Journal of Statistical and Econometric Methods, vol.1, no.3, 2012, 1-15
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2012

Disclosure of cheating by statistical methods

Gustav Kristensen[1]

Abstract

Statistics can be used to reveal cheating and corruption. Statistics can see what cannot be perceived with the naked eye.
This article discusses the disclosure of cheating by statistics seen in relation to similar studies.
The evaluation process of a semester course at a university is a data generating process.
After the rules the teacher hands out questionnaires to the students at the end of a lecture and leaves the room. While seated in rows in the auditorium the students fill in the blanks. The completed questionnaires are collected in bunches for each row and then combined to one bunch, which a student brings to the secretariat. At the secretariat the questionnaires are numbered in the order they came in. The data are entered in a computer where average and standard deviation are calculated. The result is sent to the teacher along with the filled out questionnaires for his comments.
The question is here: Could the teacher have filled out the questionnaires himself? If the answer is yes, this paper shows how he can be revealed.

Keywords: Input output analysis, cheating, teachers, students, geometric distribution, group creation

---

[1] University of Southern Denmark, Odense, e-mail: guk@sam.sdu.dk

Article Info: *Received* : June 21, 2012. *Revised* : August 9, 2012
Published online : November 30, 2012

1 Introduction

Shuffling a deck of cards does not easily lead to randomness. It is notable that when shuffling a deck of cards the first shuffling does little to disturb the original order. Thereby a skilled cheat can exploit the lack of randomness to gain advantage over other players. Diaconis and Graham (2012) describe in detail how a few easy-to-grasp mathematical principles are the basis for many of the most astonishing tricks.

Benford's law (1938) says that people have more occasions to calculate with numbers beginning with 1 or 2 than with 8 or 9. Benford found that the frequency of the digit 'a' being the first digit of a decimal number was closely approximated by: $F_a = \log(a+1) - \log a$. This has become known as Benford's law. More work with Benford's law was done by Leemis, Schmeiser, and Evans (2000) who offer a probability distribution function that obeys Benford's low, while Janvresse and de la Rue (2000) offer a probabilistic explanation where Benford can be seen as a mixture of uniform distributions.

Related to the Euro crisis Anja Kelber (2011) writes about how the mathematicians prove "creative bookkeeping" in Greece, using Benford's law.

Scheers and Dayton (1987) estimate students cheating. Jacob and Levitt (2003), develop an algorithm for determining the prevalence of teachers cheating. Similarly, Levitt and Dubner (2005) show how teachers' cheating to get better results for their pupils by adjusting their pupils' answers rapidly results in a system that is fairly easy to reveal.

2 The Input Output Analysis

At the end of each semester university students evaluate their courses and teachers. The question is now: Could a teacher make his "own" evaluations without being revealed?

In a classroom with e.g. 100 students the students are asked to evaluate the teacher on a scale from 1-6, where 6 is the highest/best.

The teacher asks the students to fill in the questionnaires for the evaluation and leaves the room. An appointed student collects the evaluations, while the teacher is outside the room, and delivers them to the study secretariat. The secretariat which handles the evaluations gets the bunch of evaluations from the students and provides each with a number according to the order of receipt.

The teacher, when he sees the evaluations, is surprised to see sequences of highly similar evaluations, as if the students' evaluations were influenced by their right-hand and left-hand neighbors.

To control for this, the observations must of course be compared with a *random* distribution. The big question is, however, what is the *random* distribution to be compared with?

On this basis he forms the hypothesis: Deviations from the *geometric distribution* in the sequences of evaluations can be seen as an expression that students with the same attitude towards the teacher tend to group together – or those students who group together get the same opinion.

As in the deck of cards the order is surprisingly well preserved during the collection of questionnaires.

Suppose that the collection is well ordered. The questionnaires are collected in exactly the order they lie in on the rows. Then of course the order is preserved. However, if each group takes the bunch of the group and puts it in between other groups' bunches, the (group) order is likewise preserved no matter how the groups join their bunches of papers. The order in which the papers are collected among the rows likewise does not influence the groupings.

3 The Relevant Random Distribution Function

A data generating process which obviously comes near to the description of group formation on a line is the Negative Binomial distribution.

The core of the creation of a negative binomial distribution can be described in the following way: Throw a dice until "r" outcome of "6" has appeared and see the distribution of the sequences of "non-6". Here you can read "r" as the number of successes.

