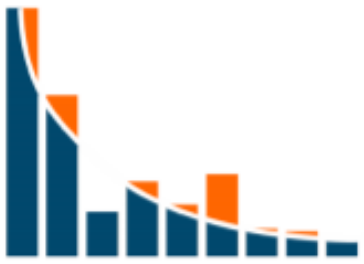


Benford's Law Conference

10-12 July 2019 - Stresa, Italy

# The TC Data 360 dataset and the Benford's Law

FRANCESCO PORRO

UNIVERSITA' DEGLI STUDI DI MILANO-BICOCCA

# The TC Data 360 dataset - introduction

## What is the TC Data 360 dataset?

"The TC data 360 is an initiative of the World Bank Group's Macroeconomics, Trade & Investment Global Practice, which helps countries achieve the Bank Group's twin goals, ending extreme poverty and boosting shared prosperity, through rapid and broad-based economic growth, centered on strong contributions from the private sector around the world"

# The TC Data 360 dataset - introduction

## What is the World Bank?

It is not a bank in the ordinary sense but a unique partnership to reduce poverty and support development. The World Bank Group comprises five institutions managed by their member countries.

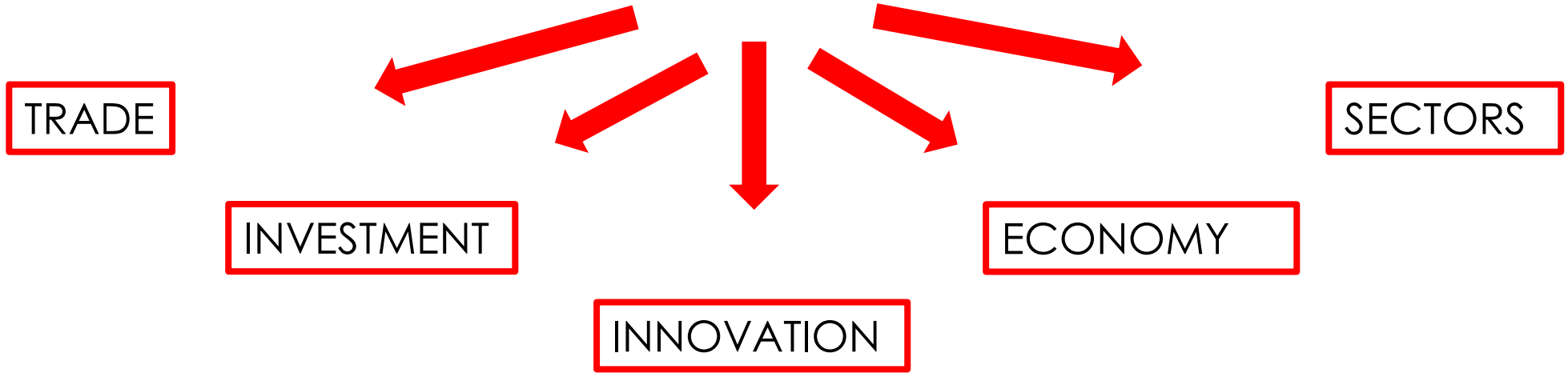
Established in 1944, the World Bank Group is headquartered in Washington, D.C.

It has more than 10,000 employees in more than 120 offices worldwide.

# The kind of the data

**Which kind of data are they?**

The TC data 360 currently features over 2400 indicators for more than 200 countries in the time range 1960-2018:



# The kind of the data

**Which kind of data are they?**

- **Trade**
  - Trade Outcomes
  - Trade Barriers
  - Trade Facilitation
  - E-trade
  - ….
- **Investment**
  - Reform Progress (Investment)
  - Risk and Policy Uncertainty
  - Entry and Investment
  - …
- **Innovation**
  - Innovation Outputs
  - Entrepreneurship
  - …
- **Economy**
  - Economic Outcomes
  - Economic and Social Context
  - Monetary and Fiscal
  - …
- **Sectors**
  - Manufacturing
  - Tourism (Sector)

# The sources of the data

## Where do the data come from?

The key sources of data on TCdata360 include:







# The main goal

**Are there any variables following the Benford's Law?**

For same suitable variables all the available values in the dataset have been considered and a statistical test on the distribution of the first significant digit has been performed. The distribution target is:

Benford's Law:

$$P(D_1 = d_i) = \log(1 + \frac{1}{d_i}) \quad where\ d_i \in \{1,2,\dots,9\}$$

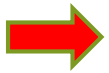
Frequency distribution:

| values | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| frequencies | 0.3010 | 0.1761 | 0.1249 | 0.0969 | 0.0792 | 0.0669 | 0.0580 | 0.0512 | 0.0458 |

