

# Critical issues with the Pearson's chi-square test

VLADIMIR GURVICH<sup>✉</sup> AND MARIYA NAUMOVA\*<sup>✉</sup>

**ABSTRACT.** Pearson's chi-square tests are among the most commonly applied statistical tools across a wide range of scientific disciplines, including medicine, engineering, biology, sociology, marketing and business. However, its usage in some areas is not correct. For example, the chi-square test for homogeneity of proportions (that is, comparing proportions across groups in a contingency table) is frequently used to verify if the rows of a given nonnegative  $m \times n$  (contingency) matrix  $A$  are proportional. The null hypothesis  $H_0$ : " $m$  rows are proportional" (for the whole population) is rejected with confidence level  $1 - \alpha$  if and only if  $\chi^2_{stat} > \chi^2_{crit}$ , where the first term is given by Pearson's formula, while the second one depends only on  $m, n$ , and  $\alpha$ , but not on the entries of  $A$ .

It is immediate to notice that the Pearson's formula is not invariant. More precisely, whenever we multiply all entries of  $A$  by a constant  $c$ , the value  $\chi^2_{stat}(A)$  is multiplied by  $c$ , too,  $\chi^2_{stat}(cA) = c\chi^2_{stat}(A)$ . Thus, if all rows of  $A$  are exactly proportional then  $\chi^2_{stat}(cA) = c\chi^2_{stat}(A) = 0$  for any  $c$  and any  $\alpha$ . Otherwise,  $\chi^2_{stat}(cA)$  becomes arbitrary large or small, as positive  $c$  is increasing or decreasing. Hence, at any fixed significance level  $\alpha$ , the null hypothesis  $H_0$  will be rejected with confidence  $1 - \alpha$ , when  $c$  is sufficiently large and not rejected when  $c$  is sufficiently small. Yet, obviously, the rows of  $cA$  should be proportional or not for all  $c$  simultaneously. For this reason, Pearson's test certainly cannot be applied to "physical data", which are obtained by measurements. Indeed, in this case matrix  $A$  depends on the unit of measurement. The test can be applied only to categorical data and even then some further limitations are required, which we consider in this paper.

**Keywords:** Pearson chi-square test, difference between two proportions, goodness of fit, contingency tables.

**2020 Mathematics Subject Classification:** 62-00, 62-07XX, 62P10, 62P20, 62P25.

## 1. INTRODUCTION

The chi-square test for proportionality determines whether there is a statistically significant difference between the proportions of a certain outcome in two independent groups.

The original chi-square test, commonly referred to as Pearson's chi-square, originated from Karl Pearson's papers in the late 1800 - early 1900s [11, 12, 13, 14]. It is used both as a "goodness of fit test" - where data are classified along a single dimension - and as a test for contingency tables, where classification occurs across two or more dimensions. A historical overview of the development of the test is provided in [2, 15]. Pearson suggested but did not provide a proof that the test statistic follows the chi-square distribution [2]. The correction regarding the number of degrees of freedom was addressed in Fisher's papers published in 1922 and 1924 [5, 6].

Currently, the chi-square test is extensively utilized across a broad spectrum of research disciplines in analyzing categorical data. Its applications span fields such as social sciences, biomedical research, economics, education, and marketing, where it is employed to assess associations between variables and evaluate the goodness of fit of observed data to expected

distributions. This widespread adoption underscores its substantial role in empirical research and statistical inference. According to [1], the chi-square test has seen more use than any other statistical test beside possibly Student's  $t$ -test. However, despite its widespread application, multiple authors (see for ex., [2, 7, 9, 10]) mention that the chi-square test presents several methodological limitations, such as (not theoretically proven) lower and upper bounds on the sample size, the assumption that the data are derived from independent observations, etc.

In this paper, we demonstrate an extreme sensitivity of the test to the sample sizes and data representation. Specifically, we show that if the given contingency matrix is multiplied by a constant  $c$ , the test statistic is multiplied by  $c$  as well. The results of the test can therefore be easily manipulated (to reject or not reject  $H_0$ ).

