Package 'benford'

July 20, 2021

Type Package

Title Benford's Analysis on Large Data Sets

Version 1.0.1

Author Vitor Hugo Moreau

Maintainer Vitor Hugo Moreau <vitorhmc@ufba.br>

Description Perform the Benford's Analysis to a data set in order to evaluate if it contains human fabricated data. For more details on the method see Moreau, 2021, Model Assist. Statist. Appl., 16 (2021) 73–79. <doi:10.3233/MAS-210517>.

License Creative Commons Attribution 4.0 International License

Encoding UTF-8

LazyData true

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2021-07-20 18:10:02 UTC

R topics documented:

---

benford   *Benford's analysis*

Package 'benford'

July 20, 2021

Type Package

Title Benford's Analysis on Large Data Sets

Version 1.0.1

Author Vitor Hugo Moreau

Maintainer Vitor Hugo Moreau <vitorhmc@ufba.br>

Description Perform the Benford's Analysis to a data set in order to evaluate if it contains human fabricated data. For more details on the method see Moreau, 2021, Model Assist. Statist. Appl., 16 (2021) 73–79. <doi:10.3233/MAS-210517>.

License Creative Commons Attribution 4.0 International License

Encoding UTF-8

LazyData true

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2021-07-20 18:10:02 UTC

R topics documented:

---

benford   *Benford's analysis*

Description

Benford's analysis makes use of a statistic property of natural data sets called Benford's Law. Benford's Law (also called "first digit phenomenon") is a statistical phenomenon that describes the frequency of a given integer, from 1 to 9, to be in the first significant digit in the numbers of a large data set. The Benford's law has been most practically used to detect fraud or rounding errors in real world numbers. This is possible by examining departures in the frequencies of individual digits from those predicted by Benford. This only makes sense once it is established (often empirically) that the data follow the law under normal circumstances (Sambridge et al., 2011). This is true because human pseudo-random productions are in many ways different from true randomness (Nickerson, 2002). As a consequence, fabricated data might fit to the Benford's Law to a lesser extent than genuine data (Banks and Hill, 1974; Gauvrit et al., 2017). Benford package is able to analyze the frequence of the first, second, first-two and first-three digits in large data sets.

Usage

```
benford(x, plot = FALSE, mode = 1)
```

Arguments

x	A numeric vector with the data set numbers to be analyzed
plot	A logic that control whether the resulting first digit distribution and the Benford's distribution would be ploted
mode	A numeric value (1, 2, 12 or 123) to select, respectively, first digit, second digit, first-two digits or first-three digits analysis

Value

LIST countaining: 1. Named vector with three elements: the Chi Square test p value (p), the root mean square deviation (RMSD) from the Benford's distribution, and the log of the likelihhod of the first digit distribution in relation to the Benford's distribution; 2. Matrix, with three columns, countaining the first digits ([,1]), the frequency counts of the first digit in the data set ([,2]) and the frequency count of the first digit in a classic Benford's distribution ([,3])

Note

RMSD and likelihood are not formal statistic tests, so it may be evaluated only in a comparative way. To perform analysis in order to get to absolute conclusion on the veracity of the data set, Chi square p value is more trustable. For first-two and first-three digits analysis, the number of observation in the data set must be large enough to permit good Chi-square calculation. Otherwise, benford will return a warning message.

Author(s)

Vitor Hugo Moreau, Ph.D. Department of Biotechnology Federal University of Bahia, Brazil

References

Banks WP, Hill DK. 1974. The apparent magnitude of number scaled by random production. J. Exp. Psychol. 102:353–376. <http://content.apa.org/journals/xge/102/2/353>. Benford F. 1938. The Law of Anomalous Numbers. Proc. Am. Philos. Soc. 78:551–572. <http://www.jstor.org/stable/984802>. Gauvrit N, Houillon J-C, Delahaye J-P. 2017. Generalized Benford's Law as a Lie Detector. Adv. Cogn. Psychol. 13:121–127. <http://ac-psych.org/en/download-pdf/id/214>. Moreau, V. H. 2021. Inconsistencies in countries COVID-19 data revealed by Benford's law. Model Assisted Statistics and Applications 16 (2021) 73–79. <http://dx.doi.org/10.3233/MAS-210517> Nickerson RS. 2002. The production and perception of randomness. Psychol. Rev. 109:330–357. <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.109.2.330>. Sambridge M, Tkalcic H, Arroucau P. 2011. Benford's Law of First Digits: From Mathematical Curiosity to Change Detector. Asia Pacific Math. Newsl. 1:1–5.

Examples

```
#Computer generated random data do not conform to the benford law
result <- benford(seq(1,10000)+rnorm(10000,0,100), TRUE)
#Natural data set, countaining the number of daily new cases of COVID-19 in Switzerland
##conform to the Benford' Law
result <- benford(switz.data, TRUE)
##conform to second digit analysis of the Benford' Law
result <- benford(switz.data, TRUE, 2)
```

---

switz.data *Number of daily new cases of COVID-19 in Switzerland*

Description

A data set with the number of new cases of COVID-19 in Switzerland. A natural data set that conform very well to the Benford's law

Usage

```
data("switz.data")
```

Format

A numeric vector with 383 observations.

x a numeric vector

Source

Our World in Data COVID-19 project: https://ourworldindata.org/coronavirus-data

References

Roser M, Ritchie H, Ortiz-Ospina E, Hasel J. 2020. Coronavirus Pandemic (COVID-19). https://ourworldindata.org/coronavi

Examples

```
data(switz.data)
## maybe str(switz.data) ; plot(switz.data) ...
```

Index