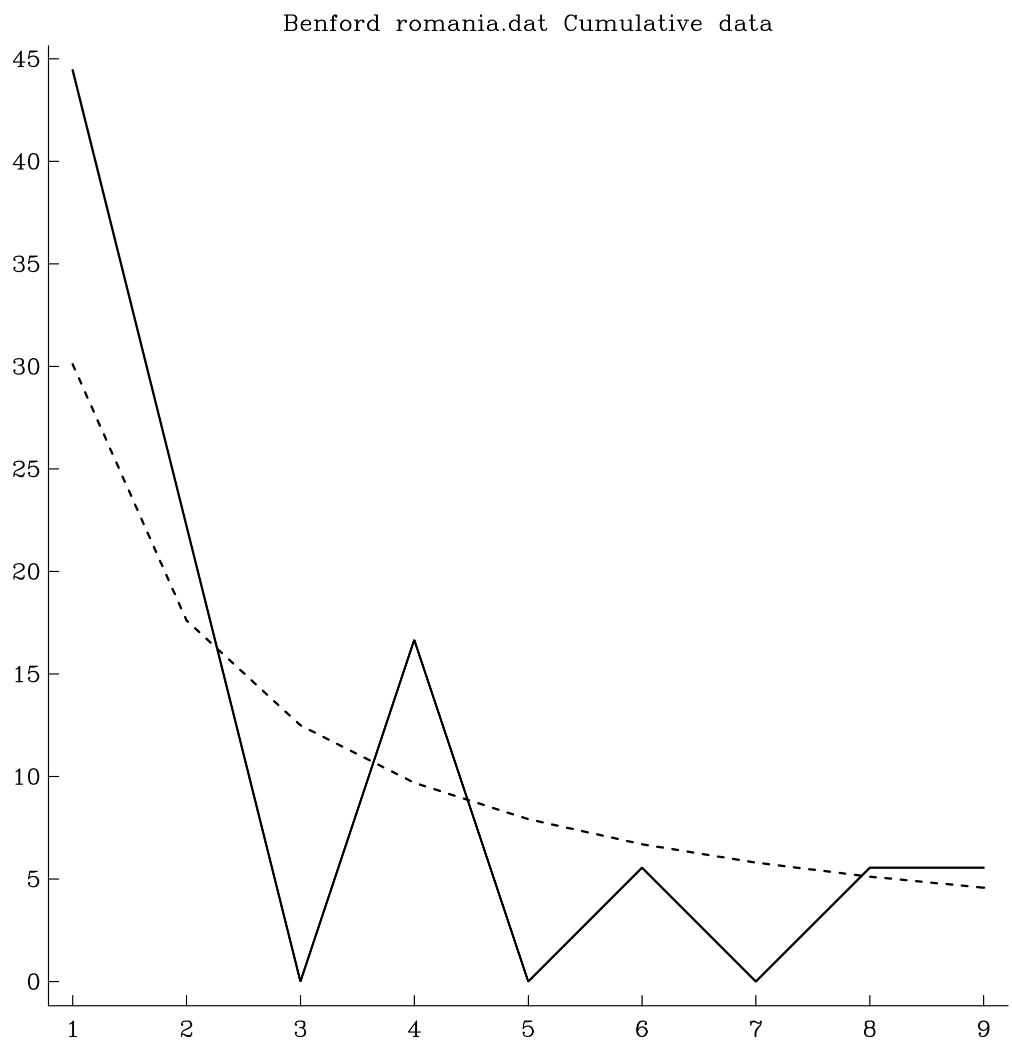


Benford poland.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9


Benford portugal.dat Cumulative data
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

