

Primary analysis of disclosure risk in tabular data from a Brazilian economic survey

Sâmela Batista Arantes and Maysa S. de Magalhães (Brazilian Institute of Geography and Statistics, Brazil)

samela.arantes@ibge.gov.br, maysa.magalhaes@ibge.gov.br

Abstract and Paper

The Innovation Survey (IS), realized by the Brazilian Institute of Geography and Statistics (IBGE), is an economic survey carried out by sampling, in every three years, raising information for a three-year period (triennium), which produces information on characteristics such as investment on innovative activities; sources of funding for these investments; impact of innovations on company performance. The last dissemination refers to the triennium 2012-2014. This survey has some quantitative variables, among them the investment in innovative activities. This variable is considered the most sensitive in the scope of the survey. The protection of confidentiality of the companies is one of the institute's legal obligations. The results of the survey are published in tables. This work has the objective of evaluating the risk of disclosure of companies through the analysis of disclosure risk in the disseminated tables. This analysis will be done through one of the concentration rules, the rule $p\%$, considering the sample weights. This concentration rule is used by several countries and is recommended in the literature. Different values of p will be used to verify the number of sensitive cells.

Primary analysis of disclosure risk in tabular data from a Brazilian economic survey

Sâmela Batista Arantes* and Maysa S. de Magalhães**

* Brazilian Institute of Geography and Statistics, samela.arantes@ibge.gov.br

** Brazilian Institute of Geography and Statistics, maysa.magalhaes@ibge.gov.br

Abstract: The Innovation Survey (IS), realized by the Brazilian Institute of Geography and Statistics (IBGE), is an economic survey carried out by sampling, raising information for a three-year period (triennium), which produces information on characteristics such as investment on innovative activities; sources of funding for these investments; impact of innovations on company performance. Its last dissemination refers to the triennium 2012-2014. This survey has some quantitative variables, among them the investment in innovative activities. This quantitative variable is considered the most disaggregated in the scope of the survey, since it's considered different types of investments. The results of the survey are published in tables. This work has the objective to perform a sensitivity analysis of investment variables with respect to changes in the value of parameter p of the $p\%$ rule considering the table of magnitude studied that contains the types of investments in innovation activities by economic activity. This study considers the sample weights. This concentration rule is used by several National Statistics Offices and is recommended in the literature.

1 Introduction

Statistical Disclosure Control (SDC) is a field of study that has gained importance in the National Statistics Offices (NSOs) in recent decades. For this reason, several NSOs, and the Brazilian Institute of Geography and Statistics - IBGE, have invested in improvements to ensure the protection of the identity of the respondents, whether individuals, companies or other organizations providing information. Economic entities are also respondent units whose confidential variables are protected. As described in O'Keefe and Shlomo (2012), the economic surveys have some characteristics that differ from household surveys. These characteristics make it difficult to disseminate microdata, as they facilitate to disseminate of confidential information. An example of such characteristics is the asymmetry of quantitative variables, such as revenues and investments, due to differences between companies. For this reason, several NSOs, like IBGE, do not make microdata of economic survey available to the public. Consequently, the tables are the main form of dissemination from IBGE economic data and others NSOs.

One of the economic surveys conducted by the Institute is the Industrial Technological Innovation Survey (PINTEC), from now on called Innovation Survey. This survey investigate the types of innovation proposed and developed by Brazilian companies

related to extractive industries, manufacturing, electricity and gas and some selected services in the survey. Some of the variables of this survey are quantitative variables, such as the company's net revenues and investment in innovation activities. These variables have asymmetric distributions due to differences in size and production between companies. Thus, large companies represent the highest percentages of total quantitative variables when the cells of a magnitude table are disclosed.

Investment in innovation activities is considered the most detailed quantitative variable of the Innovation Survey. The most disaggregated information of investments in innovation activities is a table of magnitude that contains these investments by economic activities. This table consists of 69 rows corresponding to economic activities and 8 columns corresponding to 8 types of investments in innovation activities.

