Benford’s Project

Figs 4,7: Summary Charts

Edward Stansfeld

January 2022

## Introduction

Nigrini suggests that Benford’s studies start with the Mantissa Arc test. In this test, the logarithms in base 10 of the data points are calculated. The mantissae are the part of the logarithm to the right of the decimal place. Benford’s Law can be defined as “The law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally probable” (Newcomb 1881), so the mantissae of a dataset, arranged in order of size, that follows Benford’s Law should plot along a straight diagonal line.

A further requirement for a Benford’s analysis is that the data should cover at least two orders of magnitude. As Kossovsky (2014) notes, the data at the extreme ends of the distribution produce different Benford’s distributions, with the smallest exhibiting a tendency to be uniform, while the largest can be characterised as Benford’s extreme. Accordingly, the number of data items in each order of magnitude were plotted as a test of suitability.

## Results

### Is the data suitable for a Benford’s analysis?

The mantissae are graphed below for the daily case data and the weekly deaths data. The line for the daily cases is close to the ideal and indicates broad suitability. The mantissa line for deaths is stepped and bowed and indicates that the number of records or range of data is deficient. This will have a bearing on two digit tests but the first digit should be OK.

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 1 rows containing non-finite values (stat\_smooth).

## Warning: Transformation introduced infinite values in continuous y-axis

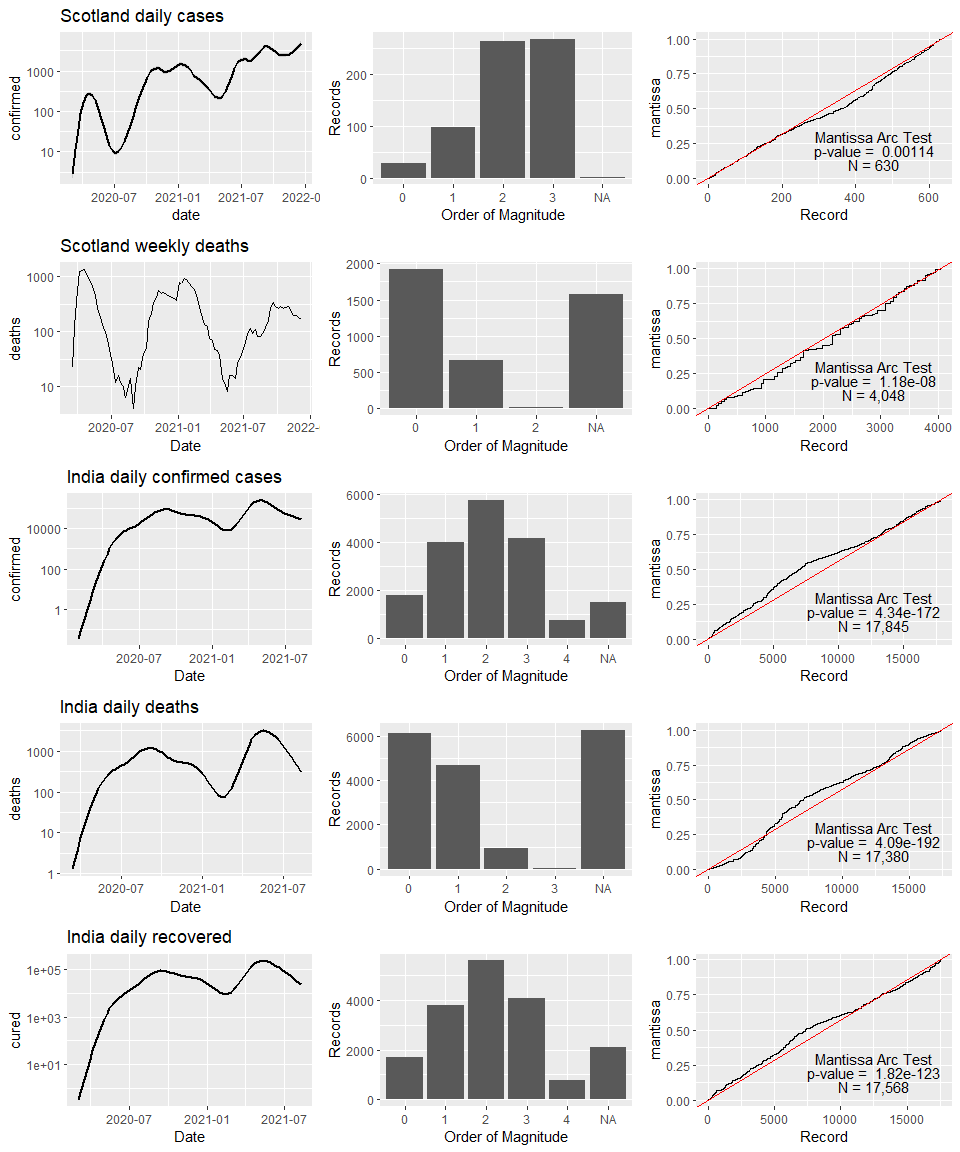
## Warning: Removed 75 rows containing non-finite values (stat\_smooth).

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 249 rows containing non-finite values (stat\_smooth).

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 227 rows containing non-finite values (stat\_smooth).



## Conclusion

The Scotland deaths dataset are deficient for tests of Benfords Law owing to their limited range and preponderance of low values. This causes the distributions to be heavily skewed by the “Benford’s extreme” effect.

It was established that the pandemic dataset is expected to follow Benford’s Law for natural data, because of the exponential way in which an epidemic spreads. It was also established that there are sufficient records and that they have sufficient range for a Benford’s study.

## References

Nigrini, M. (2012) Benford’s Law: Applications for Forensic Accounting, Auditing, and Fraud … - Mark J. Nigrini - Google Books. Available at: <https://books.google.co.uk/books?hl=en&lr=&id=Bh5Vr_I1NZoC&oi=fnd&pg=PP11&ots=qcigiizrOi&sig=Zyv0EV3snhVKH3yOC9z117ieSlI&redir_esc=y#v=onepage&q&f=false> (Accessed: 11 August 2021).