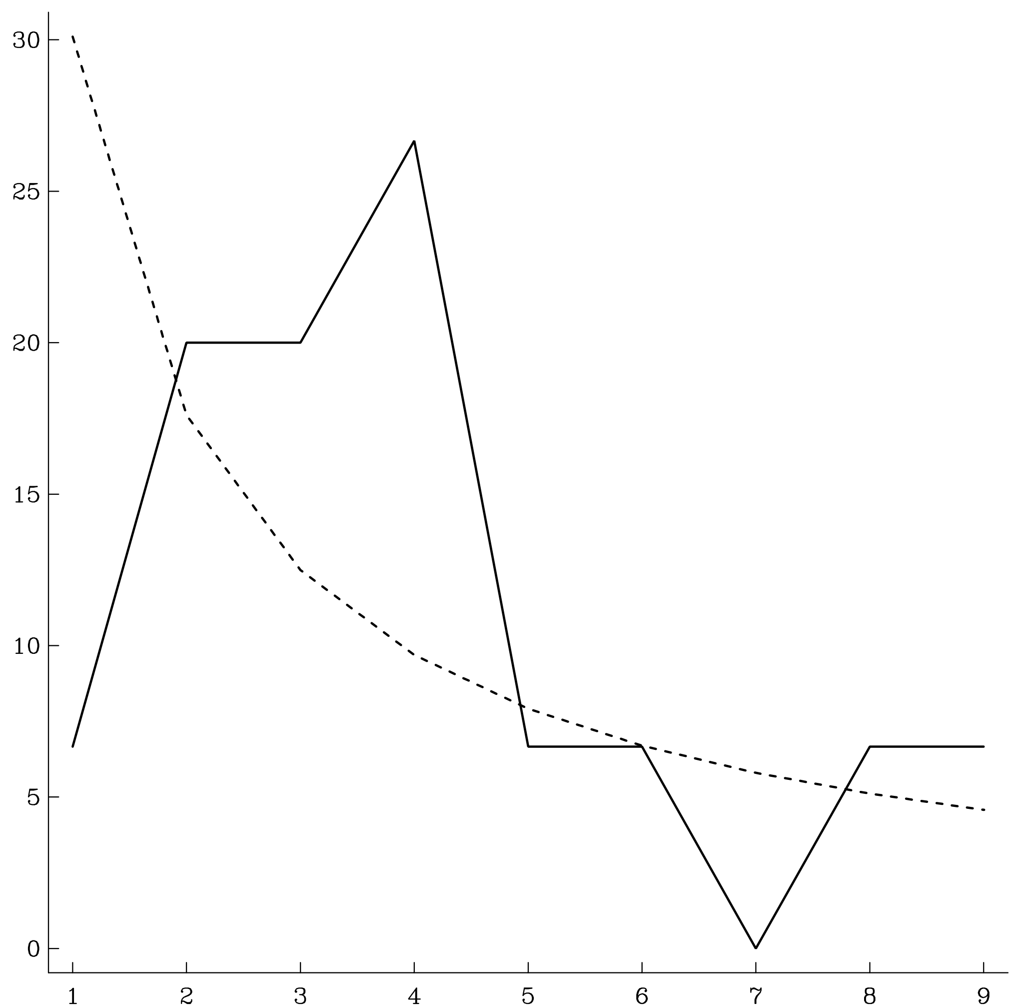


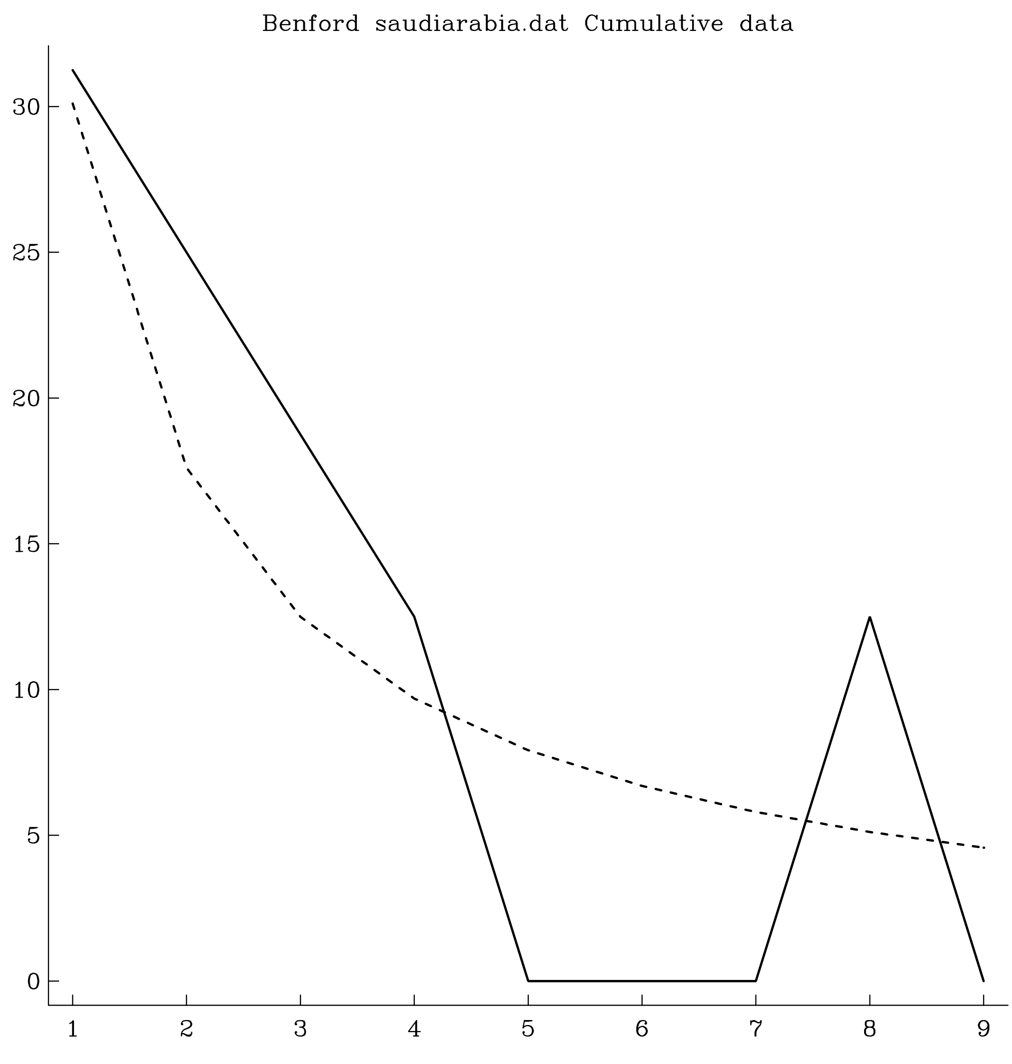
Benford qatar.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

