



# The Applicability of Benford's Law

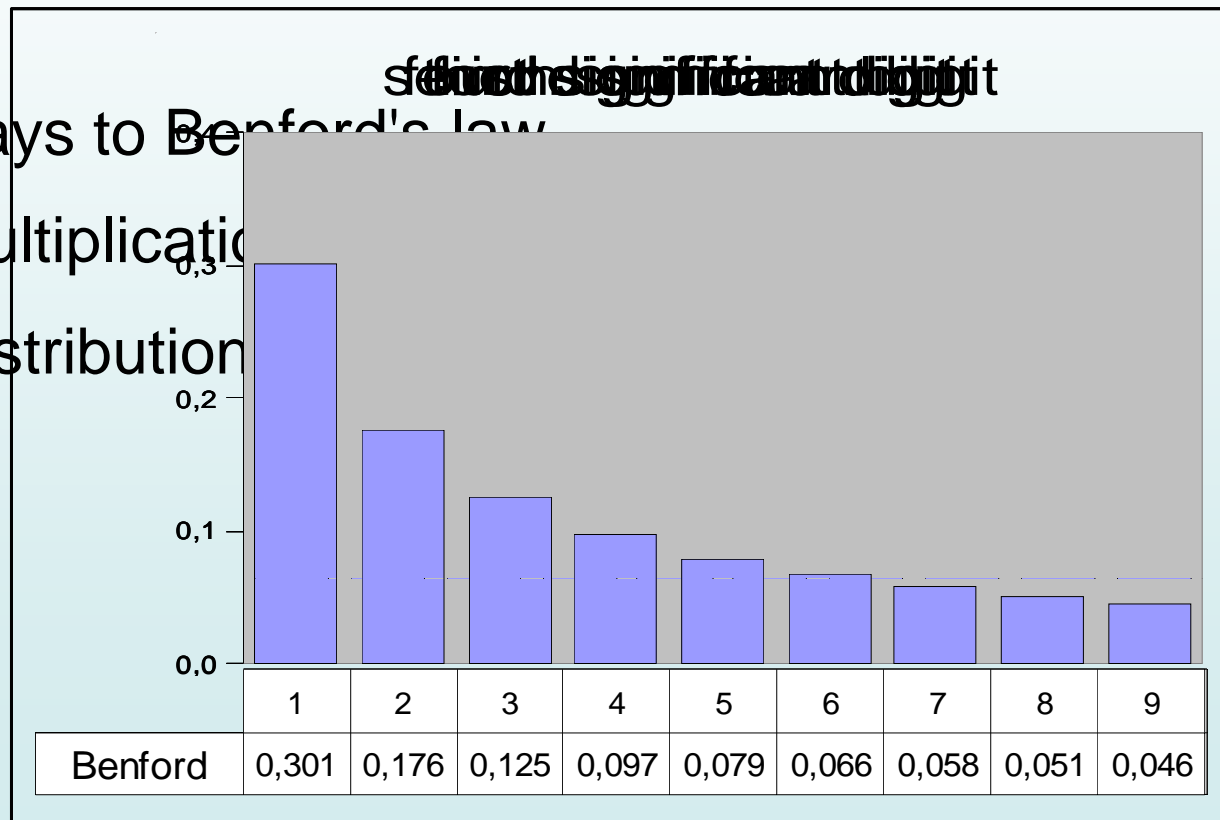
Fraud detection in the social sciences

# Benford distribution

$$P(D_1=d_1 \dots D_k=d_k) = \log_{10} \left[ 1 + \left( \sum_{i=1}^k d_i \cdot 10^{k-i} \right)^{-1} \right]$$

Two ways to Benford's law

- Multiplication
- Distribution



# Benford distributed data



- Front pages of newspapers
- Farmers in Paraguay
- The surface of rivers
- House numbers
- Company balance sheets