**Remark 1.1.** *Similar problems were recently detected for the analysis of variance (ANOVA) and the Tukey-Kramer test. Tukey-Kramer's formula (see (4) in [8]) of the critical range for groups of observations  $i$  and  $j$  depends not only on these two groups but also on all other groups. This contradicts common sense, since these other groups may be not related to groups  $i$  and  $j$  at all, and immediately leads to some logical contradictions in ANOVA with more than two groups of observations; see [8] for definitions and more details.*

## 2. PEARSON CHI-SQUARE TEST FOR HOMOGENEITY OF PROPORTIONS

All three omnibus tests in the Pearson family - goodness of fit, independence, and homogeneity of proportions - share basically the same underlying formula for the test statistic. We will concentrate on the homogeneity test.

Consider a scenario in which  $T$  outcomes of multinomial trials are classified according to two distinct criteria, into one of the same  $m$  categories, in one of  $n$  groups. The outcomes can be presented in a two-way contingency table with  $m$  rows and  $n$  columns, as shown in Table 1. For example, in [4], a population-based case-control study of prostatic cancer in Alberta

	Group 1	Group 2	...	Group $j$	...	Group $n$	Row Total
Category 1	$O_{11}$	$O_{12}$	...	$O_{1j}$	...	$O_{1n}$	$O_{1\cdot}$
Category 2	$O_{21}$	$O_{22}$	...	$O_{2j}$	...	$O_{2n}$	$O_{2\cdot}$
...	...	...		...		...	...
Category $i$	$O_{i1}$	$O_{i2}$	...	$O_{ij}$	...	$O_{in}$	$O_{i\cdot}$
...	...	...		...		...	...
Category $m$	$O_{m1}$	$O_{m2}$	...	$O_{mj}$	...	$O_{mn}$	$O_{m\cdot}$
Column Total	$O_{\cdot 1}$	$O_{\cdot 2}$	...	$O_{\cdot j}$	...	$O_{\cdot n}$	$T$

TABLE 1. General structure of an  $m \times n$  contingency table. Here,  $O_{ij}$  denotes the observed (absolute) frequency in cell  $(i, j)$ ,  $O_{i\cdot}$ ,  $O_{\cdot j}$  denote the marginal totals for row  $i$  and column  $j$  respectively, and  $T$  is the grand total.

uses the contingency table related to 376 newly diagnosed prostatic cancer patients and 620 controls, group-matched based on their ethnicity (see Table 2). Let  $O_{ij}$  denote the observed (absolute) frequency in the cell corresponding to row  $i$  and column  $j$ , for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ .

	Cases	Controls	Row Total
British	200	279	479
French	16	20	36
German	55	93	148
Ukrainian	31	79	110
Others	74	149	223
Column Total	376	620	996

TABLE 2. The contingency table for a population-based case-control study of prostatic cancer in Alberta, with 376 newly diagnosed prostatic cancer patients and 620 controls of different ethnicities [4].

Let

$$\begin{aligned}
 R_i &= \sum_{j=1}^n O_{ij}, \quad i \in \{1, 2, \dots, m\} && \text{(row sum for population } i), \\
 C_j &= \sum_{i=1}^m O_{ij}, \quad j \in \{1, 2, \dots, n\} && \text{(column sum for category } j), \\
 T &= \sum_{i=1}^m \sum_{j=1}^n O_{ij}. && \text{(total sample size).}
 \end{aligned}$$

The null hypothesis is:

$$H_0 : \pi_{1j} = \pi_{2j} = \dots = \pi_{mj}, \quad \forall j \in \{1, \dots, n\},$$

where  $\pi_{ij} \in [0, 1]$  denotes the proportion of individuals in group  $i$  falling into category  $j$ , and  $\sum_{j=1}^n \pi_{ij} = 1$  for each  $i$ . Under  $H_0$ , the expected frequency  $E_{ij}$  in cell  $(i, j)$  is given by:

$$E_{ij} = \frac{R_i \cdot C_j}{N}, \quad \forall i = 1, \dots, m; \quad j = 1, \dots, n.$$

Define the test statistic:

$$(2.1) \quad \chi_{stat}^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

Under the test assumptions (see Section 3), the test statistic  $\chi_{stat}^2$  approximately follows a chi-square distribution with  $(m-1)(n-1)$  degrees of freedom under  $H_0$ . Let  $\alpha \in (0, 1)$  be the significance level. Then the decision rule is:

$$\text{Reject } H_0 \quad \text{if} \quad \chi_{stat}^2 > \chi_{1-\alpha, (m-1)(n-1)}^2,$$

where  $\chi_{1-\alpha, (m-1)(n-1)}^2$  is the  $(1-\alpha)$ -quantile of the chi-square distribution with  $(m-1)(n-1)$  degrees of freedom.

