TESTING AND IDENTIFYING VARIABLE DEPENDENCY THROUGH THE FISHER EXACT TEST IN CENTRAL EUROPE ENTERPRISES

Sebastian Kot[1,a,*] and Ismi Rajiani[2,b]

[1]The Management Faculty, Czestochowa University of Technology, 42-201 Czestochowa, Poland
[2]Management Department, Universitas Muhammadiyah Gresik, Gresik, East Java, 61121, Indonesia
[a]sebacat@zim.pcz.czest.pl, [b]ismi.rajiani@umg.ac.id
*Corresponding author

Cite as: *Kot, S., Rajiani, I. (2020). Testing and identifying variable dependency through the fisher exact test in central Europe enterprises. Ekonomicko-manazerske spektrum, 14(1), 10-18.*

Available at: *dx.doi.org/10.26552/ems.2020.1.10-18*

Abstract: The main problem of scientific thinking is the search for truth. There are two main tools used in the process of drawing scientific conclusions. It is an observation of the world around us and the ability to generate 'generalizing knowledge' from these observations. For example, monitoring the development of a company's financial indicator makes it impossible to control all companies in the Czech Republic. The process of generalizing knowledge, for example, by transferring the conclusions from the selection to the whole country, is called an inductive method of reasoning, and this method is used in the research paper. The preference for processing was a data file consisting of 286 companies operating on the Central Europe market. The file was compiled using basic statistical methods, and based on the results, a conclusion can be drawn that can be generalized to all enterprises. There is a brief description of Fisher's exact test in the introductory part of the work, which was the primary tool for data processing. The methodology is followed, where the basic principles of the Fisher's exact test are explained. Subsequently, the final results of SPSS processing are available in the results, and in the end, the results are evaluated.

Keywords: default, solvency, Fisher Exact Test, Chi-Square

JEL Classification: C12, G33, G30

1. Introduction

Hypothesis testing is an essential part of statistical analysis. Statisticians face the difficult task of making decisions. If their approach is responsible, they will try to solve this problem' in a way that is beneficial to the organization they work for. (Zarrin and Lim, 2009) The leaders of this organization have some ideas about the impact of their decisions, but the reality may not at all correspond to their expectations. This is because, without statistical procedures, it is possible to objectively determine the risk of a biased rejection of our assumption. (Higgins et al., 2011).

Statistical considerations base on mutual observation of individual statistics. In this way, we want to verify some assumptions about the properties of the underlying mathematical set, random variables, or random vector. When comparing, we must consider that statistics are always burdened by an inevitable variability of the values from which they were calculated. (Thompson, 2002)

The consequence of this variability (i.e., random errors) may be a difference between the implementations of the two statistics or a real difference in the parameters. Many random influences cause random errors. One of them may be the influence of man in measuring, observing, and recording the process (Scherbaum and Ferreter, 2009). Furthermore, it may be the influence of the human environment on the process he is researching. The resulting values are also affected by the method of measurement and the meter itself. To find out whether the difference between the implementations of the two statistics is statistically significant, i.e., it is not affected only by random effects. We use the testing of statistical hypotheses. (Deleryd, 1999)

The principle of Fisher's exact test is that combinatorial considerations are used to calculate the probabilities that, for a given limited number, we obtain tables that deviate from the null hypothesis as much as the given table (Poon et al., 2018; Bilder and Loughin, 2007). The sum of these probabilities is the p-value of the test, and the given table of observed frequencies acts as a test statistic.

In the research paper we deal with the bankruptcy of companies and their possibilities of statistical processing. As an indicator of economic failure, the risk of default is defined as the probability that the originator will not be able to meet its financial obligations, i. J. Cannot repay principal and / or interest (Duffie and Singleton, 2012; Lando, 2009). Therefore, it is important to better understand the risk of failure (Altman et al., 1977; Ohlson, 1980; Ślusarczyk and Grondys, 2019; Ślusarczyk et al., 2019) and its relationship to firm growth (Fu et al., 2005).