$r > 0$ number of successes ("6") ("success")

$0<p<1$ probability of failures ("non-6") ("failure")

The number of "non-6" before "6" occurring for the "r'th" time is "k". $k \in \{0, 1, 2, \ldots\}$

The probability mass function is *the Negative Binomial function*:

$$f(x) = \frac{(k+r-1)!}{(r-1)!(k)!} p^{k} (1-p)^{r} \quad (1)$$

The focus of this article is the distribution function of "failures" or the length of limited sequences.

3.1 The Geometric distribution. *Assume* r = 1

When a dice is thrown repeatedly until "success" appears for the first time; that is, r = 1, the sequences of "failures" ("non-6") will then follow a *geometric distribution:*

$$f(x) = p^{k} * (1-p) \quad (2)$$

The distribution of sequences of different lengths can be created by repeating the process e.g. n times.

Intuitively, it can be understood in this way: What is the probability of "k" tries before the outcome "6", where "6" occurs in trail (k+1). "Non-6" has the probability p = 5/6 while "6" has the probability (1-p) = 1/6.

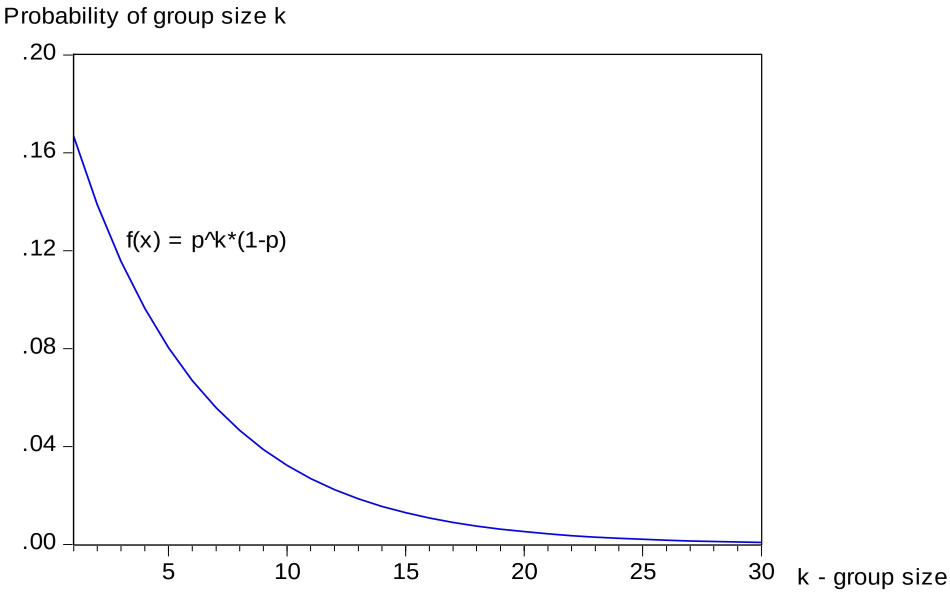

Figure 1: The probability distribution of trails as a function of k, using formula (2). For simplicity here shown as a continuous curve, actually a *discrete* distribution.

Figure 1 shows that a sequence length of 0 has the probability of .166666, while a sequence length of 4 has the probability of .08, when p = 5/6.
The process can be described as:

Model A Series of failures # Success

Or as the pattern

X X X X **Z**
X X X X X X X X X X X X **Z**
Z
X X X X X X X X **Z**
:::::

4 A Two-Dimensional Distribution Function

Now we reformulate the model to address *group size*. The length of "failures" is now seen as the size of groups.

Model B

XXX XXXXXXX XX XXXX …..

The blanks limit the groups.

We now reformulate the problem to: What is the probability function for an outcome 'X' to be a *member of a group of size k*?

Actually, to answer that question, multiply the group size with k and, in order to keep the sum of probabilities the same, adjust by multiplying again with (1-p), (see appendix). We call this frequency function g(k), [RAN_{kp}]. Thus, the Geometric distribution function is multiplied with k*(1-p) to create a new random distribution function for the probability of being a member of a (random) group of a certain size:

$$g(k) = RAN_{kp} = k*(1-p)*p^{k}*(1-p) \tag{3}$$

or

$$RAN_{kp} = k*p^{k}*(1-p)^{2} \tag{4}$$

The geometric distribution of groups and the distribution function for the individuals' probability of being members of a group of size k are shown in Figure 2.