### 3. ASSUMPTIONS OF THE CHI-SQUARE TEST

While the mathematical formulation of the chi-square test statistic is relatively simple - comparing observed frequencies with expected frequencies under a specified null hypothesis- the validity of the test's conclusions rests on a set of non-trivial assumptions. These assumptions

pertain to the way the data are generated, the structure of the contingency table, and the theoretical conditions under which the chi-square distribution serves as a valid approximation to the true sampling distribution of the test statistic. Specifically, the assumptions of the test are as follows [10]:

- (i) **Random Sampling:** The data must be drawn via random sampling.
- (ii) **Categorical Data:** The variables involved must be categorical. The full set of  $m \times n$  cells represents all possible combinations of category levels, and each observation must belong to exactly one cell.
- (iii) **Expected Frequency Condition:** The expected frequency in each cell should be sufficiently large to ensure the validity of the chi-square approximation. Specifically, it is recommended that for each cell  $(i, j)$ ,

$$E_{ij} \geq 5^1, \quad \forall i \in \{1, \dots, m\}, j \in \{1, \dots, n\}.$$

- (iv) **Independence of Observations:** Each observation must contribute to exactly one cell, and observations must be independent. That is,

$$\text{Cov}(O_{ij}, O_{kl}) = 0 \quad \text{for all } (i, j) \neq (k, l).$$

- (v) **Fixed Margins (if applicable):** For the test of homogeneity or independence, it is often assumed that either the row totals, the column totals, or both are fixed by the sampling design or conditioning.

#### 4. ONLY INVARIANT TESTS CAN VERIFY PROPORTIONALITY

Obviously,  $\chi_{stat}^2(cA) = c\chi_{stat}^2(A)$  for all real  $c \geq 0$ . The null hypothesis  $H_0$  stated in Section 2 is rejected for any fixed significance level  $\alpha$ , if and only if  $\chi_{stat}^2(A) > \chi_{crit}^2(A)$ , where  $\chi_{crit}^2(A) = \chi_{1-\alpha, (m-1)(n-1)}^2$  is the  $(1-\alpha)$ -quantile of the chi-square distribution with  $(m-1)(n-1)$  degrees of freedom, which depends only on  $m, n$ , and  $\alpha$ , but not on the entries of  $A$ .

**Observation 4.1.** *The following five statements are equivalent:*

- (a) *the rows of  $A$  are exactly proportional;*
- (b) *the columns of  $A$  are exactly proportional;*
- (c)  *$O_{ij} = E_{ij}$  for all  $i = 1, \dots, m$  and  $j = 1, \dots, n$ ;*
- (d)  *$\chi_{stat}^2(A) = 0$ ;*
- (e)  *$\chi_{stat}^2(cA) = 0$  for all real positive  $c$ .*

Furthermore, if statements (a - e) fail, then  $\chi_{stat}^2(A) > 0$  and for any fixed significance level  $\alpha$ , the null hypothesis  $H_0$  will be rejected for  $cA$  with confidence  $1 - \alpha$ , if  $c$  is sufficiently large, and it will not be rejected if  $c$  is sufficiently small.

*Proof.* It is enough to notice that in the numerator and denominator of (2.1) are, respectively, a quadratic and linear functions of the entries of  $A$ . Hence,  $\chi_{stat}^2(cA) = c\chi_{stat}^2(A)$ .  $\square$

Yet, obviously, the result of testing proportionality of the rows of  $cA$  should not linearly depend on  $c$ .

<sup>1</sup>In some sources,  $E_{ij} \geq 10$  [2], but neither requirement seems to be mathematically justified.

## 5. EXAMPLES

**Example 5.1.** We consider the following  $2 \times 2$  contingency table:

	Group 1	Group 2	Row Total
Category A	1	1	2
Category B	1	11	12
Column Total	2	12	14

We test the null hypothesis that the proportions of outcomes are the same across the two groups. The expected value for each cell under the null hypothesis are

$$E_{11} = \frac{2 \cdot 2}{14} = \frac{4}{14} = \frac{2}{7}, \quad E_{12} = \frac{2 \cdot 12}{14} = \frac{24}{14} = \frac{12}{7},$$

$$E_{21} = \frac{12 \cdot 2}{14} = \frac{24}{14} = \frac{12}{7}, \quad E_{22} = \frac{12 \cdot 12}{14} = \frac{144}{14} = \frac{72}{7},$$

and the chi-square statistic is found to be

$$\chi_{stat}^2 = \frac{(1 - \frac{2}{7})^2}{\frac{2}{7}} + \frac{(1 - \frac{12}{7})^2}{\frac{12}{7}} + \frac{(11 - \frac{12}{7})^2}{\frac{12}{7}} + \frac{(1 - \frac{72}{7})^2}{\frac{72}{7}} = \frac{25}{7} \left( \frac{1}{2} + \frac{1}{6} + \frac{1}{72} \right) = \frac{175}{72}.$$