In this research paper, we focused on the size of companies as a factor that can affect the results in terms of whether a given micro-enterprise can withstand changes in the environment as well as in the case of small and medium-sized enterprises. There are many reasons for a bankrupt company. The results of the analyzes suggest that the most serious problems of bankrupt companies can be grouped into three categories: lack of knowledge, debt unavailability and the economic climate. (Carter and Auken, 2006) Bankrupt businesses also appear to be older, more likely in the retail industry, and are organized as equity or partner companies as bankrupt businesses. They are also less likely to use the Internet in their business operations as businesses without bankruptcy. One surprising finding is that while both subgroups considered knowledge important, the non-bankruptcy sample considered it significantly more important than bankrupt companies. This evidence provides governments and academic institutions with information on their efforts to provide resources that can help reduce the incidence of bankruptcy, especially in times of deteriorating economic health (Carter and Auken, 2006, Podobnik et al., 2010).

In investigating the causes of bankruptcy by Bradley and Cowdery (2004) in a small business, it was concluded that the distinction between the use of the terms failure and bankruptcy is almost impossible to define. Although only about fifteen percent of all small businesses are bankrupt, it can be said that bankruptcy is a direct cause of failure. Sheperd in the article Trade Credit and Small Business: The Cause of Business Failure? Bradley and Rubach (2002) define failure as "an organization ceases to perform the functions expected of it." Bradley and Rubach also cited Sherperd in identifying bankruptcy as the subject of the failure because "bankruptcy reveals the organization's failure and the bankruptcy is recognized as an indication that the organization no longer functions properly." The use of bankruptcy statistics and special surveys will allow a closer look at real bankruptcy. The researcher will link the use of default and bankruptcy and provide bankruptcy-specific material when identifiable.

2. Literature review

A similar issue was addressed in a study by Kamei et al. (2017), where they examine bankruptcy in terms of financial variables as well as in terms of network structure variables. They first binarize the variables by introducing a threshold and then select the appropriate set of variables that minimize the p-value in Fisher's exact test. Financial variables related to loans and savings are strongly correlated with bankruptcy, but the variables of the business network and the capital network, including chain bankruptcy, have weaker but significant correlations. Finally, they predict bankruptcy with selected variables and the second-order Ising model, and we confirm that the Ising model has a relatively higher predictive power than the logit model.

Other authors who have focused on how to manage bankruptcy, specifically in Tunisia, are Blazy and Letaief (2017). They collected data on a group of bankrupt companies for the years 1995-2009. They addressed several issues, such as whether Tunisian bankruptcy proceedings generate significant recoveries? Are secured creditors sufficiently protected in bankruptcy and do they influence court decisions? And other questions. Despite the high level of competition between groups of claim holders, the recovery rate of secured creditors remains similar to that of unsecured creditors.

The authors of Bodobnik et al., (2010) analyze the magnitude of dependence and time stability of bankruptcy risk in the US economy using Zipf scaling techniques. They focus on a single risk factor - the debt-to-assets ratio. They found that the exponent of Zipf increases during market crashes. Based on Zipf's analysis, we use Bayes' theorem to combine the conditional probability that a bankrupt company has the debt-to-asset ratio (R) with the conditional probability of bankruptcy of a company with a given value of R. In the case of 2,737 bankrupt companies, they demonstrate a change in bankruptcy proceedings. The pre-litigation company's assets and the assets of the petitioning company follow the distribution of Zipf, but with different exporters, which means that companies with smaller assets adjust their assets more during bankruptcy than companies with larger assets. They found that both assets and liabilities are governed by the Pareto distribution. This finding is not a trivial consequence of the scalable relationship of Zipf's size quantified by the employees. They propose a combined Simon model that simultaneously develops assets and debts with the possibility of bankruptcy.