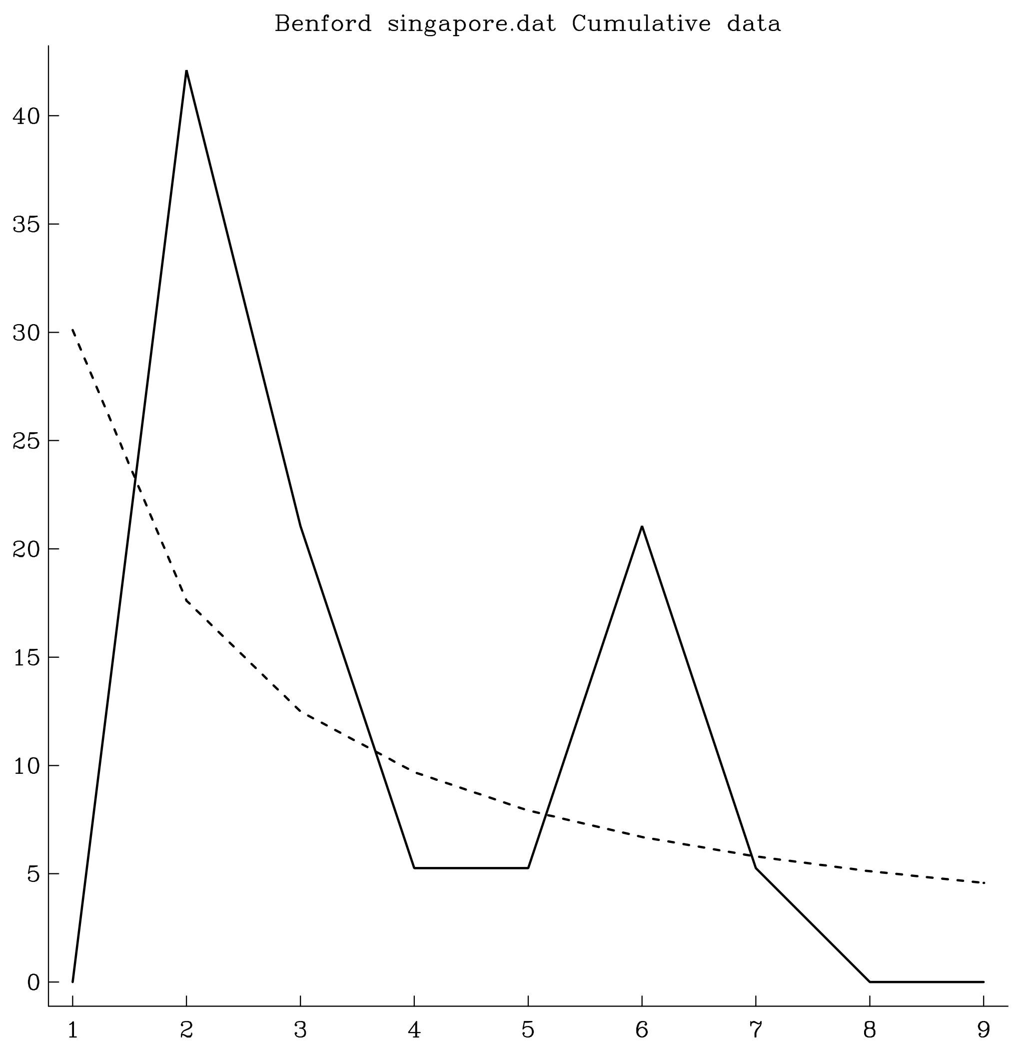


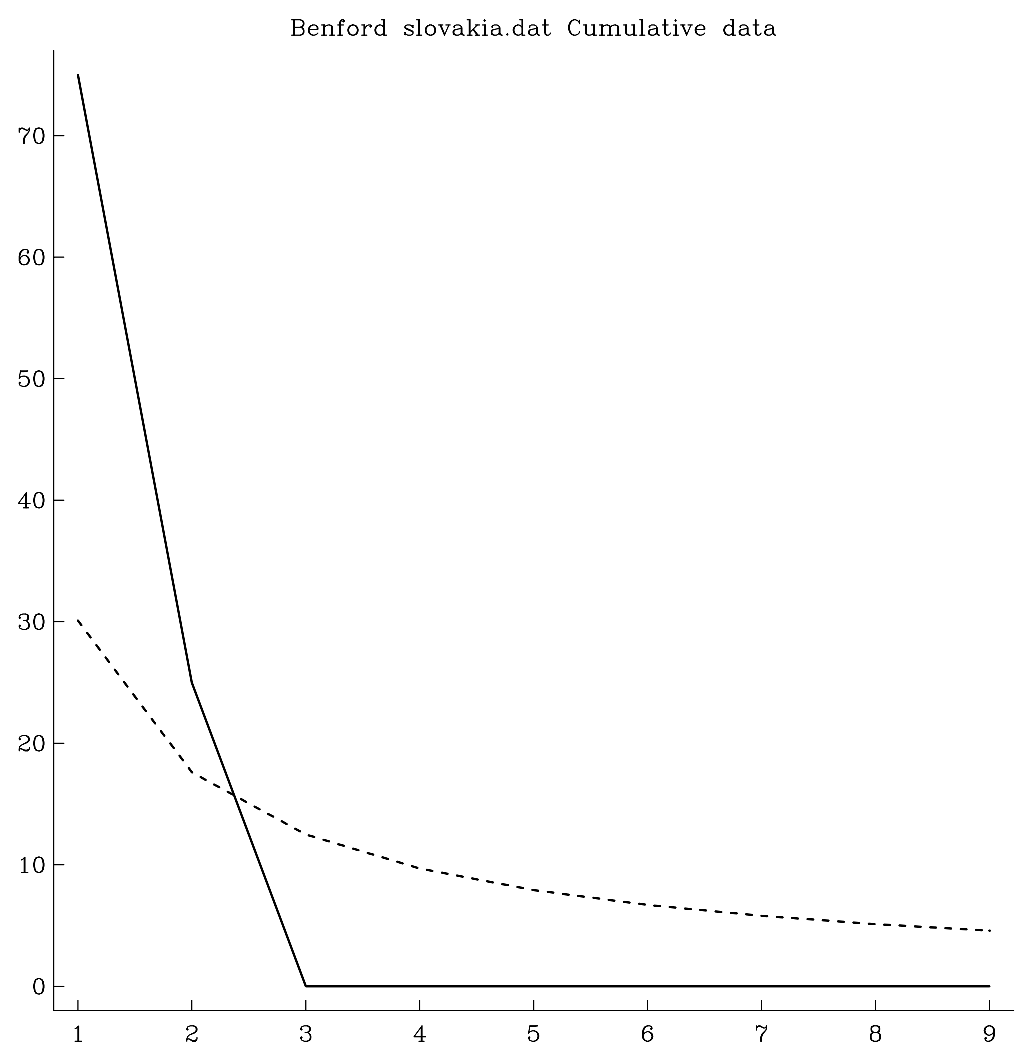
Benford romania.dat Cumulative data
45
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