With 1 degree of freedom, and using the critical value from the chi-square distribution table at significance level  $\alpha = 0.05$ , which is approximately 3.841, we have that

$$\chi_{stat}^2 \approx 2.43 < 3.841.$$

Thus, we fail to reject the null hypothesis. There is not enough evidence to suggest a significant difference in proportions between the two groups. Let's modify the original table by multiplying each entry by 2:

	Group 1	Group 2	Row Total
Category A	2	2	4
Category B	2	22	24
Column Total	4	24	28

Performing the same operations, we find the expected counts under the null hypothesis:

$$E_{11} = \frac{4 \cdot 4}{28} = \frac{16}{28} = \frac{4}{7}, \quad E_{12} = \frac{4 \cdot 24}{28} = \frac{96}{28} = \frac{24}{7},$$

$$E_{21} = \frac{24 \cdot 4}{28} = \frac{96}{28} = \frac{24}{7}, \quad E_{22} = \frac{24 \cdot 24}{28} = \frac{576}{28} = \frac{144}{7}.$$

Then the chi-square statistic is computed as

$$\chi_{stat}^2 = \frac{(2 - \frac{4}{7})^2}{\frac{4}{7}} + \frac{(2 - \frac{24}{7})^2}{\frac{24}{7}} + \frac{(2 - \frac{24}{7})^2}{\frac{24}{7}} + \frac{(22 - \frac{144}{7})^2}{\frac{144}{7}} = \frac{100}{7} \left( \frac{1}{4} + \frac{1}{12} + \frac{1}{144} \right) = \frac{175}{36}.$$

Comparing this value to the critical value  $\chi_{0.05,1}^2 = 3.841$ , we see that

$$\chi_{stat}^2 \approx 4.86 > 3.841,$$

and reject the null hypothesis. There is sufficient evidence to suggest that the proportions differ between the two groups.

Vectors (1, 1) and (1, 11) do not look proportional at all. So it seems a bit strange that we cannot reject proportionality, for the population, with confidence 95%, based on such a sample.

Some researchers (see, for ex. [2, 3, 7]) recommend to apply the Pearson's test only to samples with "large enough" entries. In particular, the expected frequencies in each cell should

generally be at least ten, or in some sources, no less than 5 (both numbers seem to be arbitrarily chosen) [2], or otherwise, the test can result in inaccurate  $p$ -values and reduced statistical power. Indeed, doubling the values in the first sample, we obtain the second one, with row vectors (2, 2) and (2, 22), which already enables us to reject proportionality with confidence 95%. However, one should note that with very large samples, even trivial associations may yield statistically significant results, potentially leading to overinterpretation.

These restrictions are ambiguous and do not solve the problem in general, since, as we already mentioned,  $\chi_{stat}^2(cA) \equiv c\chi_{stat}^2(A)$ .

**Example 5.2.** We illustrate the test for the following  $2 \times 2$  contingency table:

	Group A	Group B	Row Total
Category 1	22	18	40
Category 2	18	22	40
Column Total	40	40	80

Under the null hypothesis of homogeneity, expected counts for all cells are

$$E_{ij} = \frac{40 \cdot 40}{80} = 20, i, j \in \{1, 2\}.$$

The chi-square statistic:

$$\chi_{stat}^2 = \frac{(22 - 20)^2}{20} + \frac{(18 - 20)^2}{20} + \frac{(18 - 20)^2}{20} + \frac{(22 - 20)^2}{20} = \frac{4}{5} = 0.8.$$

Since  $0.8 < 3.841$ , we do not reject the null hypothesis. There is no significant difference in proportions between the groups.

