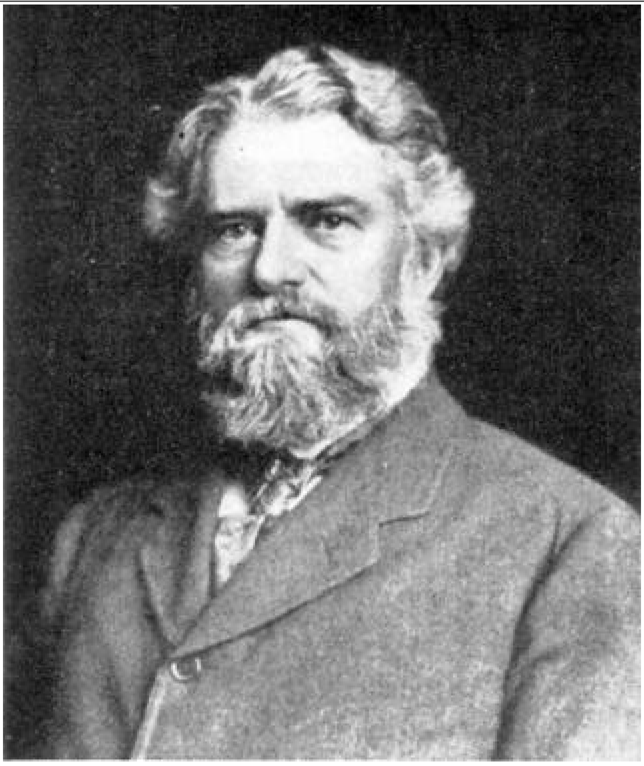
t

ESTATÍSTICA

Statisticum

DIVINUS

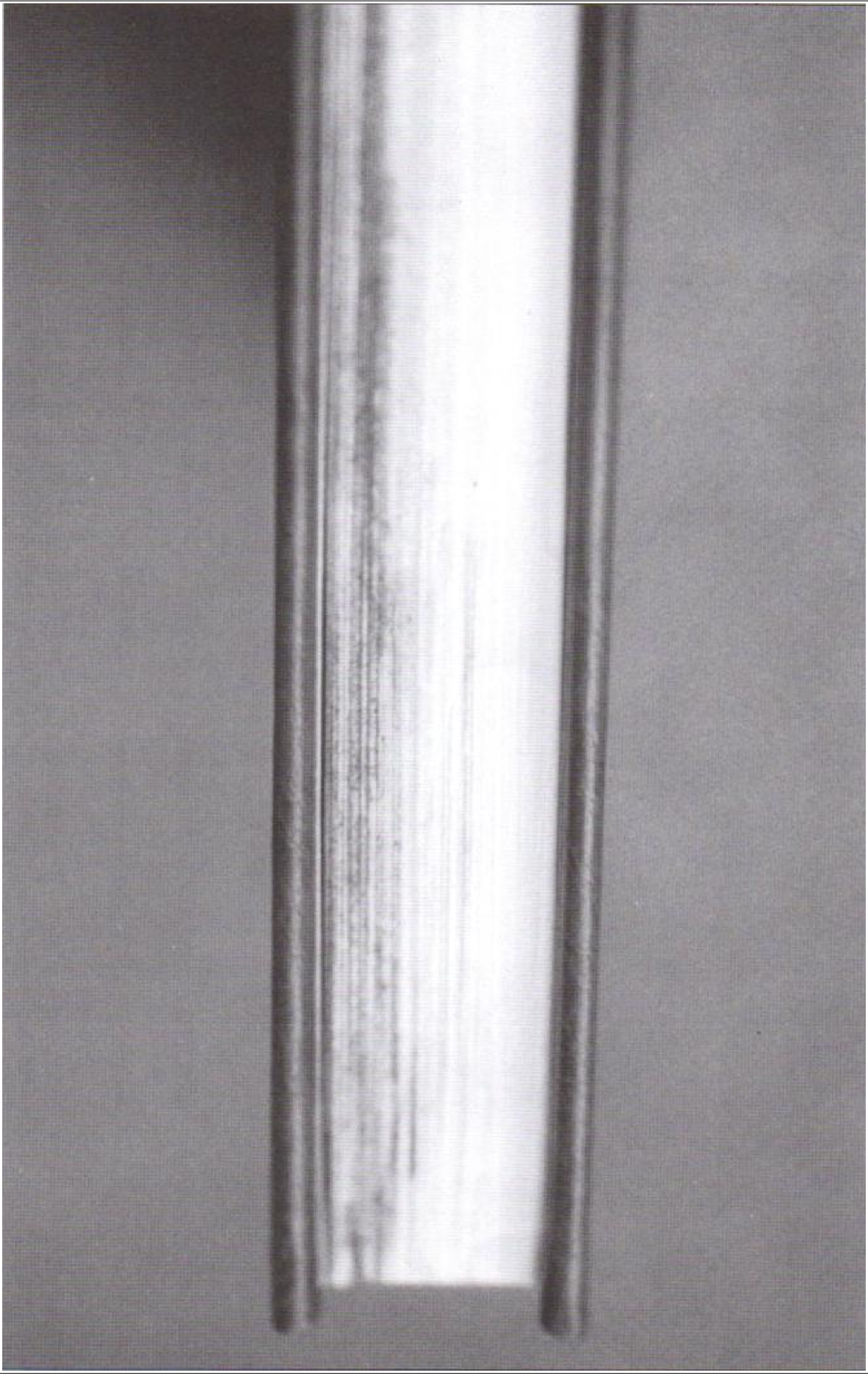
PROBO



Simon NewcomB

Frank Benford

t



Primeiro Dígito

Fenômeno Estatístico Modelado por Função

A Lei de Benford/Newcomb, também conhecida como a "Lei do Primeiro Dígito", foi descoberta pelo astrônomo americano Simon Newcomb em 1881, ao notar que as páginas dos livros de logaritmos que traziam números iniciados pelo algarismo 1 ficavam mais sujas e danificadas do que as páginas correspondentes aos números 2, e assim por diante até o algarismo 9, cujas páginas eram as mais limpas e novas, sendo o desgaste do livro proporcional à freqüência de uso. Newcomb concluiu que os cientistas que compartilhavam o livro trabalhariam com dados que refletiam a distribuição dos algarismos, sendo o 1 mais frequente do que o dois, e assim por diante.

Frank Benford, físico empregado pela General Electric, notou este mesmo fato em 1938, ao examinar as tabelas de logaritmos do laboratório de pesquisas da GE.

Nenhum dos dois foi capaz de provar a lei. Isso ocorreu apenas no ano de 1995 no Instituto de Tecnologia da Geórgia pelo matemático Ted Hill.

Segundo a Lei do Primeiro Dígito, em qualquer conjunto de dados coletados do mundo real (a maioria mas não todos), os nove algarismos não aparecem com a mesma frequência: o número 1 ocorrerá como o primeiro dígito em cerca de 30% dos números, o