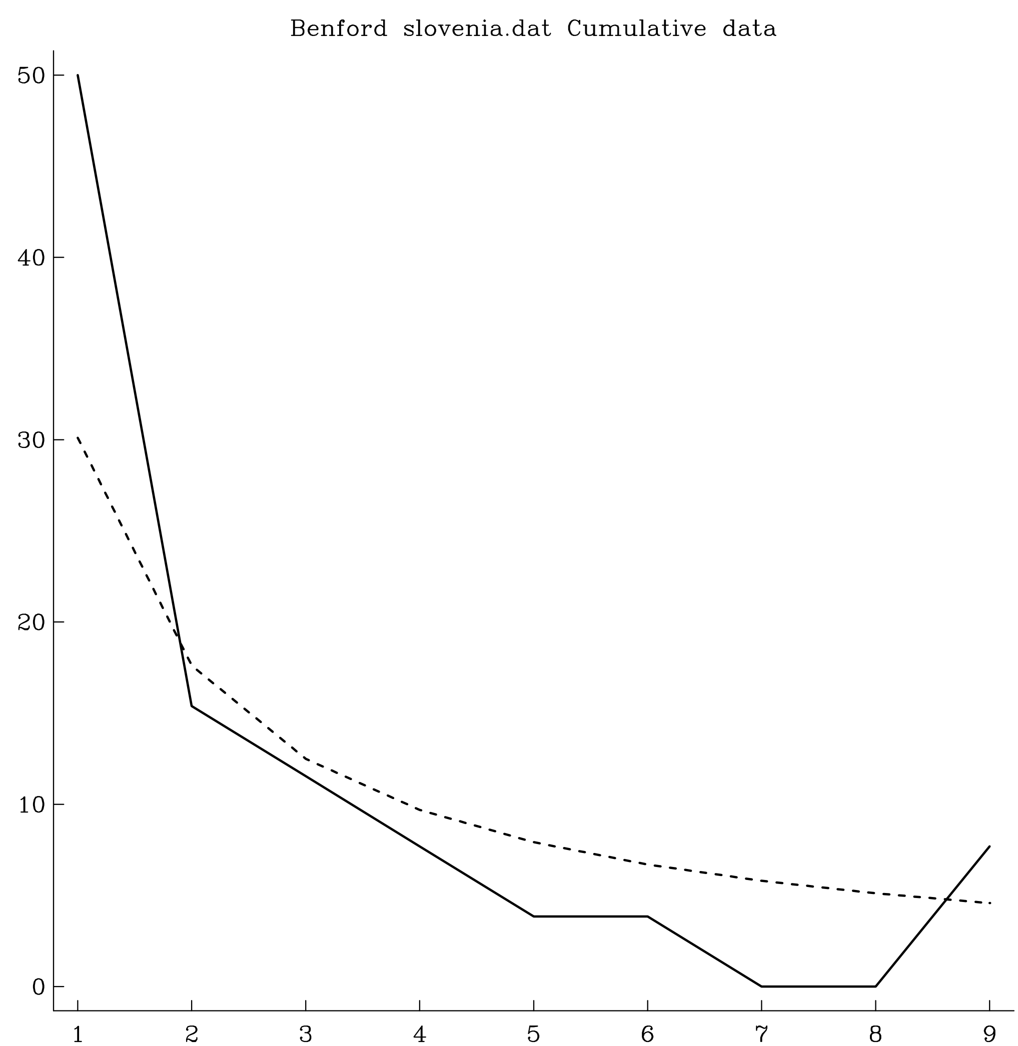
Benford russia.dat Cumulative data

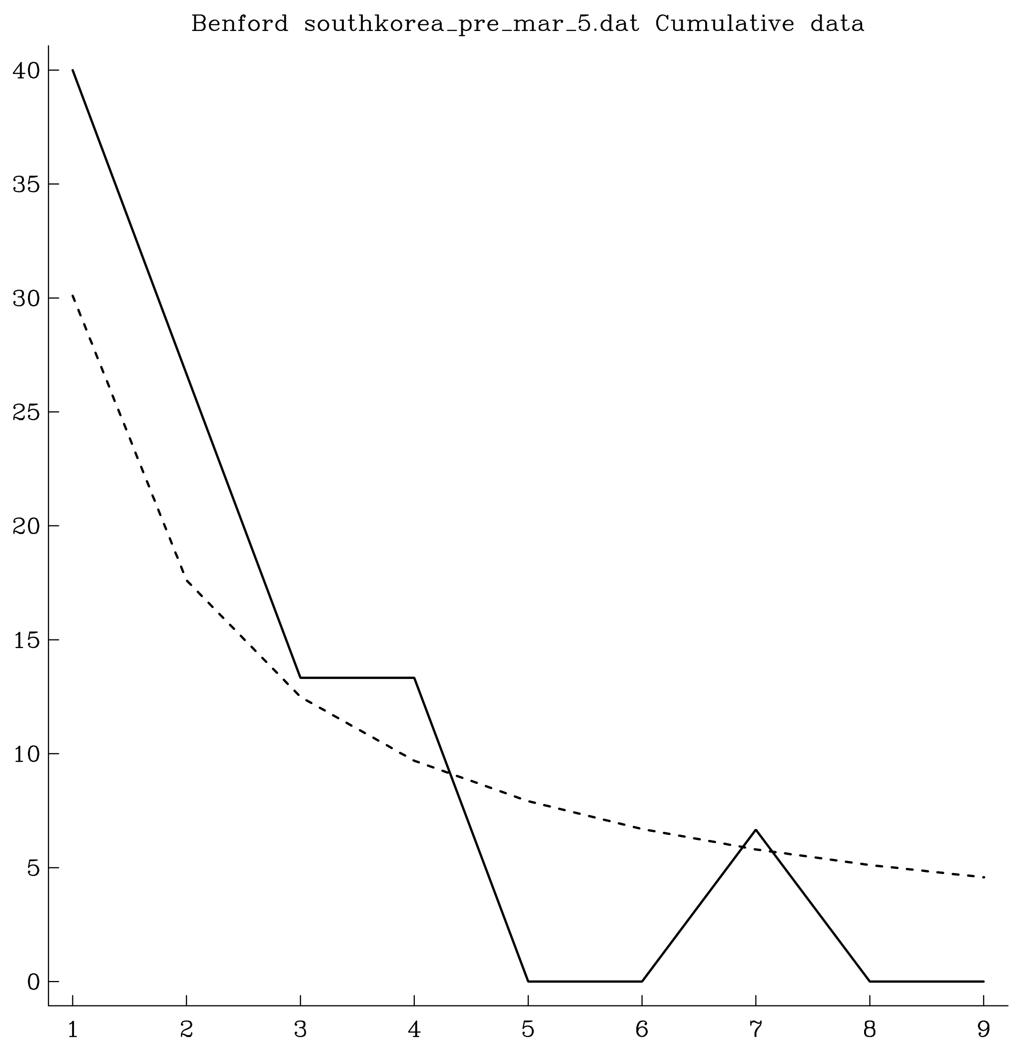



Benford saudiarabia.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

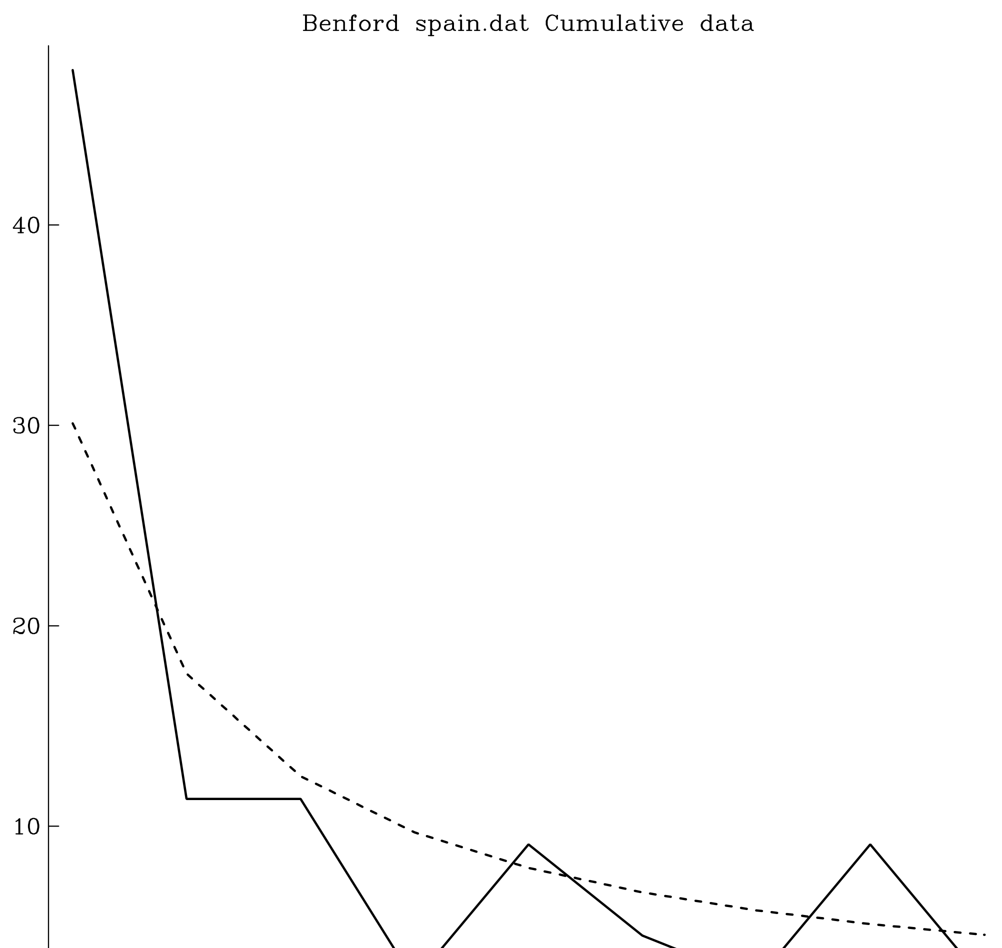