If we modify the original contingency table by multiplying all elements by 1,000 (for example, assuming that the elements of the matrix represent the actual number of cases, while before they represented the number of cases in thousands). This yields the following scaled observed frequencies:

	Group A	Group B	Row Total
Category 1	22000	18000	40000
Category 2	18000	22000	40000
Column Total	40000	40000	80000

Under the null hypothesis of homogeneity, all expected counts are

$$E_{ij} = \frac{40000 \cdot 40000}{80000} = 20000, i, j \in \{1, 2\},$$

and the chi-square statistic:

$$\chi_{stat}^2 = \frac{(22000 - 20000)^2}{20000} + \frac{(18000 - 20000)^2}{20000} + \frac{(18000 - 20000)^2}{20000} + \frac{(22000 - 20000)^2}{20000} = 800.$$

Note that the new  $\chi_{stat}^2$  is 1,000 times greater than the original one. Comparing  $\chi_{stat}^2$  to the critical value ( $800 > 3.841$ ), we reject the null hypothesis. There is a significant difference in proportions between the groups.

In this example, the null hypothesis  $H_0$  claims that all entries of the population are equal. For the first (small) sample, deviations  $\pm 2$  can appear with probability  $> 5\%$ , so we cannot reject  $H_0$  with confidence 95%. Yet, the second sample is 1,000 times larger, so deviations  $\pm 2,000$  have chances  $< 5\%$ , so we reject  $H_0$  with confidence 95%. It is known that the standard deviation of entries in  $cA$  is proportional to  $\sqrt{c}$ , rather than to  $c$ . And still, the value of  $c$  must be limited, even for the categorical data.

**Example 5.3.** In this case, we work with the following  $3 \times 4$  contingency matrix of observed frequencies:

	Group 1	Group 2	Group 3	Group 4	Row Total
Category A	98	86	79	71	334
Category B	78	82	88	51	299
Category C	75	62	82	77	296
Column Total	251	230	249	199	929

The matrix of the expected frequencies is given by

$$E = \begin{bmatrix} 90.3 & 82.7 & 89.5 & 71.6 \\ 80.9 & 73.9 & 80.2 & 64.1 \\ 79.9 & 73.3 & 79.2 & 63.4 \end{bmatrix},$$

and the chi-square statistic is found to be  $\chi_{stat}^2 \approx 11.475$ .

For the given  $3 \times 4$  contingency matrix, the number of degrees of freedom is  $(3 - 1)(4 - 1) = 6$ . So, the  $p$ -value is about 0.07 and for  $\alpha = 0.05$  we don't reject the null hypothesis. In this case, the critical value is 12.59. It is easy to verify that multiplying all entries, for example, by 2 yields the doubled test statistic ( $\chi_{stat}^2 \approx 22.95$ ) and the much lower  $p$ -value ( $\approx 0.0008$ ).

**Example 5.4.** We now provide a construction of a matrix for which a proportionality test could be designed. Suppose that we want to analyze the joint probabilistic distribution of two random variables taking  $m$  and  $n$  distinct values, say,  $X = (x_1, \dots, x_m)$  and  $Y = (y_1, \dots, y_n)$ . Our goal is to estimate the two corresponding probabilistic distributions  $(p_1, \dots, p_m)$  and  $(q_1, \dots, q_n)$  and decide if  $X$  and  $Y$  are independent. A given sample consists of  $T$  trials, each of which gives us the values of  $X$  and  $Y$ , simultaneously. Let  $Z_{ij}$  be the number of trials realizing  $x_i$  and  $y_j$ . Then,  $z_{ij} = Z_{ij}/T$  are the corresponding observed frequencies. Set  $p_i = \sum_{j=1}^n z_{ij}$ ,  $q_j = \sum_{i=1}^m z_{ij}$ , and treat them as the observed frequencies for  $X$  and  $Y$ , respectively. By the above definitions, we have

$$\sum_{i=1}^m p_i = \sum_{j=1}^n q_j = \sum_{i=1}^m \sum_{j=1}^n z_{ij} = 1.$$

We define the expected frequencies by setting  $z_{ij}^e = p_i q_j$ .

Although the last example is also a case of verifying proportionality, it is special. Since the sum of all entries equals 1, multiplication of all observed frequencies  $z_{ij}$  by a constant  $c \neq 1$  is impossible. More generally, the chi-square test will work when the sum of all entries must be constant, or almost constant.

## 6. CONCLUSIONS

The Pearson chi-square test is one of the most widely used tools for analyzing categorical data, particularly in contingency tables. As demonstrated through both theoretical formulation and numerical examples, even under its common assumptions, including sufficiently large expected cell frequencies and the independence of observations, the test is extremely sensitive to the input data, and its inaccurate application can lead to misleading conclusions. The interpretation of the results must therefore be approached with caution, as the heuristic nature of the test may introduce critical biases. Consequently, findings should be considered exploratory rather than definitive.