# Detecting fraud



## Approach:

- Distribution of digits with specific values
- Uncovering of fraudulent data

# First step: What is Benford distributed



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

	N	1.digit	2.digit	3.digit	4.digit
Not standardised regressions	2180	X	X	X	X
Logistic regressions	2251		X	X	X
T-values	1325			X	X
Cox regressions	506			X	X
Pseudo r <sup>2</sup>	239		X	X	X
Chi-square values	188		X	X	X
R <sup>2</sup>	131		X		(X)

Datasource: Kölner Zeitschrift für Soziologie und Sozialforschung Feb.1985 to Mar.2007  
with support by the chair of professor Braun

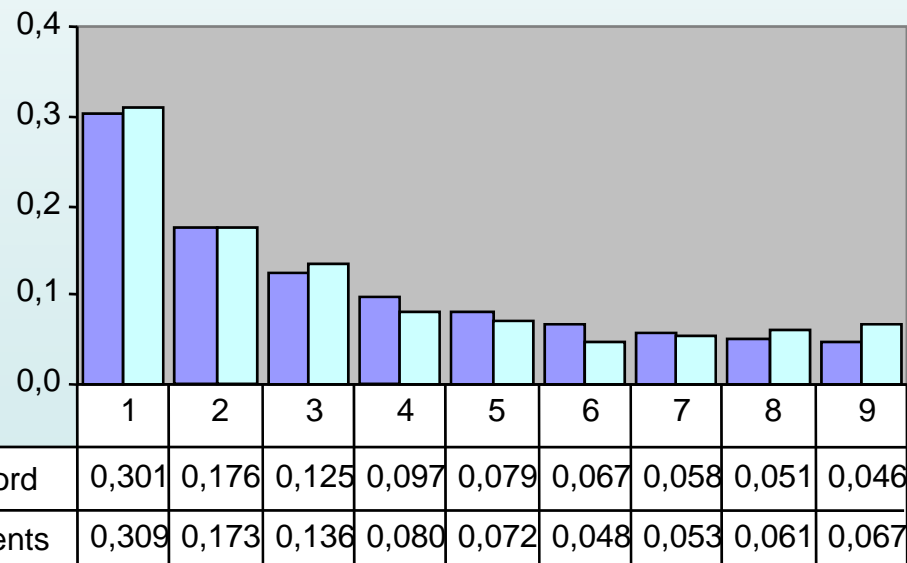
# Research survey by the chair of prof. Braun



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

Hypothesis: 'The higher the education a person has, the fewer cigarettes he consumes per day'

Fraudulent regressions coefficients  
Distribution of *first digit* (n= 4621)



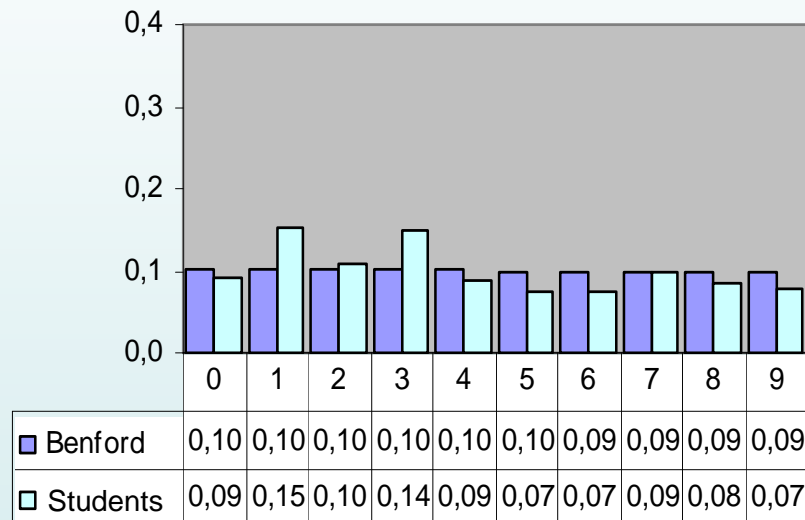
1. digit:  $H_0$  rejected  
( $X^2=103.39, df = 8, p = 0.000$ )

2. digit:  $H_0$  rejected  
( $X^2=122.59, df = 9, p = 0.000$ )