Benford singapore.dat Cumulative data
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

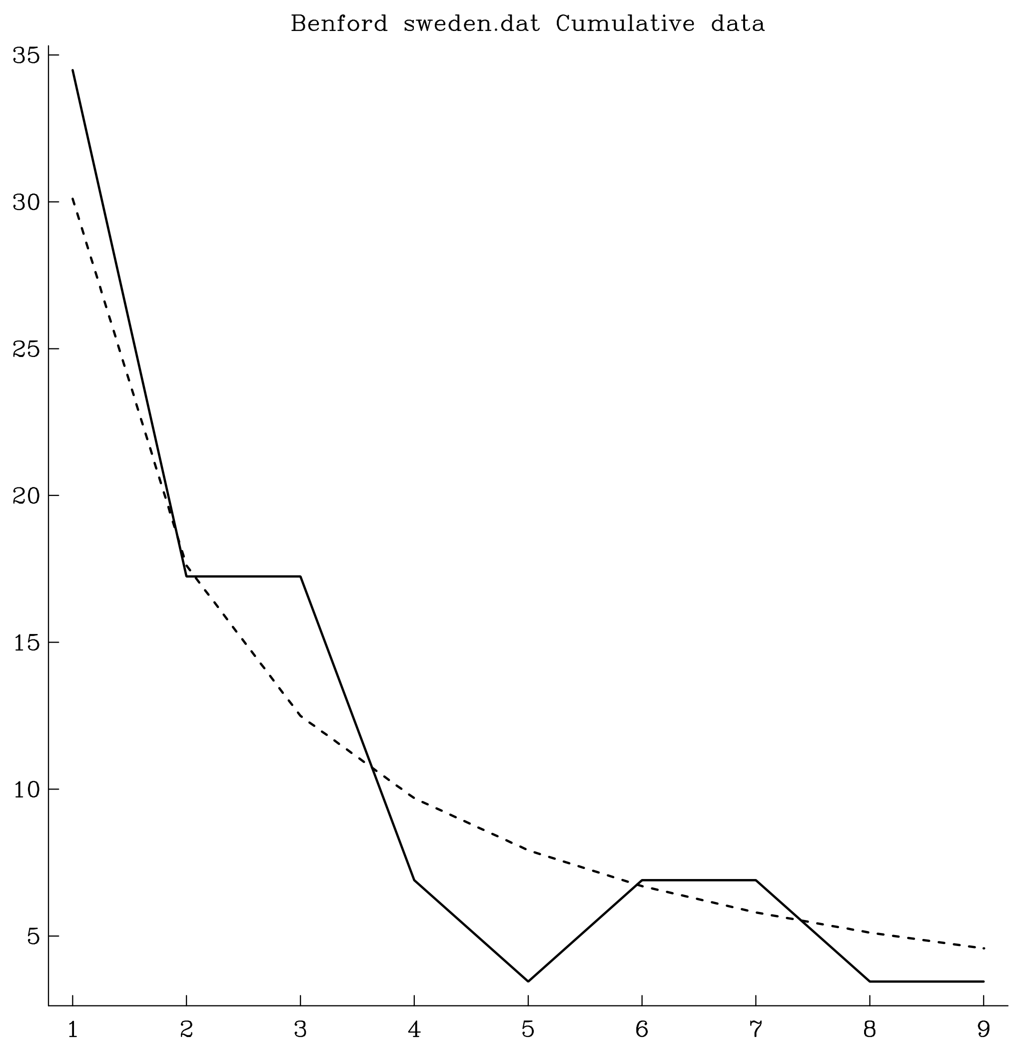

Benford slovakia.dat Cumulative data
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9