The main objective of this study is to perform a sensitivity analysis of investment variables in relation to the changes in the value of parameter p of the $p\%$ rule, considering the magnitude table that contains the types of investments in innovation activities by economic activity.

The data source used is the Innovation Survey referred to the three-year period 2012-2014. In section 2 is described the data source utilized. In section 3 is presented the $p\%$ rule. Section 4 is devoted to sensitivity analysis of the investment variables considering $p\%$ rule. Finally, the final remarks are provided in section 5.

2 Data source

The Innovation Survey is a survey on innovation practices in Brazilian companies which are selected by sampling. The population of interest are the companies whose main economic activity belongs to any of the industries or services listed below: extractive industry, manufacturing industry, electricity and gas and some services such as telecommunications, information technology service activities and architectural services and engineering. Economic activities are presented in the Appendix.

Other requirements of the participating companies in the survey are (IBGE, 2016):

- To be headquartered in the national territory;
- Have 10 or more persons employed by December 31 of the base year of the basic survey selection record and
- Be legally organized as a business entity.

The survey is conducted every three years and the latest dissemination refers to the triennium period 2012-2014. The sample is selected from a business register called Central Registry of Enterprises (CEMPRE). The selected sample is probabilistic and stratified by company size and by main economic activity. The company's size measurement is the number of employees, and there is a stratum where the largest companies are selected with probability 1 for the sample. Other details of the sample design can be found in IBGE (2016).

The topics covered in the survey questionnaire refer to the types of innovations, such as innovations in product and process. Innovation activities are classified into two groups: Research and Development activities (R & D) (basic research, applied or experimental development) and non R & D activities (involving the procurement of goods, services and external knowledge). In addition to the items related to innovation activities, the Innovation Survey questionnaire also investigates other details of businesses as revenues and costs, for example.

The results of the Innovation Survey refer in part to quantitative variables such as net revenues, investment in innovation activities and number of employed persons. Investment in innovation activities is disclosed in more detail than the other quantitative variables and for this reason it was chosen for this study. The types of investments correspond to variables with a high concentration degree in a few companies, which is an intrinsic characteristic of these variables. In Innovation Survey, investments in innovation activities are classified into 8 types according to the Institute (IBGE, 2016), as described below:

- **Investment in internal Research and Development activities (Inv1):** The internal R & D comprise creative work activities undertaken systematically, in order to increase the stock of knowledge and use this knowledge to develop new applications, such as new or technologically improved products or processes;
- **Investment in external acquisition of Research and Development (Inv2):** External R & D include the activities described above, carried out by another organization (business or technological institutions) and acquired by the company;
- **Investment in acquisition of other external knowledge (Inv3):** includes technology transfer agreements arising from the purchase of patent royalties license and the use of trademarks, acquisition of know-how and other technical and scientific knowledge of third parties for the company to develop and implement innovations;
- **Investment in software acquisition (Inv4):** comprises the acquisition of software (design, engineering, processing and transmission of data, voice, graphics, videos, process automation, etc.), specifically purchased for the implementation of new technologically improved products or processes. It does not include those registered in internal R & D;
- **Investment in the purchase of machinery and equipment (Inv5):** Includes the acquisition of machinery, equipment, hardware, specifically purchased for the implementation of new or technologically improved products or processes;
- **Investment in training (Inv6):** Comprises oriented training to the development of products or processes which are technologically new or significantly

improved and related to the company innovative activities that may include acquisition of external technical services;

- **Investment in the introduction of technological innovations in the market (Inv7):** Includes marketing activities directly related to the launch of technologically new or improved product, which may include: market research, market testing and advertising for the launch. This does not include the construction of market distribution networks for the innovations;
- **Investment in industrial design and other technical preparations (Inv8):** Refers to the procedures and technical preparation to effect implementation of product or process innovations. It includes oriented plans and drawings to define procedures, technical specifications and operational features necessary to the implementation of process or product innovations. Includes changes in production procedures and quality control, working methods and standards and software required to implement product or technologically new or improved processes, as well as basic industrial technology activities (metrology, standardization and conformity assessment), the trials and tests (which are not included in R & D) for final registration of the product and the effective start of production.