# Research survey: 3. and 4. digit

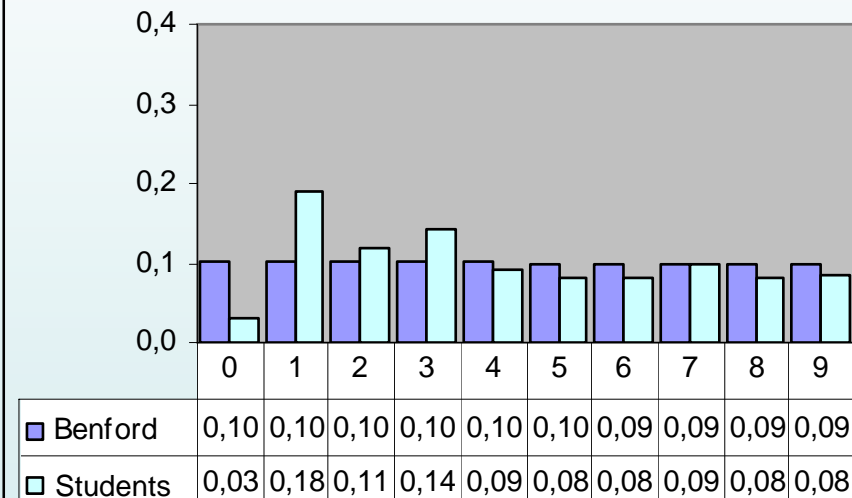


Fraudulent regressions coefficients  
Distribution of *third digit* (n= 4541)



$H_0$  rejected  
( $X^2 = 304.89$ ,  $df=9$ ,  $p= 0.000$ )

Fraudulent regressions coefficients  
Distribution of *fourth digit* (n= 4378)



$H_0$  rejected  
( $X^2 = 622.20$ ,  $df=9$ ,  $p= 0.000$ )

# Research survey: Individual data



## Deviations from the Benford distribution

47 persons	1. digit	2. digit	3. digit	4. digit
absolute	35	40	42	41
percentage	0.744	0.851	0.893	0.872



## Second step: detecting fraud



### Approach:

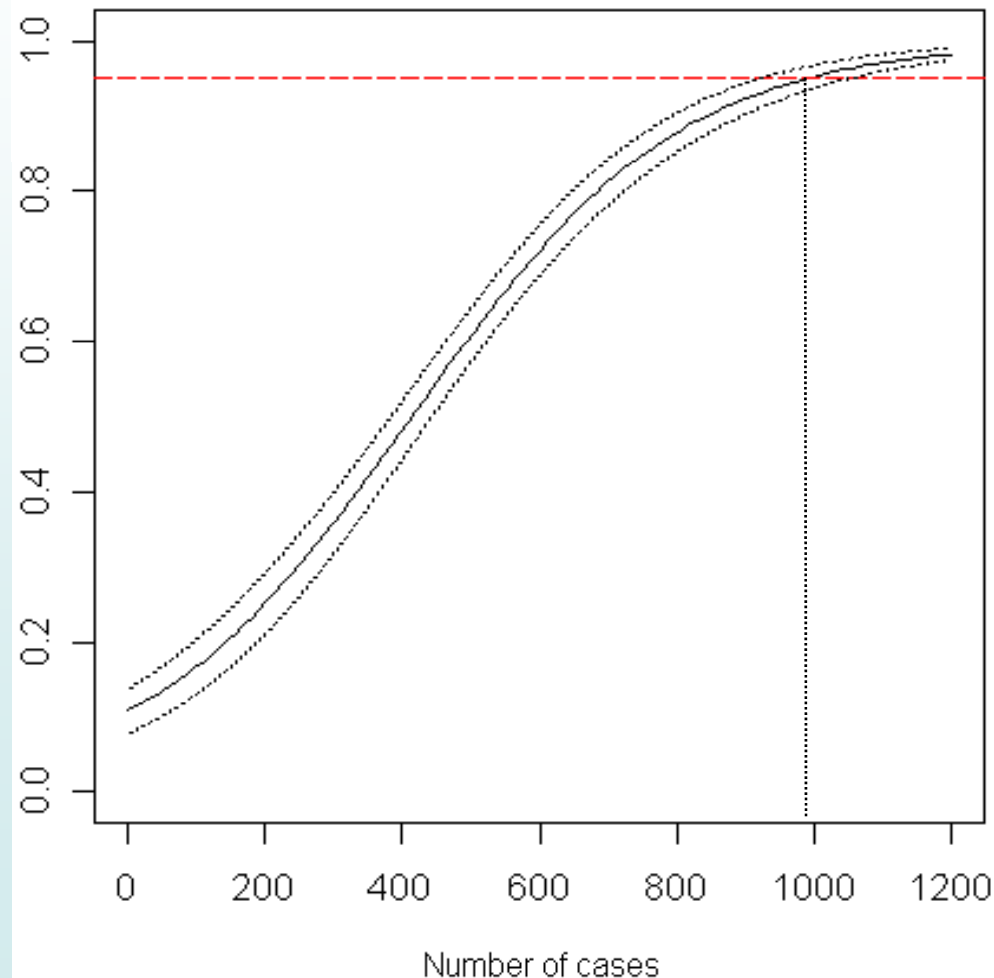
- At what point is fraud recognisable?