número 2 ocorrerá cerca de 17,5% e assim por diante, descrevendo uma função logarítmica. A Lei do Primeiro Dígito tem sido utilizada como ferramenta na detecção de fraudes em balancetes financeiros, pesquisas de opinião e relatórios similares, apontando com certa exatidão os casos de fraude no emaranhado de relatórios numéricos pesquisados.

Contudo, a Lei de Benford não se aplica em todos os casos, dependendo das características dos dados levantados.

Benford, Frank

The Law of Anomalous Numbers

Proc. American Philosophical Society, Vol. 78,
#4, March 1938, pp.551-572

#887 rec'd 500 copies 5/20/38

Benford, Frank

The Probable Accuracy of the General Physical
Constants

Physical Review, Vol. 63, Nos. 5 and 6,
p.212, March 1 and 15, 1943

No. 1174 rec'd 200 copies 6/14/43

t

1

Frank Benford

Room 503, Building 37, Phone 773

Aug 18, 1928

—

Vacation 8-18-28 to 9-3-28

—

1. Color sensitivity of photoelectric cells by filters and photometer.

Requested by Dr S.D. Sept 4, 1928.

Standard Lamp 2620

V-116 2472°K.

Working Std. 231 amb. 140.0 Dist.

56.9 "

$$90.6 = 60.6\left(\frac{83.1}{68.0}\right)^2 \text{ cd.}$$

Galv. 124432 .9640 mlgs = 10^{-10}

Shunt 2000 Ω

Coblentz cell

—

140.0 240 min

Bare cell Zero = -10 + 16.3 = 26.3-2

17.5 275-3

11.5-1

H_2O - -10

10.1-2

9.3-3

6.0-1

H_2O + Red - MRS

5.5-2

5.5-3

t

GENERAL NEWS BUREAU (JS)
GENERAL ELECTRIC
SCHENECTADY 5, N. Y.