# The assessment of the fitting

It is well-known that the Pearson's $X^2$ test is consistent:

It is not useful for "big" samples



only variables with a limited number of observations (<1600 obs) have been considered

# The assessment of the fitting

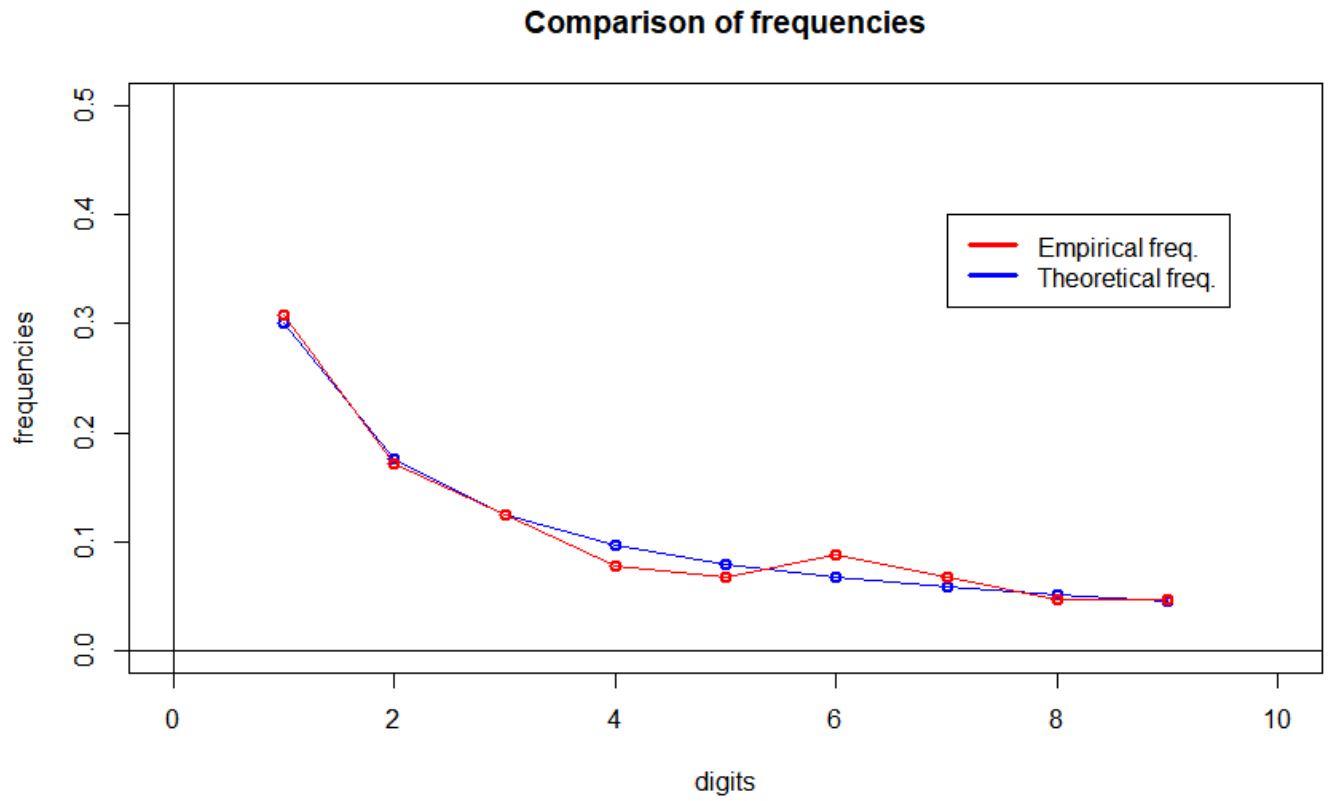
$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where:

- $O_i$ is the observed frequency for the digit *i*
- $E_i$ is the theoretical frequency of the digit *i*, according to the Benford's Law

| $\alpha$ | 0.25 | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 | 0.001 |
|---|---|---|---|---|---|---|---|
| Critical values: $\chi^2_{8;1-\alpha}$ | 10.22 | 13.36 | 15.51 | 17.53 | 20.09 | 21.96 | 26.12 |

# Investment in water and sanitation with private participation (current US$)




Time range: 1987-2018

Total number of observations:

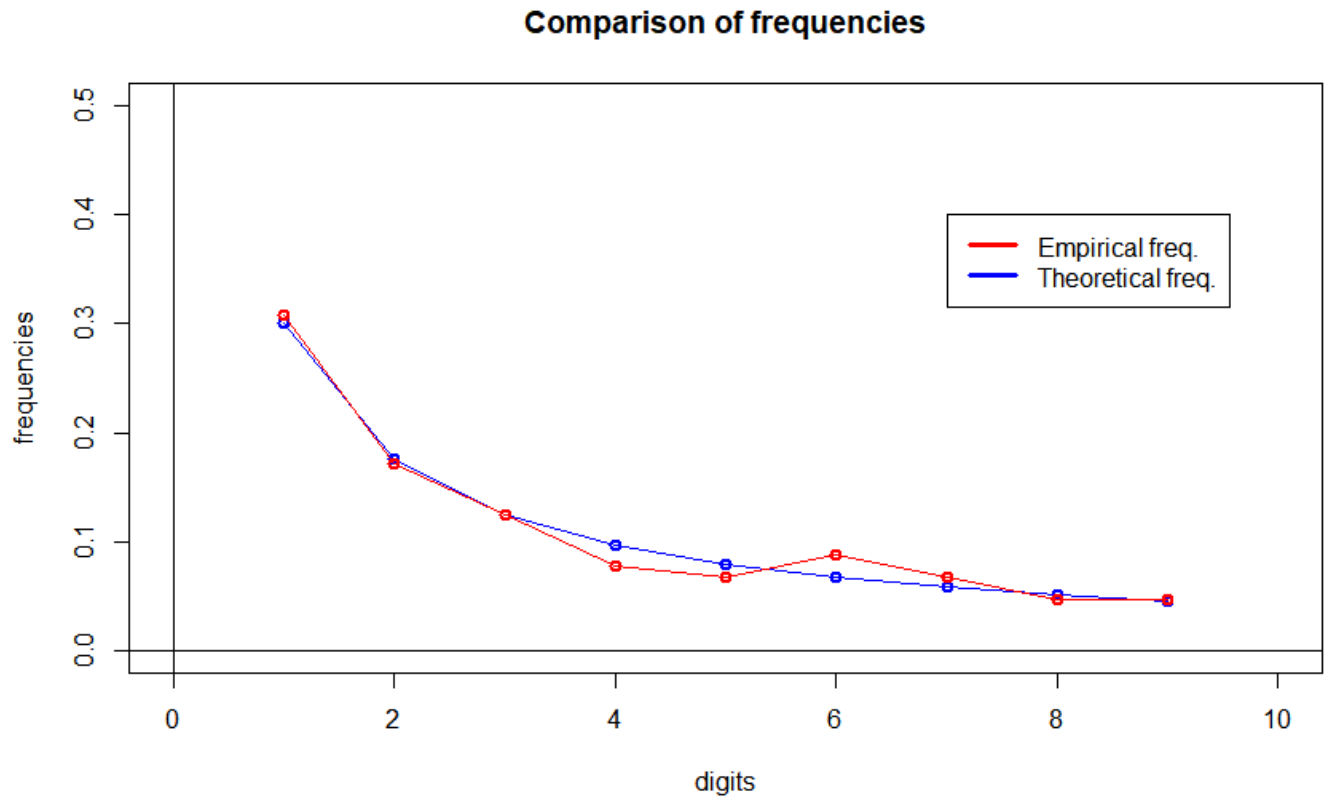
N=192

Value of $X^2$ = 2.786617

P-value= $P(\chi_8^2 > 2.786617)$ = 0.9470

**BENFORD'S LAW?**

# Investment in water and sanitation with private participation (current US$)




Time range: 1987-2018

Total number of observations:

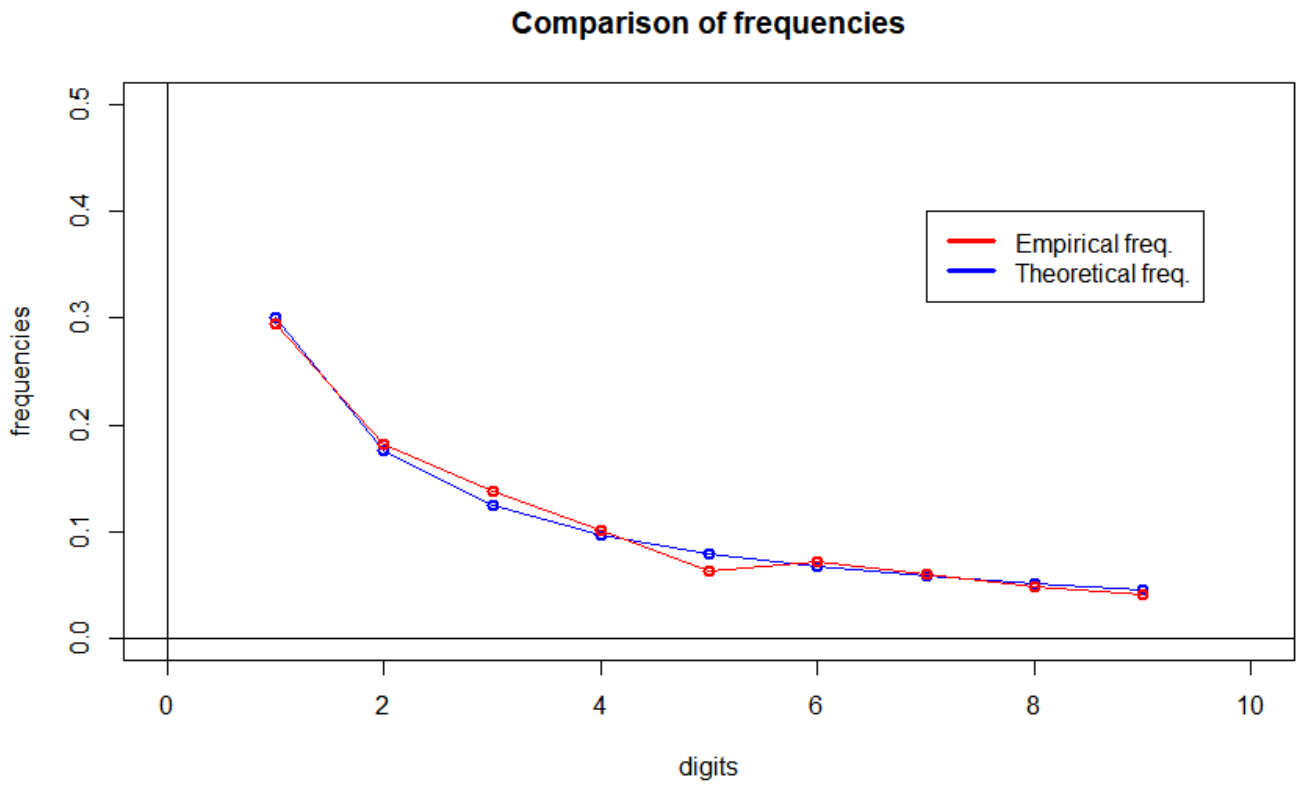
N=192

Value of $X^2$ = 2.786617

P-value= $P(\chi_8^2 > 2.786617)$ = 0.9470

**BENFORD'S LAW? YES**

# New businesses registered (number)




Time range: 2006-2015

Total number of observations:

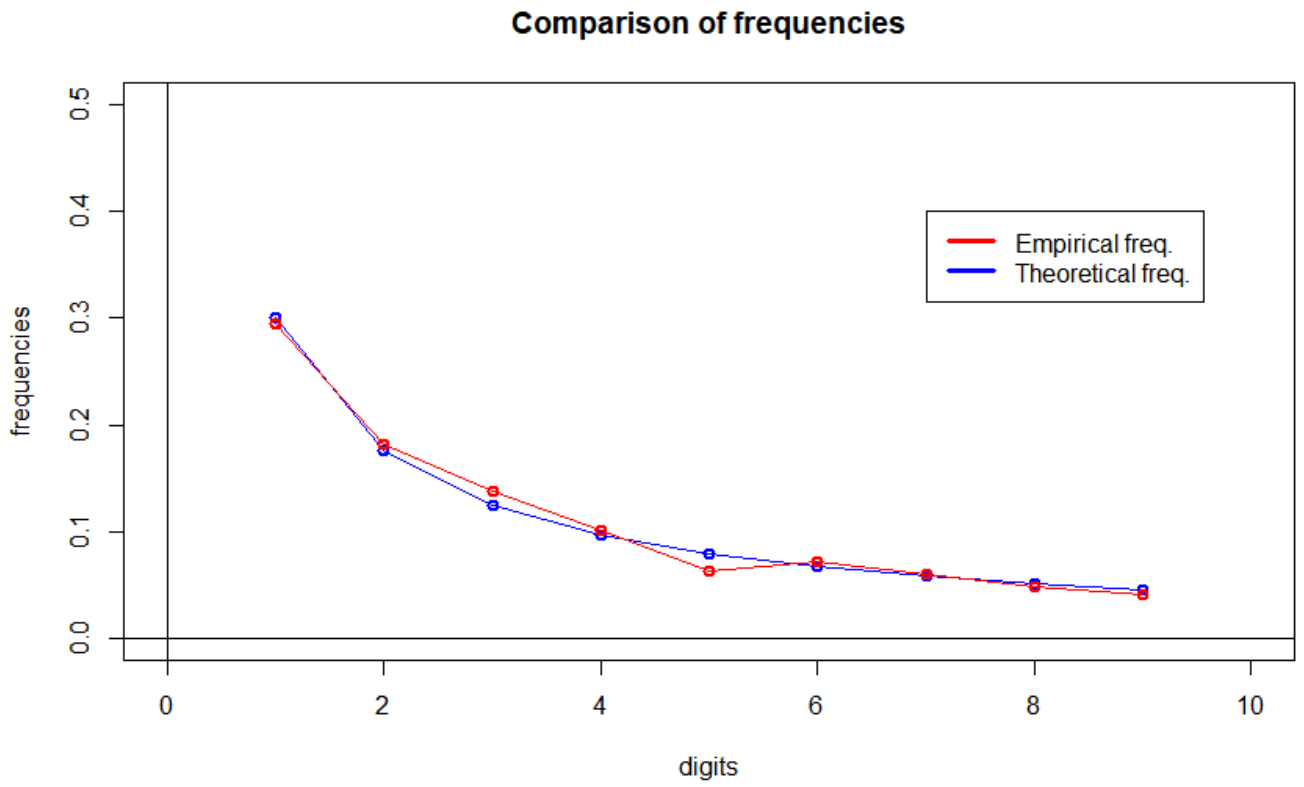
N=1293

Value of $X^2$ = 7.980378

P-value= $P(\chi_8^2 > 7.980378)$ = 0.4353

**BENFORD'S LAW?**

# New businesses registered (number)




Time range: 2006-2015

Total number of observations:

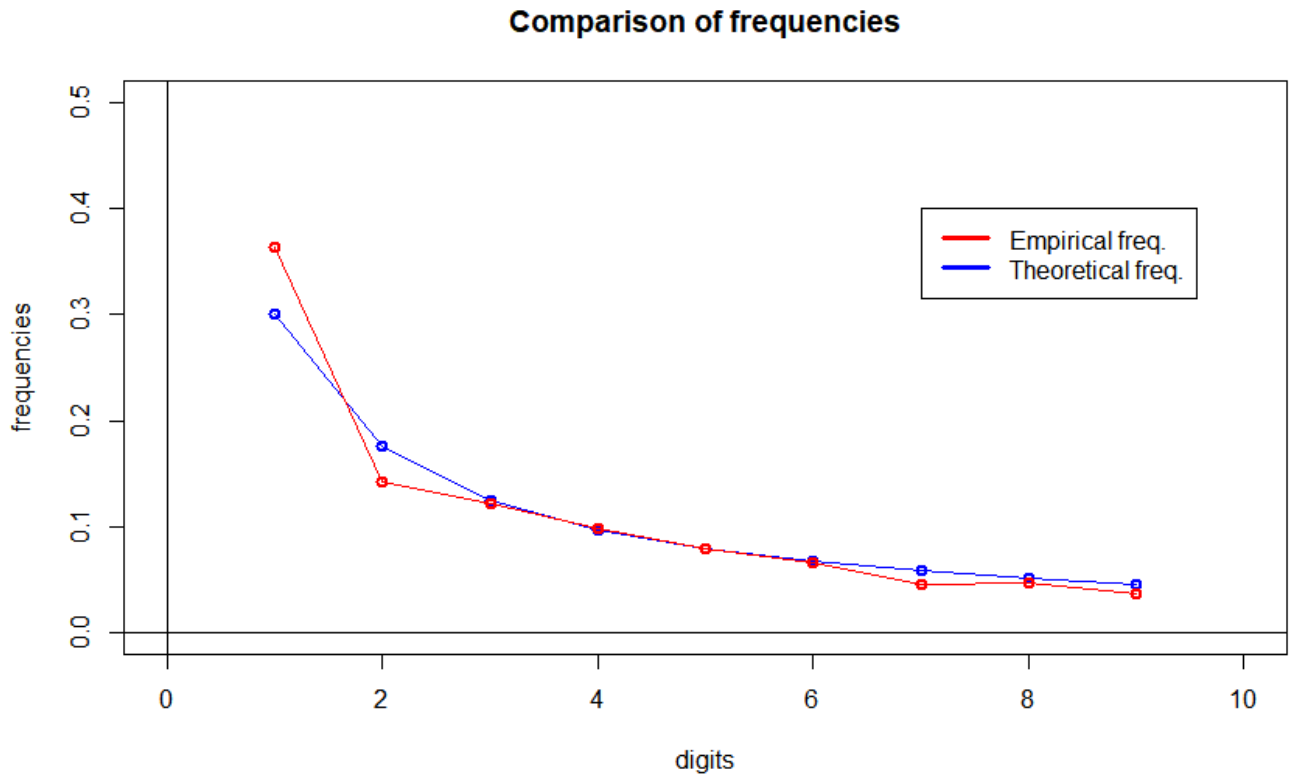
N=1293

Value of $X^2$ = 7.980378

P-value= $P(\chi_8^2 > 7.980378)$ = 0.4353

**BENFORD'S LAW? YES**

# Investment in energy with private participation (current US$)




Time range: 2003-2010

Total number of observations:

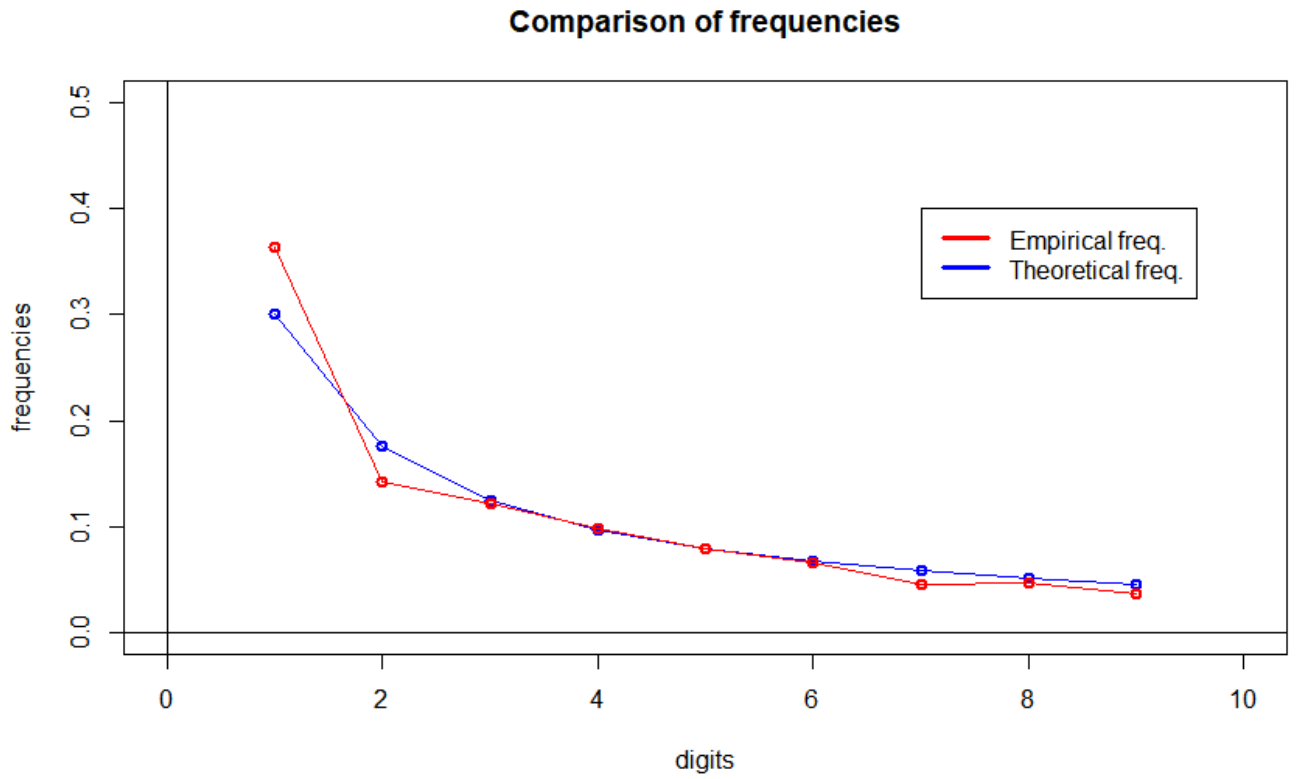
N=728

Value of $X^2$ = 21.41651

P-value= $P(\chi_8^2 > 21.41651)$ = 0.0613

**BENFORD'S LAW?**

# Investment in energy with private participation (current US$)




Time range: 2003-2010

Total number of observations:

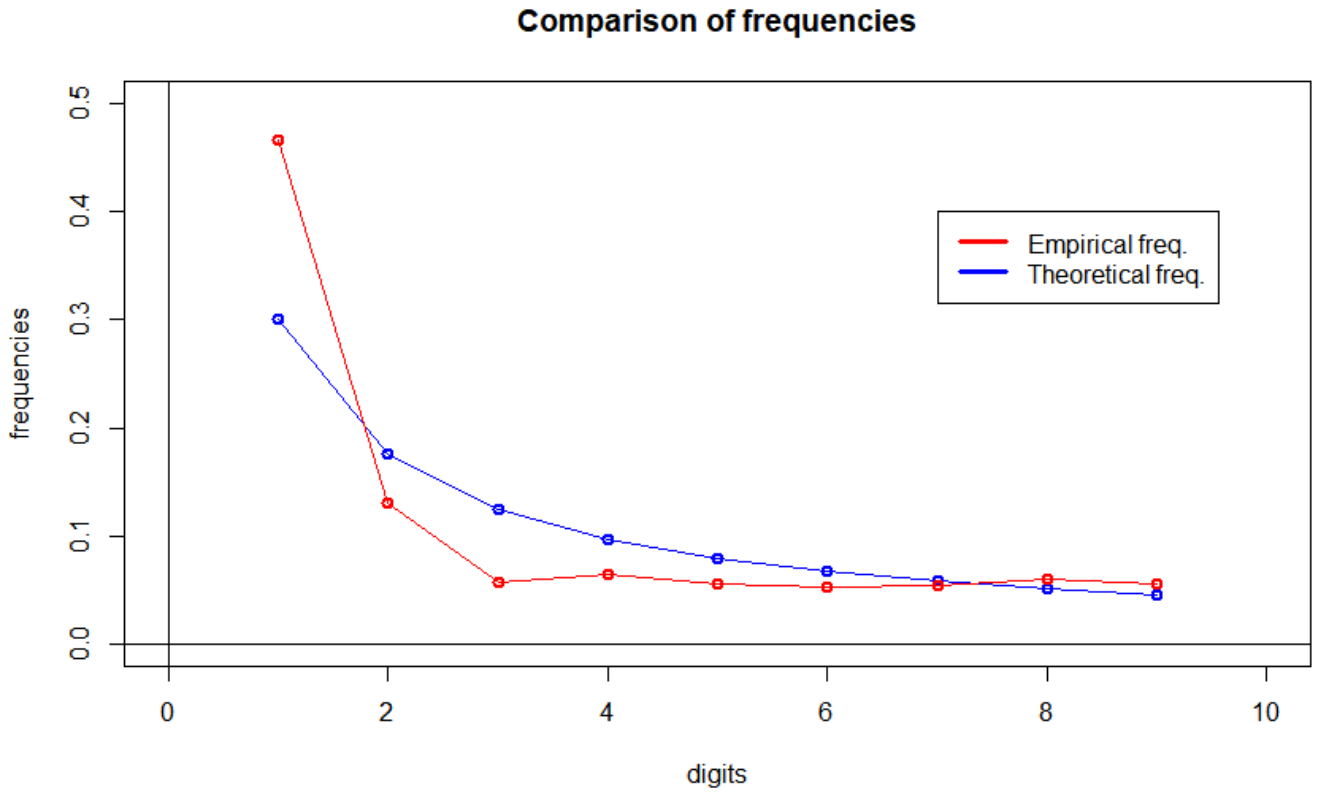
N=728

Value of $X^2$ = 21.41651

P-value= $P(\chi_8^2 > 21.41651)$ = 0.0613

**BENFORD'S LAW? NO**

# Gross national expenditure (current US$)




Time range: 1961-2017

Total number of observations:

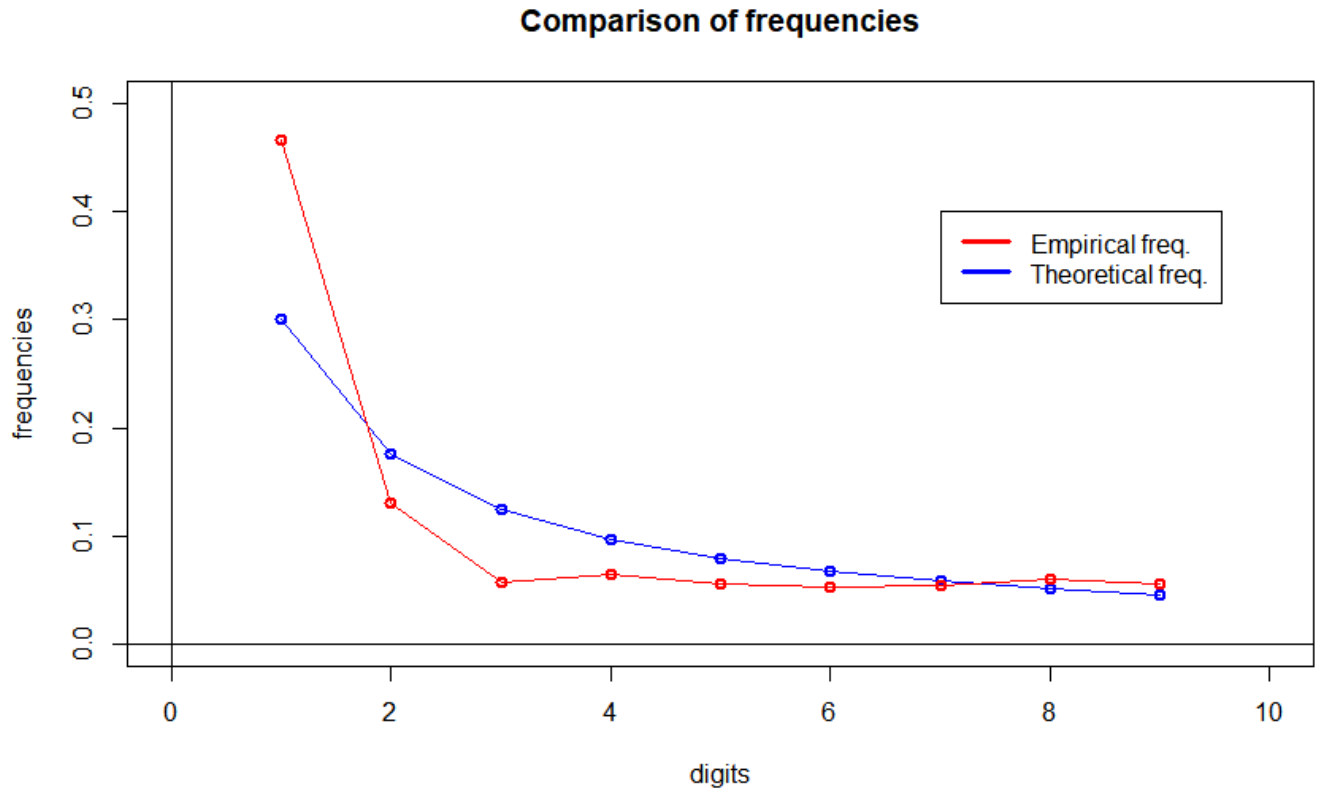
N=1564

Value of $X^2$ = 256.3665

P-value= $P(\chi_8^2 > 256.3665) = 0$

**BENFORD'S LAW?**

# Gross national expenditure (current US$)




Time range: 1961-2017

Total number of observations:

N=1564

Value of $X^2$ = 256.3665

P-value= $P(\chi_8^2 > 256.3665) = 0$

**BENFORD'S LAW? NO**

# Summary

- The TC data 360 is a very huge dataset, provided by worthy and prestigious international institutions
- There are more than 2400 indicators for more than 200 countries all over the world
- The variables range in the time interval 1960-2018 (the availability varies)
- The variables are divided into five main areas: Trade, Investment, Innovation, Economy, and Sectors
- The variables are very different: some of them are clearly not related to the Benford's law, some are clearly related to it, and the all the other ones are suitable for an investigation…
- The number of the observations is an important issue to be taken under control!

# References

Frank Benford. The law of anomalous numbers. Proceedings of the American Philosophical Society, 78(4):551–572, (1938).

Arno Berger and Theodore P. Hill. A basic theory of Benford's law. Probability Surveys, 8:1–126, (2011).

Arno Berger and Theodore P. Hill. An introduction to Benford's law. Princeton University Press, (2015).

Simon Newcomb. Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematic, 4(1):39–40, (1881)

Nigrini, Mark J. A taxpayer compliance application of Benford's Law, The Journal of American Taxation Association, Sarasota 18(1): 71-92 (1996)

Nigrini, Mark J, Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection, John Wiley & Sons, (2012)

Jessica Riccioni, Roy Cerqueti. Regular paths in financial markets: investigating the Benford's law. Chaos, Solitons and Fractals, 107 :186–194, (2018)