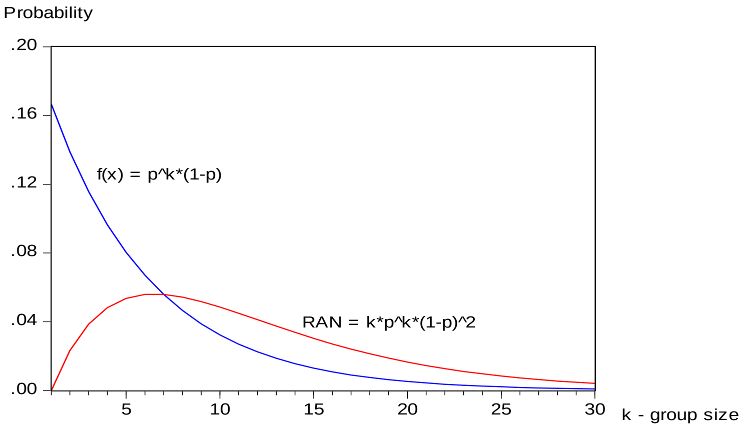

Figure 2: Probability of group size k versus the probability of being a member of a group size k

4.1 Different groups

We now include different groups. Groups which can be distinguished from each other and where the probability of being a member of one of the distinctive groups differs.

Model C

XXX Z YYYYY ZZZZ YYY X Z YYYY

where

$P(X) = p_x$; $P(Y) = p_y$; $P(Z) = p_z$ (5)

$p_x + p_y + p_z = 1$ (6)

As we see the sequences of X, Y and Z in one formula of a geometric distribution we have:

$f(p_j, k) = p_j^{\ k} (1-p_j)$ (7)

$j \in \{x, y, z\}$

$$\sum_j \sum_k p_j^{\ k}(1-p_j) = 1 \quad (8)$$

Once again we change the game and ask for the probability of a certain person of a given characteristic to be member of a certain group size.

$g(p_j, k) = k{*}p_j^{\ k}(1-p_j)^2$ (9)

$$\sum_j \sum_k k * p_j^{\ k}(1-p_j)^2 = 1 \quad (10)$$

The new model we will discuss is based on N observations and is actually formed by a number of outcomes limiting each other. In the case of three types of outcome, we have:

Each letter, e.g. Y, can be seen as being limited by e.g. non-Y. “k” now only refers to the number in each section of identical letters. Thus, in this case we “stay” in the geometric distribution (8).

4.2 From frequency to actual number

The two-dimensional frequency function on random data can be formed as:

$N{*}g(p_j, k) = NRAN_{kp} = N{*}\ k * p_j^{\ k} * (1 - p_j)^2$ (11)

$$\sum_j \sum_k N * k * p_j^{\ k}(1-p_j)^2 = N \quad (12)$$

NRAN is the total number of *calculated* observations in groups of probability p_j, and the number of members, k. As seen below in TABLES 1-4, NACT is the matching number of *actual* observations in a given subgroup that is easy to collect empirically.

5 How to create your own Evaluations - or to reveal your colleague

Now model C is supposed to indicate the students' evaluation of their teacher. The letters could be the marks he gets

XXX Z YYYYY ZZZZ YYY X Z YYYY

6 6 6 4 5 5 5 5 5 4 4 4 4 5 5 5 6 4 5 5 5 5

Let us say you want the distribution of your evaluations to be:

6	50% of the evaluations
5	33% of the evaluations
4	17% of the evaluations
3	0% of the evaluations
2	0% of the evaluations

0% of the evaluations

There are two good methods for creating that result in a random procedure.

5.1 Throw a dice (DICE)

If you have e.g. 100 students, you can throw a dice 100 times and register the evaluations as:

6	at dice outcomes 6, 5, and 4	$p_{.50}$
5	at dice outcomes 3 and 2	$p_{.33}$
4	at dice outcome 1	$p_{.17}$

Then you have the average evaluation of 5.33 – not bad. And the sequences of "evaluations" from an ordered registration will follow the distribution described by formula (2).

5.2 Randomized "evaluations" (RAN)

A more simple procedure is, however, randomization. In this process you enter 50 times 6, 33 times 5, and 17 times 4 into the computer. Put in a random number, and sort the evaluations.

Again you have the average evaluation of 5.33 – also not bad. And the sequences of "evaluations" will again follow the distribution in (2).

An extra benefit of this method is that a random element will be reduced because the method matches stratified sampling, unlike dice rolling.

A (continuous) *dice procedure* as well as a *randomizing procedure* both follow the distribution in (4) with respect to members in groups and (2) with respect of length. However, the dice procedure gives the greatest residual variance.