The Appendix describes economic activities and its related sub-activities. Economic activities are classified into four groups: extractive industries, manufacturing, electricity and gas and selected services as shown in the Appendix.

3 The p% rule

To assess whether a magnitude table needs protection, the first step is to analyse each cell individually. This requires the primary methods of risk assessment. In this case the table is classified as safe or unsafe according to the protection level of its cells. The p% rule is one of the concentration rules used to analyse each cell of the magnitude table.

There are other concentration rules to assess the disclosure risk of cells, such as the dominance rule and the pq-rule (Cox, 2001; Loeve, 2001; O'Malley and Ernst, 2007). However, the p% rule has been the most recommended in the literature (Jewett, 1993; Zayatz, 2007; Wolf and Hundepool 2012).

It is important to highlight that no risk assessment rule is considered unanimous among the NSOs. One of the most widely used rules for analysing table cells is the minimum frequency rule, ie cells are considered sensitive when they do not have a minimum number of respondents.

As the p% rule will be the concentration rule used in this study, it is defined as follows:

p% Rule: Let X be the total of one of the cells of the magnitude table considered. Let X_1 and X_2 be, respectively, the first and second largest contributions of the analysed cell. The cell is considered sensitive if the total (X) minus the first and second largest contributions (X_1 and X_2) is less than $p\%$ of the largest contribution, that is, $X - X_2 - X_1 < (p/100) X_1$.

Hundepool and De Wolf (2012) suggest that NSOs adopt the $p\%$ rule with p in the range of 5% to 15%. In an analysis of the 1992 Economic Census, the US Census Bureau chose the $p\%$ rule and adopted p equal to 15% for making calculations involved in determining sensitive cells (Jewett, 1993). Zayatz (2007) confirms that the US Census Bureau uses the $p\%$ rule, but the p value is not disclosed. In a study applied to economic data of industries in South Korea (Kim, 2009) was adopted p equal to 20%.

According to the studies presented, in this work a comparison of the p values suggested in the literature will be made, analysing the percentage increase of sensitive cells of the investment types considered. The $p\%$ rule will be used for the following p values: 5%, 10%, 15%, 20% and 25%, being $p = 5\%$ the basis for comparison to other p values.

4 Sensitivity analysis of variables

4.1 Description of magnitude table

This section presents the studied magnitude table (Table 1). The notation adopted in Table 1 is shown below:

The eight types of investments were presented in the previous section and denoted by Inv1, Inv2, Inv3, Inv4, Inv5, Inv6, Inv7 and Inv8.

The four major groups of economic activities will be denoted by A.1, A.2, A.3 and A.4, noting that:

- A.1 corresponds to extractive industries;
- A.2 corresponds to the manufacturing industries;
- A.3 corresponds to electricity and gas companies;
- A.4 corresponds to selected services companies.

The presentation of economic activities in detail as well as the notation of each activity within each economic activity group is in the Appendix. Table 1 is a magnitude table, and the concept of magnitude table is presented below:

Magnitude table: Tables that present in each cell sums of observations of quantitative variables where each sum represents a group of observations defined by grouping (spanning) variables that classify a set of respondents (Hundepool *et al.*, 2010).

Respondents are typically companies but can be individuals or households. The variables used to build the crossings in the tables are called spanning, grouping or

classification variables. The rows and columns names correspond to the classifications or categories of these variables.

In this study, the classification variables consist of the economic activities (rows) and the types of investment (columns). Each cell displays the sum of a quantitative variable, also called response variable of magnitude table, which here consists of companies' investments in each crossing of the classification variables.

The magnitude table (Table 1) under analysis is a hierarchical table, that is the classification variable “Economic activity” contains classes or categories which contain subcategories, such as: Within activity A.2 there are Activity A.2.8 “Manufacture of pulp, paper and paper products” and within it Activity A.2.8.1 “Manufacture of pulp and other pulp” and Activity A.2.8.2 “Manufacture of paper, packaging and artifacts of paper”.