### Method:

- Random selection of fraudulent digits
- Goodness of fit test used on a Benford distribution
- Logistic regression

## Second step: results

statistische Signifikanz

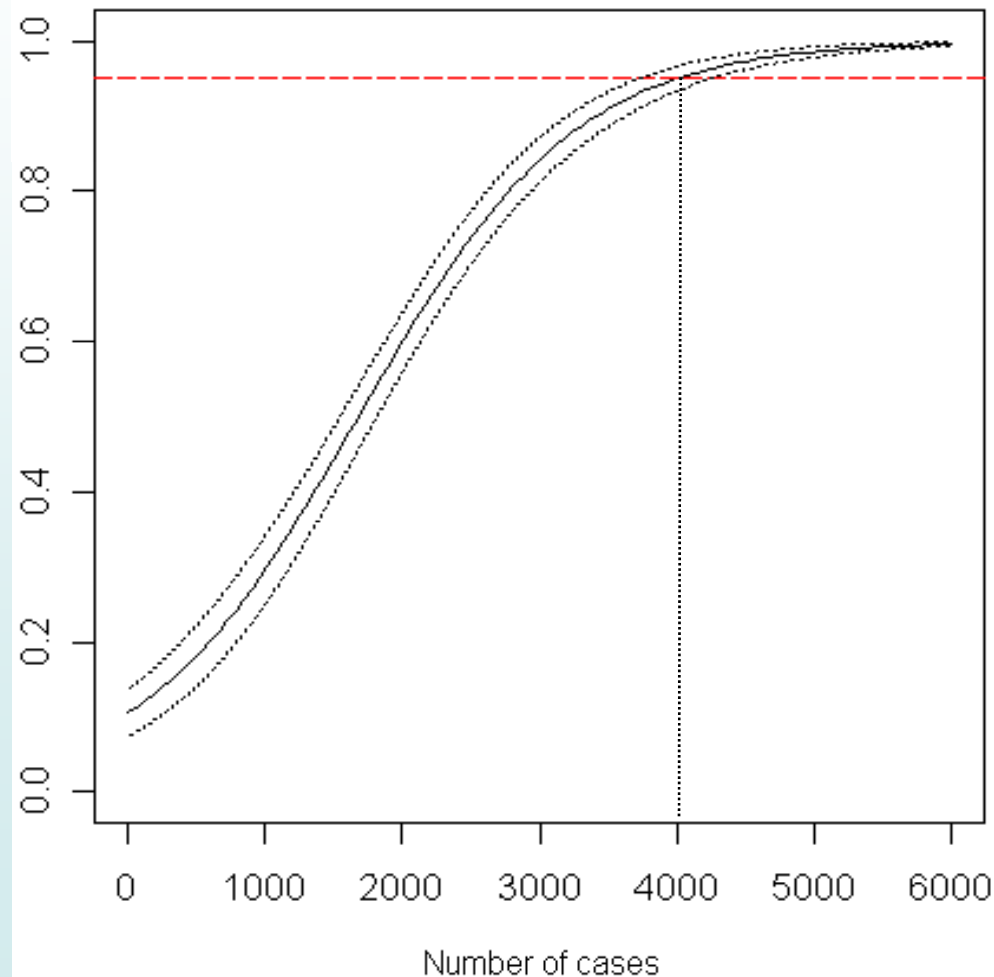


Mean number of cases in order to reject  $H_0$  with a probability of 95 %:

1. digit: 989 cases
2. digit: 766 cases
3. digit: 351 cases
4. digit: 138 cases

## Second step: results

first significant digit ~ 50 % fabricated data

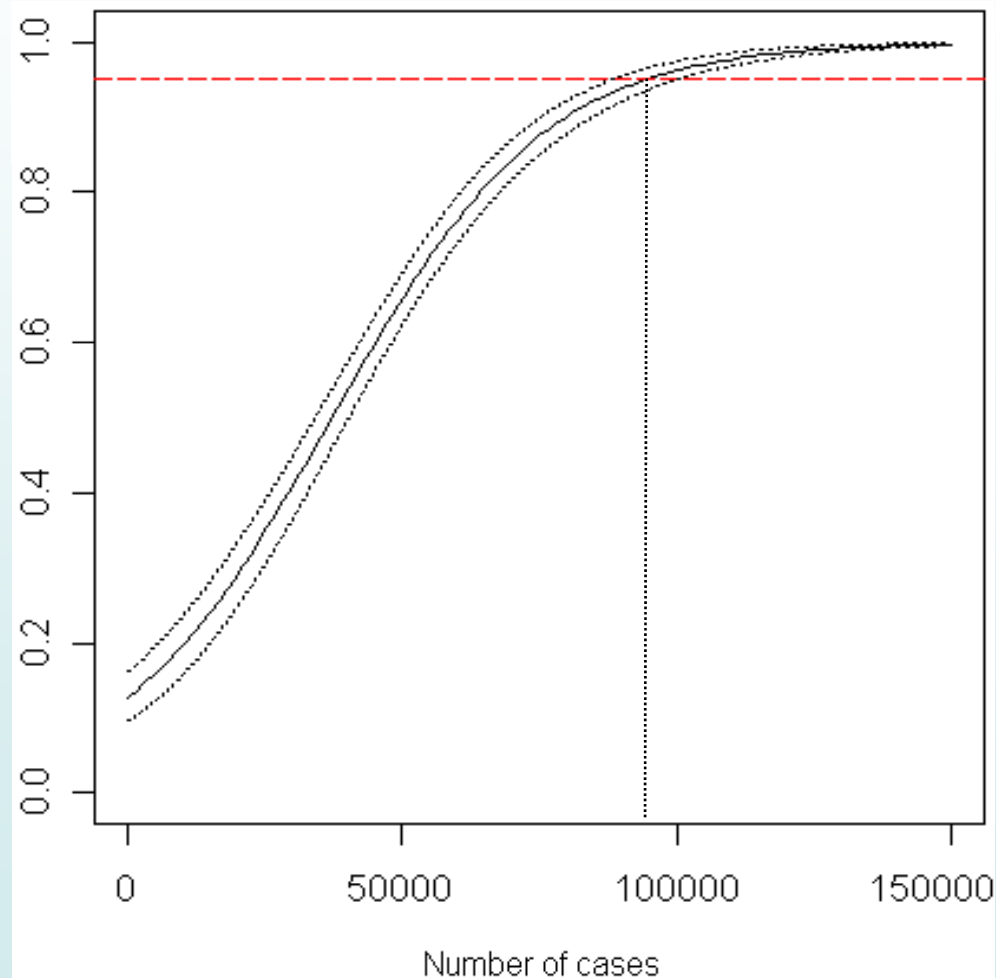


Mean number of cases in order to reject  $H_0$  with a probability of 95 %:

1. digit: 4001 cases
2. digit: 3308 cases
3. digit: 1351 cases
4. digit: 585 cases

## Second step: results

first significant digit ~ 10 % fabricated data

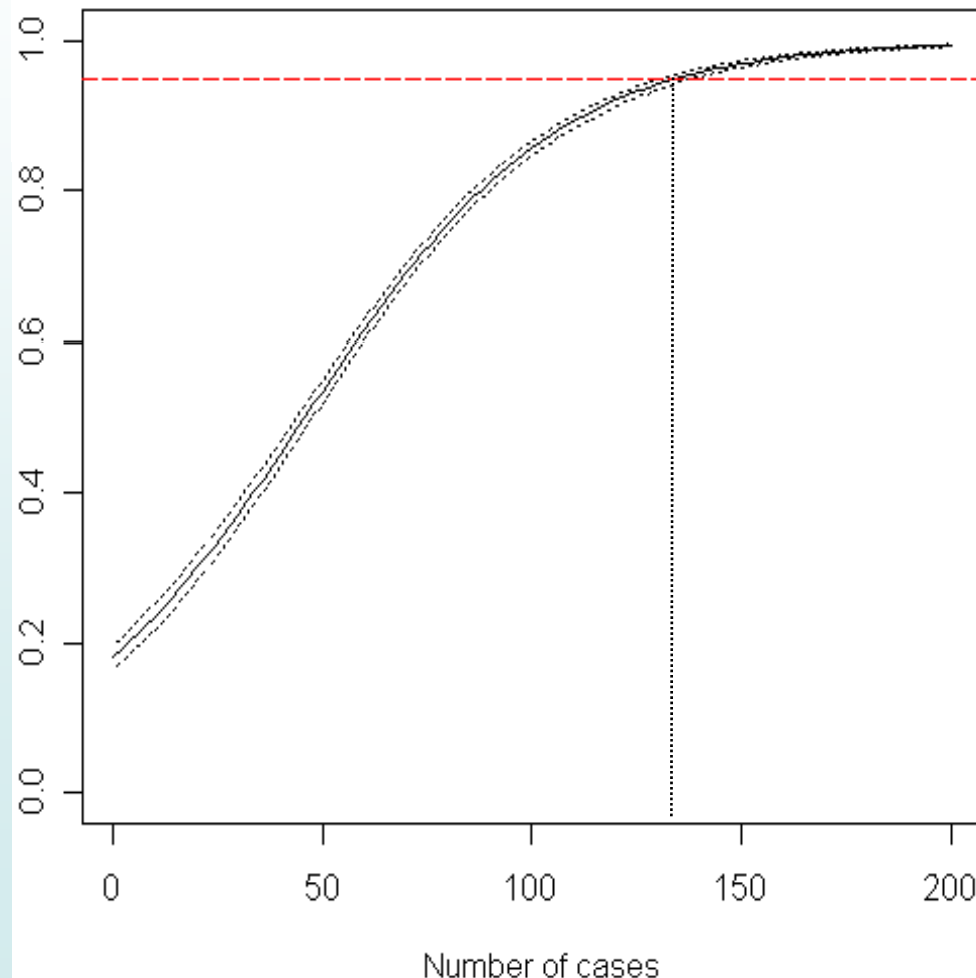


Mean number of cases in order to reject  $H_0$  with a probability of 95 %:

1. digit: 94439 cases
2. digit: 78883 cases
3. digit: 31266 cases
4. digit: 12592 cases