For release after 7 P.M.
Friday, July 30, 1948

SCHENECTADY, N.Y., July 31--Frank Benford, member of the Research Laboratory of the General Electric Company, retires today after 38 years of service, 20 of which have been in the Research Laboratory, the company announced today.

Noted as an expert on optics, he is the author of 109 published papers on optics and mathematics and has been granted 20 patents on optical devices.

A native of Johnstown, Pa., he was graduated from the University of Michigan with a bachelor of science degree in electrical engineering in 1910.

He joined General Electric in Schenectady the same year as a student on the G-E Test Course, and then became a physicist in the Illuminating Engineering Laboratory. During World War I he worked on the theory and design of searchlights. In the early 20's he published a series of 22 papers on this subject. An integrating hemisphere to measure light distribution in a searchlight beam was another Benford development which is still in use.

(more)

THE LAW OF ANOMALOUS NUMBERS

FRANK BENFORD

Physicist, Research Laboratory, General Electric Company,
Schenectady, New York

(Introduced by Irving Langmuir)

(*Read April 22, 1937*)

Abstract

It has been observed that the first pages of a table of common logarithms show more wear than do the last pages, indicating that more used numbers begin with the digit 1 than with the digit 9. A compilation of some 20,000 first digits taken from widely divergent sources shows that there is a logarithmic distribution of first digits when the numbers are composed of four or more digits. An analysis of the numbers from different sources shows that the numbers taken from unrelated subjects, such as a group of newspaper items, show a much better agreement with a logarithmic distribution than do numbers from mathematical tabulations or other formal data. There is here the peculiar fact that numbers that individually are without relationship are, when considered in large groups, in good agreement with a distribution law—hence the name " Anomalous Numbers."

A further analysis of the data shows a strong tendency for bodies of numerical data to fall into geometric series. If the series is made up of numbers containing three or more digits the first digits form a logarithmic series. If the numbers contain only single digits the geometric relation still holds but the simple logarithmic relation no longer applies.

An equation is given showing the frequencies of first digits in the different orders of numbers 1 to 10, 10 to 100, etc.

The equation also gives the frequency of digits in the second, third $\cdots$ place of a multi-digit number, and it is shown that the same law applies to reciprocals.

There are many instances showing that the geometric series, or the logarithmic law, has long been recognized as a common phenomenon in factual literature and in the ordinary affairs of life. The wire gauge and drill gauge of the mechanic, the magnitude scale of the astronomer and the sensory response curves of the psychologist are all particular examples of a relationship that seems to extend to all human affairs. The Law of Anomalous Numbers is thus a general probability law of widespread application.

Part I: Statistical Derivation of the Law

It has been observed that the pages of a much used table of common logarithms show evidences of a selective use of the natural numbers. The pages containing the logarithms of the low numbers 1 and 2 are apt to be more stained and frayed by use than those of the higher numbers 8 and 9. Of

PROCEEDINGS OF THE AMERICAN PHILOSOPHICAL SOCIETY,
VOL. 78, NO. 4, MARCH, 1938

Reprint *Printed in U. S. A.*

Histograma de Benford

d	*P*(*d*)	Relative size of *P*(*d*)
1	30.1%	
2	17.6%	
3	12.5%	
4	9.7%	
5	7.9%	
6	6.7%	
7	5.8%	
8	5.1%	
9	4.6%	