The Pearson's chi-square test cannot be applied for verifying proportionality for general contingency matrices. We show that the chi-square statistic increases linearly when the entire contingency table is multiplied by a constant factor, reflecting the proportional scaling of observed and expected frequencies. However, the test is applicable in cases where the total sum

of entries is a constant, or almost constant, for example, for testing the independence of two probabilistic distributions, given a sample of pairwise frequencies.

Pearson's test certainly cannot be applied to "physical data", which are obtained by measurements. Indeed, in this case, matrix  $A$  depends on the units of measurement. For example,  $A$  is replaced by  $3,600A$ , when we switch from hours to seconds. The test can be applied only to categorical data and even then further limitations are required; some of them are considered in this paper.

One could modify formula (2.1) making it invariant with respect to multiplication of  $A$  by a constant, for example, by squaring all denominators in (2.1) or by dividing all entries of  $A$  by their sum or maximum. However, then the distribution  $\chi^2_{crit}$  should be modified accordingly. Both above modifications deserve a separate consideration.

#### ACKNOWLEDGMENTS

This research is performed within the framework of the HSE University Basic Research Program.

#### REFERENCES

- [1] D. Baird: *The Fisher/Pearson Chi-squared controversy: a turning point for inductive inference*, Br. J. Philos. Sci., **34** (2) (1983), 105–118.
- [2] W. G. Cochran: *The  $\chi^2$  test of goodness of fit*, Ann. Math. Stat., **23** (3) (1952), 315–345.
- [3] P. E. Greenwood, M. S. Nikulin: *A guide to chi-squared testing*, John Wiley & Sons., New York (1996).
- [4] S. M. Fincham, G. B. Hill, J. Hanson and C. Wijayasinghe: *Epidemiology of prostatic cancer: a case-control study*, The Prostate, **17** (3) (1990), 189–206.
- [5] R. A. Fisher: *On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P*, J. Roy. Statist. Soc., **85** (1) (1922), 87–94.
- [6] R. A. Fisher: *The conditions under which  $\chi^2$  measures the discrepancy between observation and hypothesis*, J. Roy. Statist. Soc., **87** (3) (1924), 442–450.
- [7] T. M. Franke, T. Ho and C. A. Christie: *The chi-square test: Often used and more often misinterpreted*, Am. J. Eval., **33** (3) (2012), 448–458.
- [8] V. Gurvich, M. Naumova: *Logical contradictions in the one-way ANOVA and Tukey-Kramer multiple comparisons tests with more than two groups of observations*, Symmetry, **13** (2021), Article ID: 1387.
- [9] A. E. Kossovsky: *On the mistaken use of the chi-square test in Benford's law*, Stats., **4** (2) (2021), 419–453.
- [10] M. L. McHugh: *The chi-square test of independence*, Biochem. Med., **23** (2) (2013), 143–149.
- [11] K. Pearson: *Contributions to the mathematical theory of evolution*, Philos. Trans. R. Soc. A, **185** (1894), 71–110.
- [12] K. Pearson: *Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material*, Philos. Trans. R. Soc. A, **186** (1895), 343–414.
- [13] K. Pearson: *Mathematical contributions to the theory of evolution — On a form of spurious correlation which may arise when indices are used in the measurement of organs*, Proc. R. Soc. Lond., **60** (359-367) (1897), 489–498.
- [14] K. Pearson: *X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, London Edinburgh Philos. Mag. & J. Sci., **50** (302) (1900), 157–175.
- [15] R. L. Plackett: *Karl Pearson and the chi-squared test*, Int. Stat. Rev., **51** (1) (1983), 59–72.

VLADIMIR GURVICH  
NATIONAL RESEARCH UNIVERSITY  
HIGHER SCHOOL OF ECONOMICS  
MOSCOW, RUSSIA

RUTGERS UNIVERSITY  
RUTCOR  
PISCATAWAY, NJ, USA  
Email address: vgurvich@hse.ru, vladimir.gurvich@gmail.com



MARIYA NAUMOVA  
RUTGERS UNIVERSITY  
RUTGERS BUSINESS SCHOOL  
PISCATAWAY, NJ, USA  
*Email address:* `mnaumova@business.rutgers.edu`