Using a list of Japanese bankrupt companies, in 1997 author Fujiwara (2004) discussed with Zipf the distribution of the total liabilities of bankrupt companies in high debt. The life of these bankrupt companies has an exponential distribution in correlation with the entry rate of new companies. It also shows that indebtedness and size are highly correlated, so Zipf's law coincides with the law of size distribution.

Thornhill and Amit (2003) addressed the identification of systematic differences in the decisive factors in the failure of firms between firms that failed early in life and firms that fail after successfully negotiating timely commitments to novelty and adolescence (2003). In this case, however, not in terms of the size and duration of the market. Their analysis of 339 Canadian corporate bankruptcies suggests that the failure among younger companies can be attributed to a lack of management knowledge and financial management skills. On the other hand, the failure among older companies can be attributed to the inability to adapt to changes in the environment.

3. Methodology

Fisher's factorial test makes it possible to verify the hypothesis of independence even with small numbers. The probability of getting the frequencies $n_{11}, n_{12}, n_{.21}, n_{.22}$ for a given n is equal to:

$$P(n_{11,}n_{12},n_{21},n_{22}) = \frac{n!}{n_{11}!\,n_{12}!\,n_{21}!\,n_{22}!}p_{11}^{n_{11}}p_{12}^{n_{12}}p_{21}^{n_{21}}p_{22}^{n_{22}} \tag{1}$$

Identified as:

$$Q = p_{1.}^{n_1}p_{2.}^{n_2}p_{.1}^{n_1}p_{.2}^{n_2} \tag{2}$$

Assuming independence, we have$p_{jk} = p_{j.}p_{k}(j,k = 1{,}2)$. After simple correction we get:

$$P(n_{11,}n_{12},n_{21},n_{22}) = \frac{n!}{n_{11}!\,n_{12}!\,n_{21}!\,n_{22}!}Q \tag{3}$$

The probability of creating a table with marginal frequencies$n_1, n_2, n_{.1}, n_{.2}$ is equal to:

$$\sum_{i=\max(o,n_{.1}-n_2)}^{\min(n_1,n_{.1})} P(i, n_{.1} - i, i + n_2 - n_{.1}) =$$

$$= \sum_{i=\max(o,n_{.1}-n_2)}^{\min(n_1,n_{.1})} \frac{n!}{i!\,(n_1-i)!\,(n_{.1}-i)!\,(i+n_2-n_{.1})!}Q = Q\frac{n!}{n_1!\,n_2!}\sum_{i=\max(o,n_{.1}-n_2)}^{\min(n_1,n_{.1})}\binom{n_1}{i}\binom{n_2}{n_{.1}-i} =$$

$$= Q\frac{n!}{n_1!\,n_2!}\binom{n_1+n_2}{n_{.1}} = \frac{(n!)^2}{n_1!\,n_2!\,n_{.1}!\,n_{.2}!} \tag{4}$$

The conditional probability *P* that for a given marginal frequencies $n_1, n_2, n_{.1}, n_{.2}$a table with the frequencies $n_{11}, n_{12}, n_{21}, n_{22}$is formed, is equal to the ratio of expressions (3) and (4) . That's how we get:

$$P = \frac{n_1!\,n_2!\,n_{.1}!\,n_{.2}!}{n!\,n_{11}!\,n_{12}!\,n_{21}!\,n_{22}!} \tag{5}$$

However, we must calculate the probabilities P according to formula (5) not only for the given table, but also for all others that arise from it by gradually reducing the smallest number by one while maintaining marginal numbers. The individual probabilities *P* add up. If their sum is not greater than given α, we reject the hypothesis of level independence at thelevel of α. This test is one-sided because it investigates possible dependence in only one direction. (Martin and Bridgmon, 2012, Hanggraeni, 2019)

