Benford’s Project

Fig 14: Timing Chart

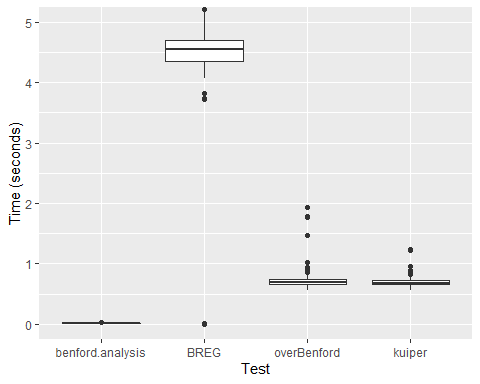
Edward Stansfeld

January 2022

## Introduction

The India data contain 36 sets of regional data, each containing around 500 records, and the Scotland 14 sets of regional data. This paper focuses on the properties of the datasets for confirmed cases, deaths and recoveries. The distributions are plotted for each region and indicative statistics are calculated.

ggplot(theTimings\_long, aes(x = Test, y = `Time (seconds)`)) + geom\_boxplot() + coord\_cartesian(ylim = c(0,  
 5))



## Results

As shown in Appendix 1, the state results have a wide variety, though most are clearly conformant over the leading digits, especially among the confirmed and cured cases. There are some notable exceptions; Jammu and Kashmir have a distribution that has uniform features (but passes Kossovsky’s test) and Tamil Nadu and Uttarakhand have a disproportionate number of “4”s and 5"s. West Bengal has a strange two-layered uniform distribution.

The deaths data has a more limited range and there is a tendency for the Benford’s extreme distribution to appear. The Benford’s extreme distribution is associated with digital progression (Kossovsky 2019). In digital progression, when the logs of sampled data covering several orders of magnitude are binned, they should approximate to a normal distribution. The rising end of the distribution tends to digital equality, whereas the falling end tends to Benford’s extreme. In the state deaths data, many of the magnitude plots have a sharply right-skewed distribution, and this is a predictor that the Benford’s distribution will be extreme.

The last digit results also have a wide variety, and most of them look conformant. The deaths plots are much affected by the limited range and in several cases, the last digit distributions are more like first digit distributions. This would seem to be a digital progression effect, with the high end effect dominant over the low end.

## Conclusion

The state-level data has a wide range of results, with most passing at least one of the tests for Benford’s conformance. There are also some that appear anomalous, in particular West Bengal, Uttarkhand and Tamil Nadu, which performed strangely in the leading digit tests. Visual inspection showed that some of the deaths graphs have the “Benford’s extreme” profile of more low digits than expected. It is particularly common with datasets that have a limited range of just one or two orders of magnitude. A Benford’s extreme profile is associated with digital progression and is caused by a reducing number of data points as the log of the data increases.