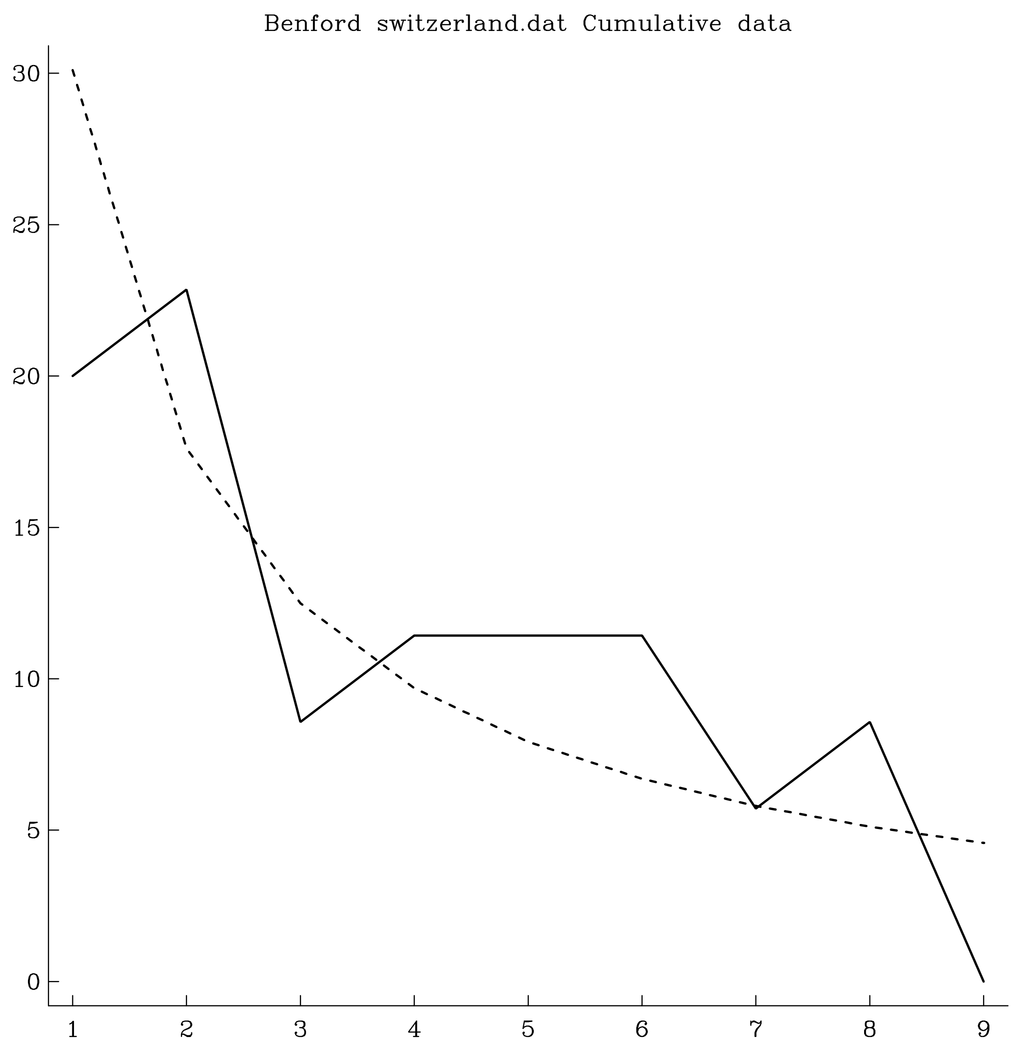

Benford slovenia.dat Cumulative data
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9