The construction of a two-sided test is much more complicated. However, because this test is very often used in practice, we will at least briefly describe its most important variants. We will find the smallest number in the given four-field table. And then we compare the marginal numbers of row and column, which intersect in the field with the smallest number found. We will choose the one with the smallest marginal number and call it a line. In the next step, however, the individual variants differ. First, we reduce the minimum number in the table gradually by one unit, while maintaining marginal numbers as well as in the one-sided test. Then we replace the original numbers in the line and again reduce the smaller one by one, while maintaining the marginal numbers of the default table. If the sum of the probabilities *P* of all the resulting tables does not exceed the number αwe reject H_0. (Overall and Starbuck, 1983)

 ISSN 1337-0839 (print) / 2585-7258(online)

Fisher's test differs from other tests in that it does not define any test statistic, the measured value of which is compared with the quantile of the relevant distribution, but that it decides to reject the null hypothesis due to the p-value, i.e. we reject the null hypothesis if $p \leq \alpha$, where α is a predetermined test level. How the p-value is calculated is determined by whether it is an H_0 test against a one-sided alternative or a two-sided alternative. (Young, 2019)

4. Results

A total of 286 companies from the Czech Republic were surveyed, of which 143 were micro-enterprises (enterprises with up to 5 employees) and another 142 were SME´s (from 5 employees to 100). In Fisher's exact test, we examined whether there was a statistically significant relationship between the size of the company and their solvency.

We therefore examined the null hypothesis:

H0: There is no statistically significant relationship between the size of the enterprise and its solvency.

An alternative to this hypothesis is hypothesis H1.

H1: There is a statistically significant relationship between the size of the enterprise and its solvency.

According to Table 1, enterprises belonging to micro-enterprises by size had a higher tendency to default. Specifically, up to 29.4% of micro-enterprises went bankrupt. While for SMEs, it is 10.5% of the sample examined.

Table 1: Default Crosstabulation

Type * Default Crosstabulation

			Default		Total
			no	yes	
Type	Micro	Count	101	42	143
		% within Type	70.6%	29.4%	100.0%
		% within Default	44.1%	73.7%	50.0%
		% of Total	35.3%	14.7%	50.0%
	MSP	Count	128	15	143
		% within Type	89.5%	10.5%	100.0%
		% within Default	55.9%	26.3%	50.0%
		% of Total	44.8%	5.2%	50.0%
Total		Count	229	57	286
		% within Type	80.1%	19.9%	100.0%
		% within Default	100.0%	100.0%	100.0%
		% of Total	80.1%	19.9%	100.0%

Source: Own output

To determine statistical significance, we have to look at the Table 2. Bellow the Table 2 there is note that 0 cells have expected count less than 5. Based on Pearson's Chi-Square, at an $\alpha = 0.05$, Pearson's Chi-Square is statistically significant at 0.001. Looking at Fisher's exact test is also statistically significant, as it is 0.002. In this case, the null hypothesis is rejected, and the alternative is accepted that there is a statistically significant relationship between the size of the enterprise and its solvency.

Table 2: Chi-Square Tests

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10.700[a]	1	.001	.002	.001
Continuity Correction[b]	9.535	1	.002		
Likelihood Ratio	10.286	1	.001	.002	.001
Fisher's Exact Test				.002	.001
N of Valid Cases	172				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 15.07.

Source: Own output

Lambda is a measure of association and compares the symmetrical ratio of Phi and Cramer V. In the case of Lambda (see Table 3), it is essential which variable we solve as dependent and independent. At the same time, Lamba measures the percentage of explained variance in you dependent by independence. The dependent is "type" (micro or MSP), and the value here is at level 0, it is significant.