The teacher, who in that way makes his own evaluations, however, can claim that his students' evaluations actually follow the same rules.

We shall now see if that is true by some empirical estimation on student data.

5.3 Actual distributions (ACT)

The above formulas referred to random distributions. In real life, randomness is seldom a part of human behavior. In order to have a frequency function that is able to measure a group formation process as the deviation from a random distribution created after the above-mentioned data generating process, we write a functional form of (4):

$$ACT_{kp} = \alpha_0 * k^{\alpha_1} * p_i^{\,k} * (1 - p_i)^2 \qquad (13)$$

We can now make a distinction between:
"Random distribution"
Random distribution for membership of group seize k appears for:

$\alpha_0 = 1$

$\alpha_1 = 1$

"Group effect"
"Group effect" is here defined as a concentration around a certain group size $1<k<n$, where n is the highest number in a group. A clustering is most likely around functional groups of 2-4 with a low number of separated individuals ('group' size 1):

$\alpha_0 < 1$

$1 < \alpha_1$

In order to emphasize (separate) the group effect we develop α_0 and α_1:

$\alpha_0 = 1 - \beta_{00}$

$\alpha_1 = 1 + \beta_{10} - \beta_{11} k$

"Anti-group effect"

"Anti-group effect" is here defined as a concentration around small groups possibly of size 1, e.g. men and women placed around a dinner table at a party.

$\alpha_0 > 1$

$0 < \alpha_1 < 1$

The negative group effect, which in this context is seen as a concentration of groups of size 1, is expressed by:

$\alpha_0 = 1 + \beta_{00}$

$\alpha_1 = 1 - \beta_{10} + \beta_{11} k$

If the values of β_{ij} above are significant, we have positive or negative group effects.

The collected data based on a *genuine* (actual data, ACT) evaluation procedure will in this section be compared with a dataset created by a *random* procedure based on equation (4).

An empirically improved model is:

$$ACT_{kp} = (1+\beta_{00})^{*} k^{(1+\beta_{10}+\beta_{21}k)} * p_i^{\,k} *(1 - p_i)^2 \quad (14)$$

The sizes of the calculated groups is *upward biased* because two groups with the same opinion/content when lying side by side can only be interpreted as one group. The upward bias grows at increasing p, and in the present dataset one outlier is dropped, indicating a group size of 14 for p=.57.

6 Actual Data and the Estimated Equations

Estimating on actual data ACT, replace RAN. Actually, we most easily find NACT and calculate ACT. Actual data are from two groups of students, here referred to as: Negot and Bachelor.

For Negot students N = 101 plus 103; p takes the interval .01- .57. Equation (14) is estimated on the two datasets together shown in the appendix, Tables 1 and 2.

Actual data for Negot

$$ACT_{kp} = (1 - .2250)^{*}k^{(1+1.1858-.3063^{*}k)} * p^{k-1} (1- p)^2$$

t (-1.94) (2.22)(-1.95)

$R^2 = .55$ Obs. = 41

Conclusion: Group formation is significant and can be observed in the estimated coefficients. The results follow the form in equation (14).

As seen in the appendix, Table 2, evaluation of professional skills, there is an outlier of a sequence on 14. Most likely this is two different student groups of the same opinion lying side by side. Omitting this outlier, we get the equation below:

6.1 Actual frequency for Negot, omitting the outlier k=14

N is now 101 pedagogical plus 89 professional evaluations:

$ACT_{kp} = (1 - .1868) * k^{(1+1.2322-.3114*k)} * p^k (1-p)^2$

t (-1.78) (2.73)(-2.36)

$R^2 = .67$ Obs. = 40

Conclusion: Group formation can again be seen. The result still follows the form in equation (14).

6.2 Actual frequency for Bachelor

Bachelor students: p takes the interval .16-.50. Equation (14) is estimated on the two last datasets shown in the appendix:

$ACT_{kp} = (1-.4246) * k^{(1+1.3746-.1836*k)} * p^k (1-p)^2$

t (-3.79) (2.64)(-1.45)

$R^2 = .53$ Obs. = 41

Conclusion: Strong group formation.

6.3 The pure group effect

The pure group effect can now be illustrated by the formula:

$$\text{Group effect} = ACT_{kp} - RAN_{kp} \tag{15}$$

$$\text{Group effect} = (1+\beta_{00}) * k^{(1+\beta_{10}+\beta_{21}k)} * p_i^k * (1 - p_i)^2 - k * p_i^k * (1-p_i)^2 \tag{16}$$

Figures 3 and 4 show the group effect for Bachelor and Negot students as group formation around group sizes of 3 and 2.

