Benford’s Project

Figure 12: Last Digit

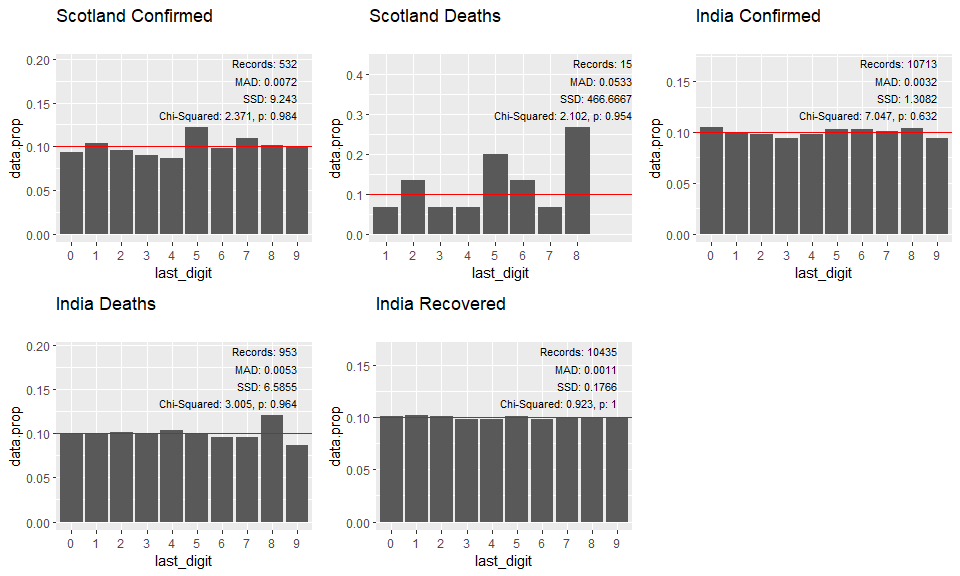
Edward Stansfeld

January 2022

## Introduction

The last digits should have a uniform distribution, and peaks at 0 or 5 would indicate estimates, while peaks at 2 or 8 could indicate a tendency to round away from psychologically important thresholds (Nigrini 2012). The modulus 10 and modulus 100 were used in a function to extract the last one and two digits respectively.

The graphs for the last two digits were prepared after excluding numbers below 99. This limits the deaths dataset to fewer than 1,000. Both distributions are generally uniform as expected for the last digits. Intriguingly, the deaths data has a peak at “8”, and a trough at “9”. However, the statistics calculated on the distribution do not indicate an issue, and they could arise by chance.



## 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

DataScienceScotland (2021) COVID-19-Management-Information: The data provide past data for the daily updates published by the Scottish Government, GitHub. Available at: <https://github.com/DataScienceScotland/COVID-19-Management-Information> (Accessed: 21 December 2021).

Scottish Government (2021) Deaths involving coronavirus (COVID-19), statistics.gov.scot. Available at: <https://statistics.gov.scot/resource?uri=http%3A%2F%2Fstatistics.gov.scot%2Fdata%2Fdeaths-involving-coronavirus-covid-19> (Accessed: 21 December 2021).