Table 3: Directional Measures

Directional Measures

			Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance	Exact Significance
Nominal by Nominal	Lambda	Symmetric	.000	.068	.000	1.000	
		Type Dependent	.000	.128	.000	1.000	
		Default Dependent	.000	.000	.[c]	.[c]	
	Goodman and Kruskal tau	Type Dependent	.062	.039		.001[d]	.002
		Default Dependent	.062	.039		.001[d]	.002
	Uncertainty Coefficient	Symmetric	.049	.031	1.593	.001[e]	.002
		Type Dependent	.048	.030	1.593	.001[e]	.002
		Default Dependent	.051	.032	1.593	.001[e]	.002

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Cannot be computed because the asymptotic standard error equals zero.
d. Based on chi-square approximation

e. Likelihood ratio chi-square probability.

Source: Own output

Table 4 shows the value of how dependent the values are. The value of Phi is 0.249, which is a weak correlation and is significant.

Table 4: Symmetric Measures

Symmetric Measures				
		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.249	.001	.002
	Cramer's V	.249	.001	.002
	Contingency Coefficient	.242	.001	.002
N of Valid Cases		172		

Source: Own output

5. Discussions

In this study, we focused on the relationship between the size of the company and its solvency. By analyzing other studies, we found that there are two camps. On the one hand, according to some authors, micro-enterprises are more likely to be a banker because they cannot respond to changes in the environment as effectively as larger ones. At the same time, they usually do not hold as much capital as larger companies.

On the other hand, some authors assume that small companies have a lower share of external resources in the structure of liabilities than large companies. Cassar and Holmes (2003) addressed the availability and cost of specific forms of external funding in their research. These authors concluded that small companies would receive less foreign capital or obtain foreign capital at higher costs than large companies, which may be reflected in their lack of interest in external financing. Erdem et al., (1996) argues that the reason for lower indebtedness of small businesses may be, among other things, their efforts to increase liquidity in times of financial difficulties. The authors Rajan and Zingales (1995) and Frank and Goyal (2003) also commented on this issue in their research papers. These authors studied the relationship between the size of the company and its indebtedness of long-term liabilities, banks, and other external financing. They concluded that there was a significant positive relationship between the extent of the company and its debt. The larger the company, the higher its tendency to finance its activities with external sources. In their work, Titman and Wessels (1988) conclude that large companies are more indebted, which they say is due to better debt diversification and thus less risk of corporate bankruptcy than smaller companies. Banking institutions consider such companies to be less risky, and therefore their availability for credit is generally better.

It is, therefore, essential to address this issue to reach a particular conclusion. However, it is complicated to approach the evaluation of companies as a whole, as each company is specific, and it is difficult to conclude.

6. Conclusion

Fisher's exact test is an inferential statistical procedure for comparing the number of people or things that fall into different categories. Using Fisher's exact test in a four-field pivot table, we verified the dependence of four pairs of variables, which acquired only two variants. The limit of this study and consequently of the possible interpretation of the results is a small set of 284 companies from various regions of the Czech Republic, so it cannot fully represent all companies in the Czech Republic. Some of the main causes of bankruptcy are poor planning, lack of funds, lack of business experience and lack of personal discipline. Anyone considering opening their own business should have not only an idea of where the business will end up, but also a strong business plan. An important part of this business plan should be the exit strategy in case of business problems. Not all small business failures should be automatically classified as negative. Some of them are designed to exist for a specified period of time. At the same time, the closure of some small businesses may be considered a success by the owner when it moves to other areas of life.

References

Altman, E. I., Haldeman, R. G. and Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, *1*(1), pp. 29-54.

Bilder, C. R. and Loughin, T. M. (2007). Modeling association between two or more categorical variables that allow for multiple category choices. *Communications in Statistics-Theory and Methods*, *36*(2), pp. 433-451.

Blazy, R. and Letaief, A. (2017). When secured and unsecured creditors recover the same: The emblematic case of the Tunisian corporate bankruptcies. *Emerging Markets Review*, *30*, pp. 19-41.

Bradley, D. and Cowdery, C. (2004). Small business: causes of bankruptcy. SBANC: Small Business Advancement National Center. University of Central Arkansas.