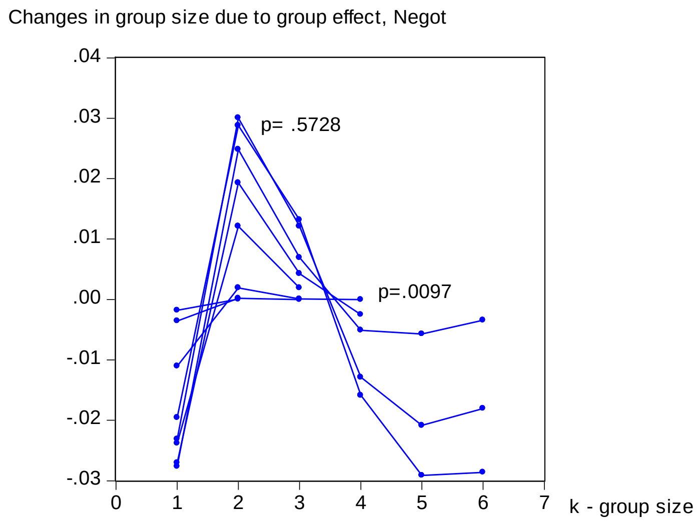

Figure 3: The group effect (Negot) defined as units moving towards medium sequences (here towards groups of size 2) compared to a random distribution of sequences.

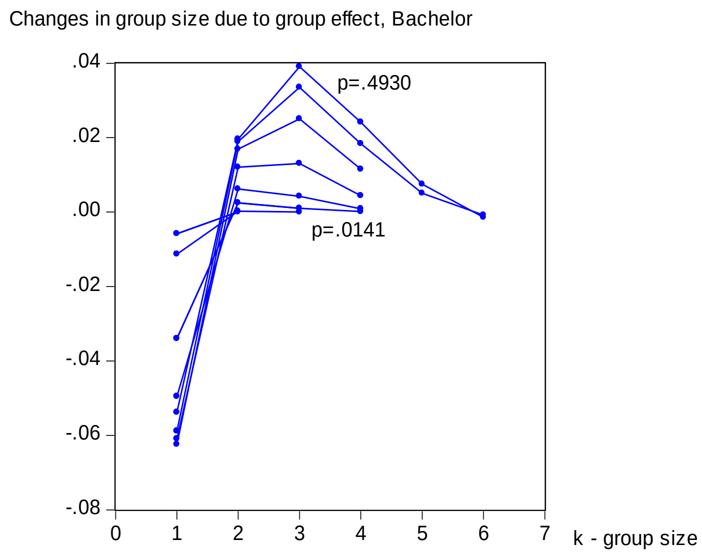

Figure 4: For categories with few members the group effect (Bachelor) is going towards groups with 2 members, and for bigger categories towards 3.

7 Conclusion

When questionnaires are ordered randomly, series of uniform answers tend to follow a geometric distribution.

When the variables, here student evaluations, are created as defined above, the evaluations will mirror cooperation between students sitting close together by creating more groups of size 2-3 (in this case) than expected. The group effect in the collected questionnaires is rather robust during the collection of questionnaires.

This pattern of group formation cannot be constructed by simple formulas, and the results (ACT) are thus a strong indication that the teacher did not create his own evaluations.

The absence of the group effect is a strong indicator of cheating.

Acknowledgements. For encouragement and good advice in writing this article I am indebted to Erik Holst, Terkel Christiansen, Gisela Hostenkamp, and Steven Levitt.

Appendix

Grouping (actual) of evaluations of the teacher - pedagogical and professional

This appendix include actual data from four teacher evaluations.

Table 1: Actual Distribution (NACT). Negot - pedagogical

Eval.	**p**	**k**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**Sum**
6	.2772		11	8	9	0							28
5	.5050		14	14	12	4	0	0	7	0			51
4	.1980		12	8	0								20
3	.0198		0	2	0								2
	1.0000		**37**	**32**	**21**	**4**	**0**	**0**	**7**	**0**			**101**

Table 2: Actual Distribution (NACT). Negot - professional

Eval.	**p**	**k**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**14**	**Sum**
6	.5728		4	14	15	4	0	0	0	8	0	14	59
5	.3495		11	10	0	4	5	6	0				36
4	.0680		4	0	3	0							7
3	.0097		1	0									1
	1.0000		**20**	**24**	**18**	**8**	**5**	**6**	**0**	**8**	**0**	**14**	**103**