## Second step: Individual data

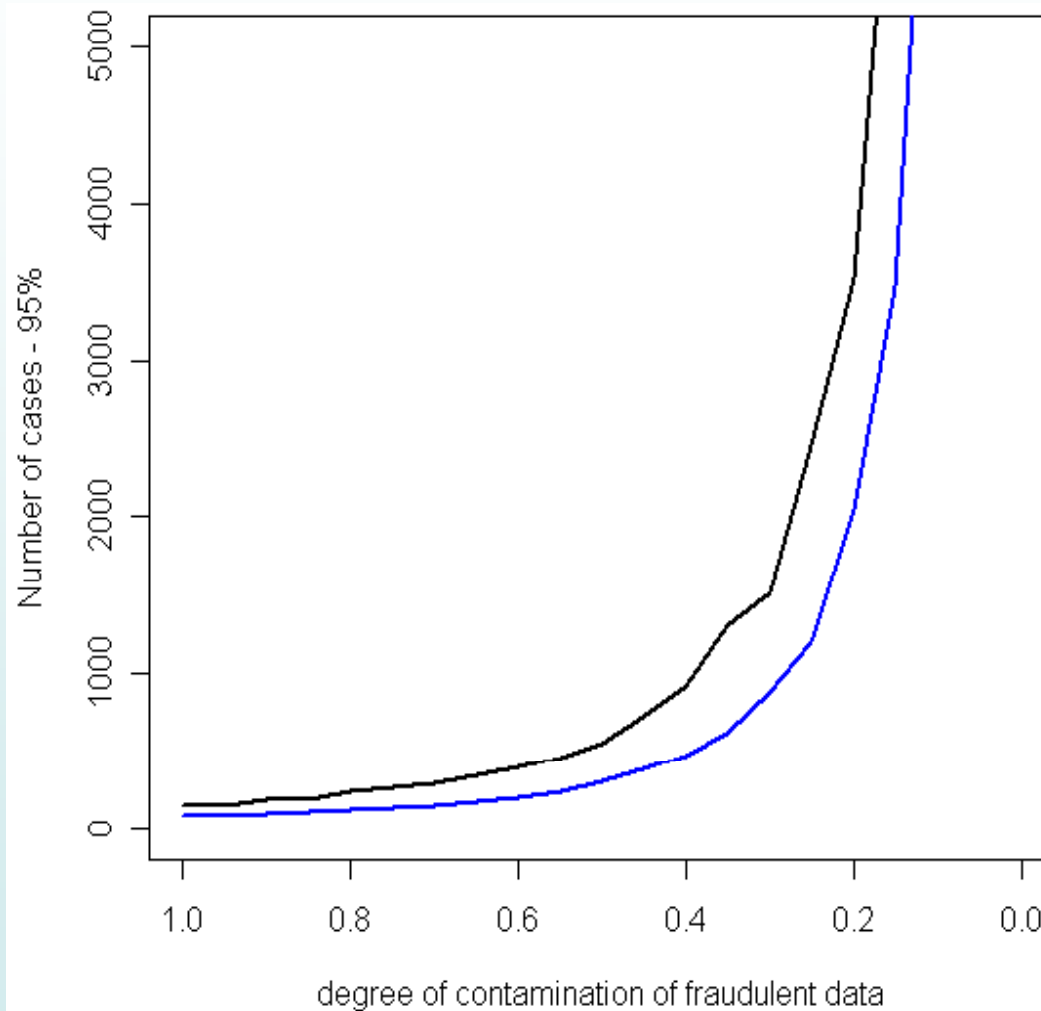
first significant digit ~ 100 % fabricated data



Mean number of cases in order to reject  $H_0$  with a probability of 95 %:

1. digit: 136 cases
2. digit: 102 cases
3. digit: 100 cases
4. digit: 69 cases

## Second step: Individual data



— first digit  
— fourth digit

## Summary of results



Fraud detection with Benford's law – proposals:

- Observation of individual data
- Observation of higher digits
- Recording of all possible metric coefficients
- Application of Goodness of fit test, which react more strongly to the sample size (i.e. chi-square g.o.f. test)
- The result is strongly dependent on the procedure of the forger

# Literature

The logo for Ludwig-Maximilians-Universität München (LMU), consisting of the letters 'LMU' in a bold, black, sans-serif font.The full name of the university, 'LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN', in a black, sans-serif font, arranged in three lines.