t

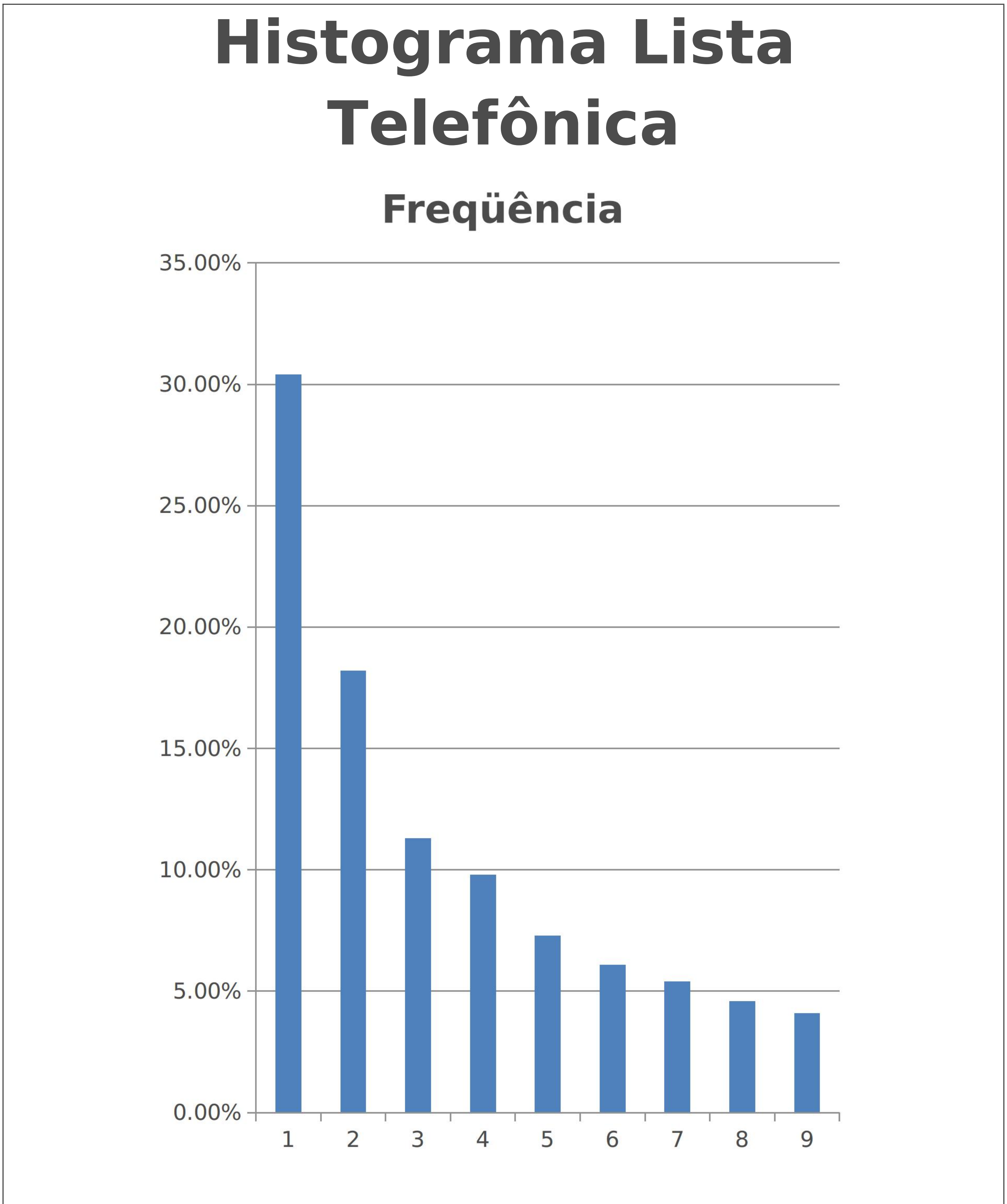
2.0 Aplicação da lei de Benford

2.1 Lista Telefônica de Ponta Grossa

Para aproximadamente 1500 (1463) assinantes da lista telefônica, foram encontrados as seguintes ocorrências

1,2,3,4,5,6,7,8 e 9

$P(1) = \frac{456}{1500}$ P(1) = 0,304 P(1) = 30,4%	$P(2) = \frac{273}{1500}$ P(2) = 0,182 P(2) = 18,2%	$P(3) = \frac{170}{1500}$ P(3) = 0,113 P(3) = 11,3%
$P(4) = \frac{148}{1500}$ P(4) = 0,098 P(4) = 9,8%	$P(5) = \frac{110}{1500}$ P(5) = 0,073 P(5) = 7,3%	$P(6) = \frac{92}{1500}$ P(6) = 0,061 P(6) = 6,1%
$P(7) = \frac{82}{1500}$ P(7) = 0,054 P(7) = 5,4%	$P(8) = \frac{70}{1500}$ P(8) = 0,046 P(8) = 4,6%	$P(9) = \frac{62}{1500}$ P(9) = 0,041 P(9) = 4,1%


Histograma Lista Telefônica
Freqüência
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
1
2
3
4
5
6
7
8
9

PROBO

2.2 Notas Fiscais

1,2,3,4,5,6,7,8 e 9

DADOS NF
P(1) = 12,0%
P(2) = 4,0 %
P(3) = 5,6 %
P(4) = 7,4 %
P(5) = 17,7%
P(6) = 2,0%
P(7) = 16,6%
P(8) = 13,3%
P(9) = 21,3%

HISTOGRAMA DE NF