Bradley, D. B. And Rubach, M. J. (2002). Trade Credit and Small Businesses: A cause of business failures. *University of Central Arkansas*.

Carter, R. and Auken, H. V. (2006). Small firm bankruptcy. *Journal of Small Business Management*, *44*(4), pp. 493-512.

Cassar, G. and Holmes, S. (2003). Capital structure and financing of SMEs: Australian evidence. *Accounting & Finance*, *43*(2), pp. 123-147.

Deleryd, M., Garvare, R. and Klefsjö, B. (1999). Experiences of implementing statistical methods in small enterprises. *The TQM Magazine*.

Duffie, D. and Singleton, K. J. (2012). *Credit risk: pricing, measurement, and management*. Princeton university press.

Erdem, A. T., Sezan, M. I. And Ozkan, M. K. (1996). U.S. Patent No. 5,550,935. Washington, DC: U.S. Patent and Trademark Office.

Frank, M. Z. and Goyal, V. K. (2003). Testing the pecking order theory of capital structure. *Journal of Financial Economics*, *67*(2), pp. 217-248.

Fu, D., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K. and Stanley, H. E. (2005). The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences*, *102*(52), pp. 18801-18806.

Fujiwara, Y. (2004). Zipf law in firms bankruptcy. *Physica A: Statistical Mechanics and its Applications*, *337*(1-2), pp. 219-230.

Hanggraeni, D., Ślusarczyk, B., Sulung, L.A.K., Subroto, A. (2019). The impact of internal, external and enterprise risk management on the performance of micro, small and medium enterprises. *Sustainability*, *11*(7), art. no. 2172.

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... and Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj*, *343*, d5928.

Kamei, H., Takayasu, H., Kabashima, Y. and Takayasu, M. (2017). Bankruptcy Prediction with Interfirm Network Structure. *Proceedings of the Asia-Pacific Econophysics Conference 2016---Big Data Analysis and Modeling Toward Super Smart Society---(APEC-SSS2016), id. 011013*, 11 pp. (No. 1).

Lando, D. (2009). Credit risk modeling. *Handbook of Financial Time Series,* pp. 787-798, Springer, Berlin, Heidelberg.

Martin, W. E. and Bridgmon, K. D. (2012). Quantitative and statistical research methods: From hypothesis to results (Vol. 42). *John Wiley & Sons*.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, pp. 109-131.

Overall, J. E. and Starbuck, R. R. (1983). F-test alternatives to Fisher's exact test and to the chi-square test of homogeneity in 2 × 2 tables. *Journal of Educational Statistics*, *8*(1), pp. 59-73.

Podobnik, B., Horvatic, D., Petersen, A. M., Urošević, B. and Stanley, H. E. (2010). Bankruptcy risk model and empirical tests. *Proceedings of the National Academy of Sciences*, *107*(43), pp. 18325-18330.

Poon, A., Jankly, S. and Chen, T. (2018, December). Privacy preserving Fisher's exact test on genomic data. *2018 IEEE International Conference on Big Data (Big Data)*, pp. 2546-2553.

Scherbaum, C. A. and Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, *12*(2), pp. 347-367.

Ślusarczyk, B. and Grondys, K. (2019). Parametric conditions of high financial risk in the SME sector. *Risks*, *7*(3), art. no. 84.

Titman, S. and Wessels, R. (1988). The determinants of capital structure choice. *The Journal of Finance*, *43*(1), pp. 1-19.

Thompson, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, *80*(1), pp. 64-71.

Thornhill, S. and Amit, R. (2003). Learning about failure: Bankruptcy, firm age, and the resource-based view. *Organization science*, *14*(5), pp. 497-509.

Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, *134*(2), pp. 557-598.

Zarrin, S. and Lim, T. J. (2009). Composite hypothesis testing for cooperative spectrum sensing in cognitive radio. *2009 IEEE International Conference on Communications*, pp. 1-5.