Benford, F. (1938): The Law of Anomalous Numbers, in: Proceedings of the American Philosophical Society, 78(4): 551-572

Diekmann, A. (2007): Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, in: Journal of Applied Statistics, 34(3): 321-329

Hill, T.P. (1996): A Statistical Derivation of the Significant-Digit Law, in: Statistical Science, 10(4) 354-363

Hammering, R.W. (1970): On the Distribution of Numbers, in: Bell System Technical Journal, 49(8): 1609-1625

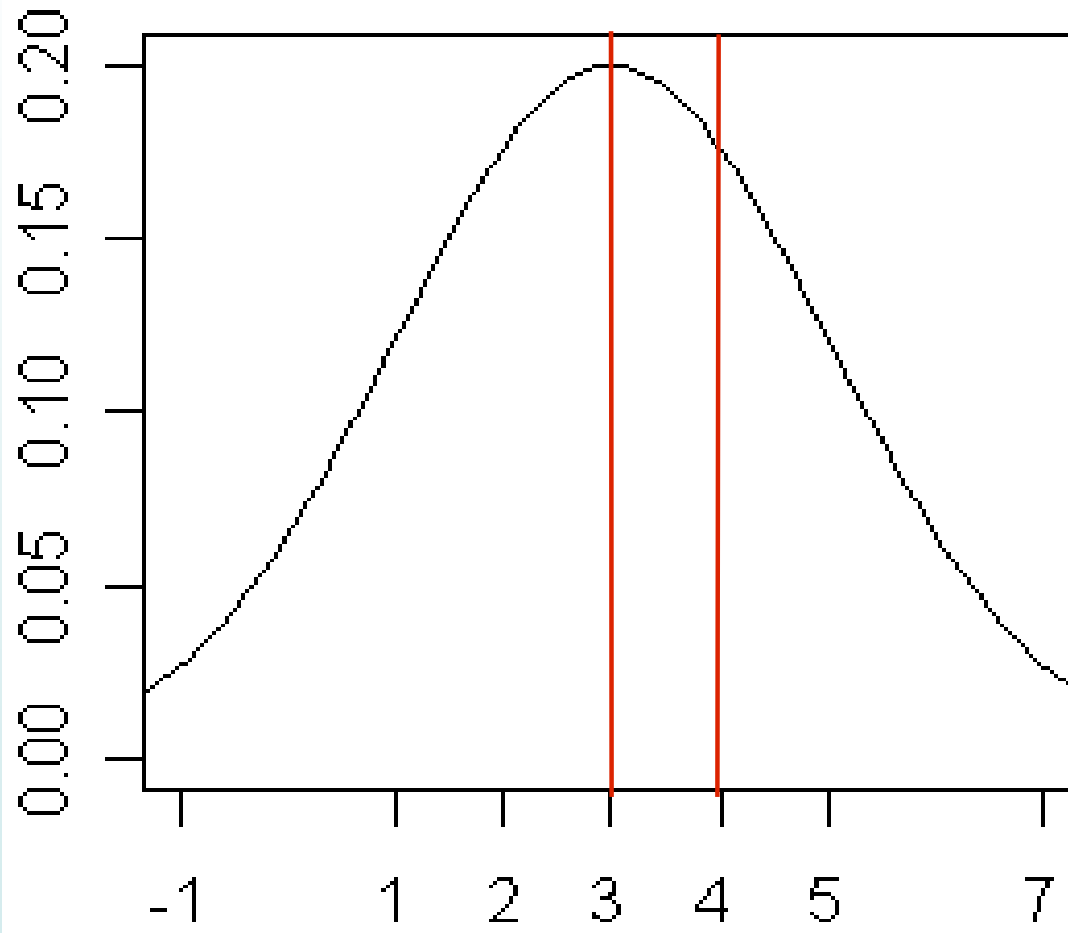
Judge, G.; Schechter, L. (2006): Detection problems in survey data using Benford's Law: Preprint.

Nigrini, M. (1999): I've Got Your Number. How a Mathematical Phenomenon Can Help CPAs Uncover Fraud and Other Irregularities, in: Journal of Accountancy, May: 79-83

Pietronero, L.; Tossati, E.; Tossati, V., Vespignani, A. (2001): Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf, in: Physica A, 293(1): 297-304



# Unified distribution of digits



## Normal distribution

- Mean: 3
- Standard error: 2