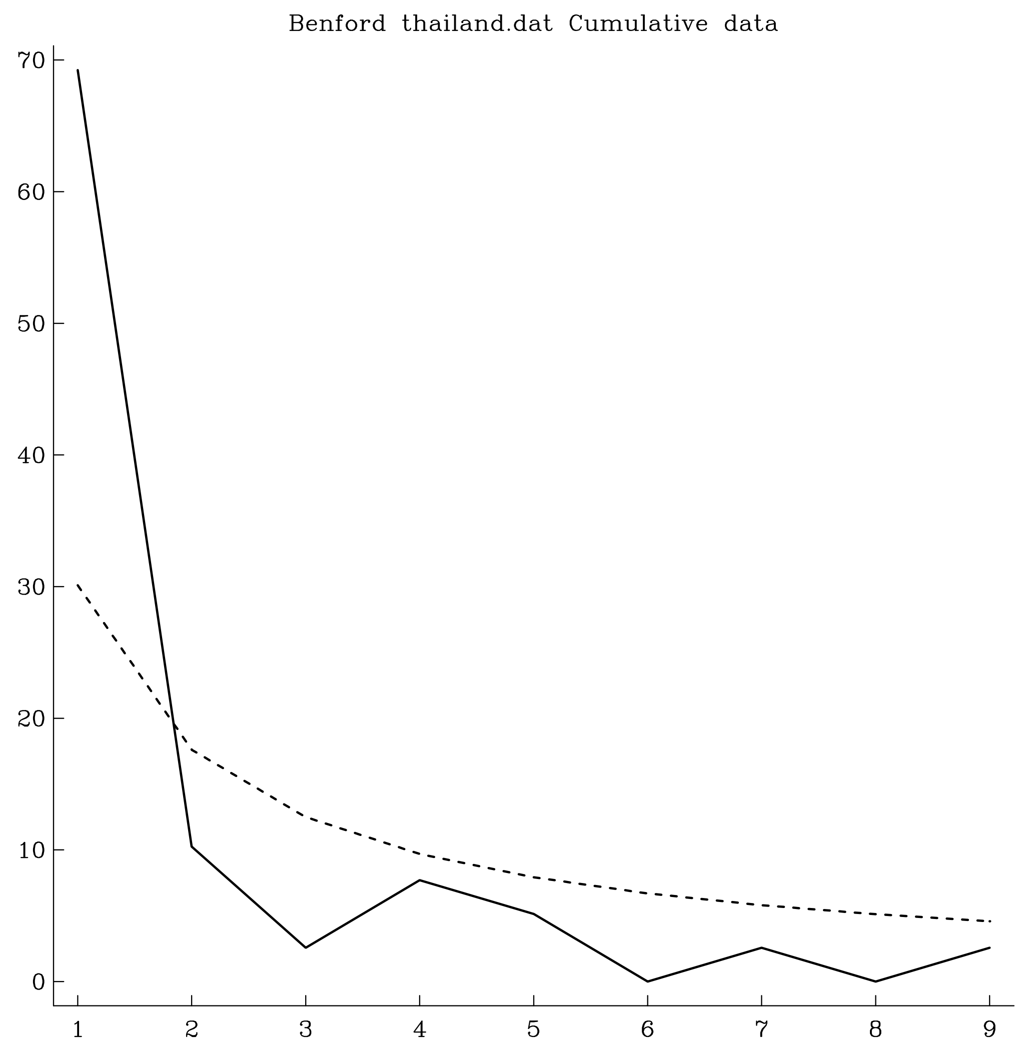


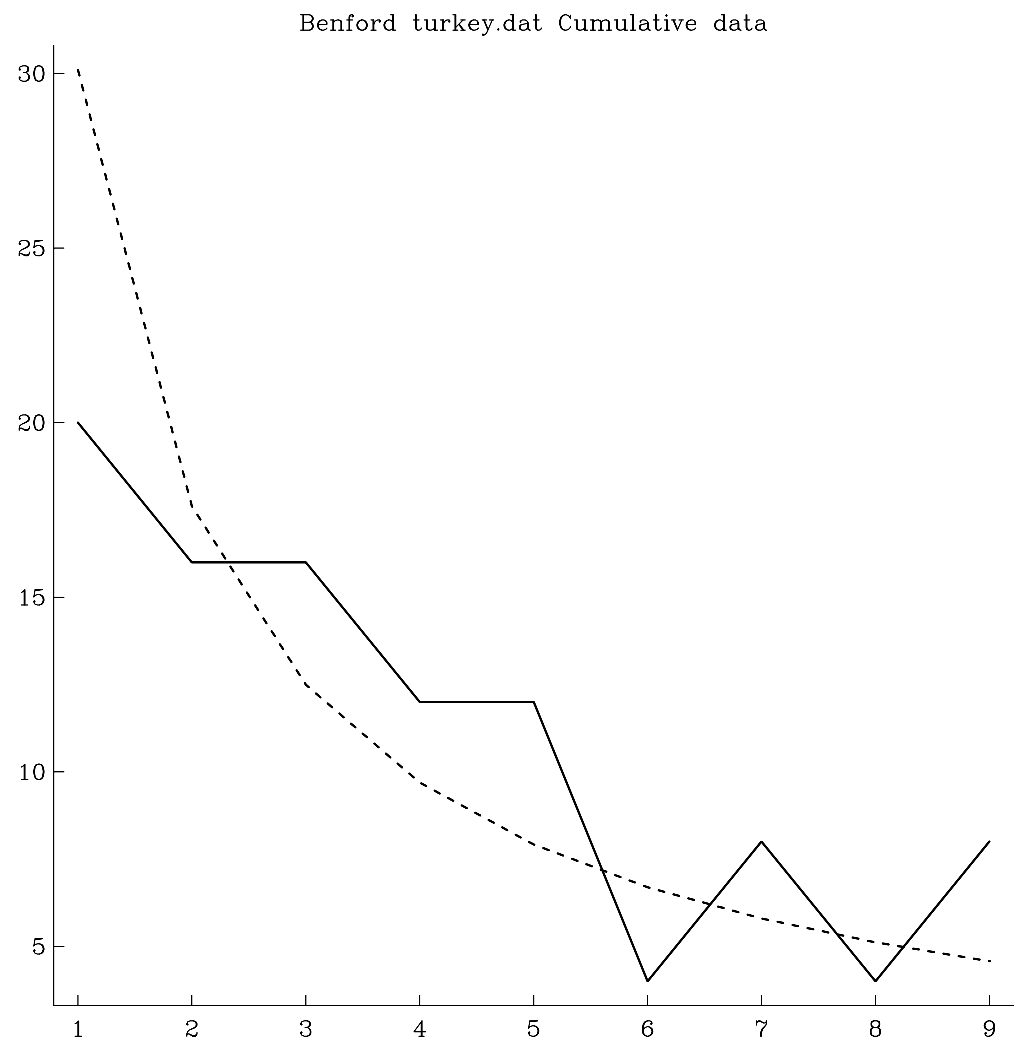
Benford southkorea_pre_mar_5.dat Cumulative data
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9