Table 3: Actual Distribution (NACT). Bachelor - pedagogical

Eval.	**p**	**k**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Sum**
6	.2676		4	12	3	0				19
5	.4366		6	6	9	0	10	0		31
4	.1690		7	2	3	0				12
3	.0986		3	4	0					7
2	.0141		1	0						1
4.5	.0141		1	0						1
	1.0000		**22**	**24**	**15**	**0**	**10**	**0**		**71**

Table 4: Actual Distribution (NACT). Bachelor - professional

Eval.	**p**	**k**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Sum**
6	.3662		6	8	12	0				26
5	.4930		6	10	9	4	0	6	0	35
4	.0986		2	2	3	0				7
3	.0282		0	2	0					2
3.5	.0141		1	0						1
	1.0000		**15**	**22**	**24**	**4**	**0**	**6**	**0**	**71**

Mathematical Appendix

Probability of membership of group size k.

Description of the move from (2) to (4). You have the formula for the sum of a quotient row

$$1 + p^1 + p^2 + p^3 + p^4 + \ldots\ldots + p^{n-1} = \frac{1-p^n}{1-p} \quad \text{(A1)}$$

$$p^1 + p^2 + p^3 + p^4 + \ldots\ldots + p^{n-1} = \frac{1-p^n}{1-p} - 1 \quad \text{(A2)}$$

For n→∞

$$\frac{1}{1-p} - 1 = \frac{1-1+p}{1-p} = \frac{p}{1-p} \quad \text{(A3)}$$

or

$$p^1 + p^2 + p^3 + p^4 + \ldots\ldots + p^{n-1} = \frac{p}{1-p} \quad \text{(A4)}$$

You differentiate the row (A1) with p and gets

$$0 + 1 + 2^* p^1 + 3^* p^2 + 4^*p^3 + 5^*p^4 + \ldots\ldots + (n-1)^*p^{n-2} = \frac{d\left(\frac{1-p^n}{1-p}\right)}{dp} \quad \text{(A5)}$$

Multiply with p, and for n→∞, we have

$$1^*p + 2^* p^2 + 3^* p^3 + 4^*p^4 + \ldots\ldots + (n-1)^*p^{n-1}$$

$$= \frac{d\left(\frac{1-p^n}{1-p}\right)}{dp} p = \frac{p}{(1-p)(1-p)} \quad \text{(A6)}$$

$$1*p*(1\text{-}p) + 2*p^2*(1\text{-}p) + 3*p^3*(1\text{-}p) + \cdots + (n\text{-}1)*p^{n-1}*(1\text{-}p) = \frac{p}{(1-p)} \qquad (A7)$$

It is seen that multiplying the right side in (A2) with k*(1-p) will for $n \to \infty$ give the right side of (A7).

References

[1] W.H. Angoff, The Development of Statistical Indices for Detecting Cheaters, *Journal of the American Statistical Association*, **69**(345), (1991), 44-49.

[2] F. Benford, The law of Anomalous Numbers, *Proceedings of the American Philosophical Society*, **78**(4), (1938), 551-572.

[3] P. Diaconis and R. Graham, *Magical Mathematics. The Mathematical Ideas that Animate Great Magic Tricks,* Princeton University Press, Princeton and Oxford, 2012.

[4] B.A. Jacob and S.D. Levitt, Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating, *Quarterly Journal of Economics,* **118**(3), (2003), 843-877.

[5] A. Kelber, Man muss nur die richtigen Leute fragen. 2011. (http://tagesschau.de/schlusslicht/benfordschesgesetz100.html)

[6] G.N. Kristensen, Group formation on a line, *Symposium i Anvendt Statistik. Red. Peter Linde, Københavns Universitet,* 219-232, 2010.

[7] L.M. Leemis, B.W. Schmeiser and D.L. Evans, Survival Distributions Satisfying Benford's Law, *The American Statistician,* **54**(4), 2000, 236-241.

[8] S.D. Levitt and S.J. Dubner, *Freakonomics. A Rogue Economist Explores the Hidden Side of Everything,* (case pp 27-38), HarperCollins Publishers, 2005.

[9] N.J. Scheers, and C.M. Dayton, Improved Estimation of Academic Cheating Behavior Using the Randomized Response Technique, *Research in Higher Education,* **26**(1), (1987), 61-69.