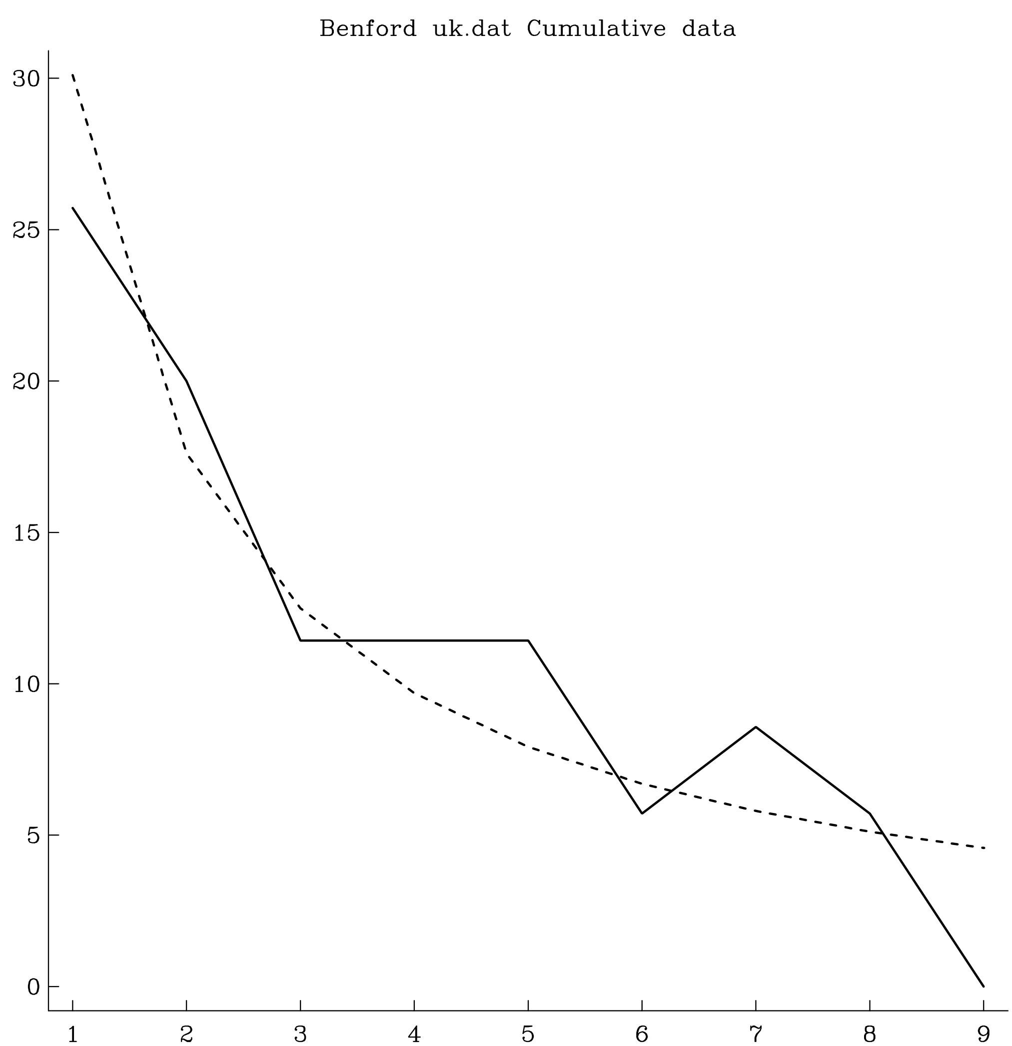

Benford spain.dat Cumulative data
40
30
20
10
1
2
3
4
5
6
7
8
9

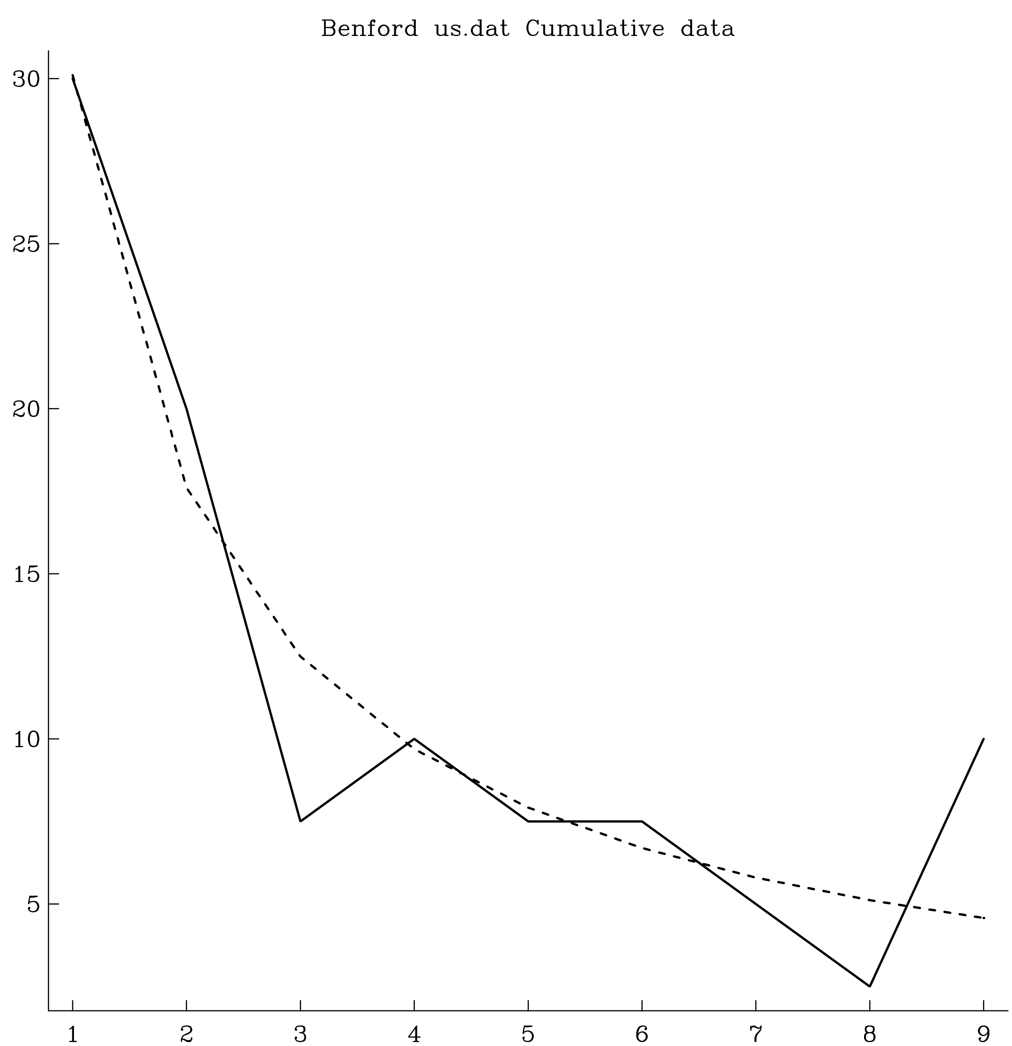

Benford sweden.dat Cumulative data
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9


Benford switzerland.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9


Benford thailand.dat Cumulative data
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9


Benford turkey.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9


Benford uk.dat Cumulative data
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9



Benford us.dat Cumulative data
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9