Each cell in Table 1 contains the sums of the eight types of investments in innovation activities according to the economic activities of the surveyed companies. The four major groups of economic activities are highlighted in bold. The missing values represented by “(x)” in Table 1 refer to values not disclosed by the IBGE for cell protection.

Economic activities¹	Inv1	Inv2	Inv3	Inv4	Inv5	Inv6	Inv7	Inv8	Total
A.1	611.399	48.066	(x)	25.346	900.740	8.191	(x)	(x)	1.746.578
A.2	17.560.176	4.535.764	2.381.504	1.263.395	22.267.198	583.353	3.304.743	3.995.625	55.891.758
A2.1	776.246	164.321	76.847	182.526	4.228.850	107.646	1.392.673	177.406	7.106.516
A2.2	62.540	7.093	23.157	6.500	1.353.779	10.039	60.539	629.124	2.152.772
A2.3	75.430	0	(x)	2.688	28.354	1.687	(x)	59.921	170.408
A2.4	59.212	12.869	26.265	26.261	534.398	14.102	30.420	57.523	761.050
A2.5	105.549	4.245	14.155	126.035	348.915	25.274	127.754	53.626	805.552
A2.6	264.533	14.051	21.040	21.325	208.040	5.306	38.273	101.134	673.702
A2.7	48.121	(x)	(x)	14.007	428.186	6.311	11.988	116.459	630.051
A2.8	275.760	30.994	4.323	21.258	755.241	8.482	61.507	61.466	1.219.031
A2.8.1	48.695	15.471	(x)	1.152	36.921	(x)	(x)	(x)	106.239
A2.8.2	227.065	15.523	(x)	20.107	718.320	(x)	(x)	(x)	1.112.792
A.2.9	55.991	(x)	23.183	18.621	277.925	10.875	(x)	45.920	444.443
A.2.10	2.665.346	731.284	4.315	21.839	1.216.452	12.165	10.357	162.151	4.823.909
A.2.10.1	44.716	6.790	4.315	7.682	902.829	4.673	6.126	60.673	1.037.806

¹ The description of the economic activity codes is in the Appendix.

A.2.10.2	2.620.630	724.493	0	14.158	313.623	7.492	4.231	101.478	3.786.104
A.2.11	1.966.468	359.557	374.181	92.912	1.278.592	46.530	406.653	224.052	4.748.945
A.2.11.1	119.798	21.402	27.235	7.366	473.453	6.501	39.741	10.785	706.282
A.2.11.2	370.714	27.166	102.938	20.565	252.097	3.343	17.567	22.953	817.343
A.2.11.3	701.542	80.547	16.746	4.621	124.173	8.032	42.799	52.456	1.030.917
A.2.11.4	457.316	209.109	223.287	35.269	231.497	16.044	264.630	111.901	1.549.052
A.2.11.5	317.098	21.333	3.975	25.090	197.373	12.609	41.916	25.957	645.351
A.2.12	1.228.473	213.481	58.761	40.375	350.915	39.283	235.867	114.469	2.281.624
A.2.12.1	3.399	(x)	(x)	12.296	123.130	307	4.391	25.325	193.589
A.2.12.2	1.225.074	(x)	(x)	28.079	227.785	38.976	231.476	89.144	2.088.035
A.2.13	467.323	28.918	220.583	36.886	1.081.198	32.567	78.085	140.786	2.086.346
A.2.14	295.289	82.610	9.833	61.177	1.376.071	24.878	56.070	463.446	2.369.375
A.2.15	558.020	42.992	34.745	106.370	1.684.621	20.124	30.599	35.675	2.513.146
A.2.15.1	351.817	34.674	7.536	89.417	1.343.939	17.177	20.265	16.239	1.881.065
A.2.15.2	206.203	8.319	27.209	16.952	340.682	2.946	10.334	19.436	632.081
A.2.16	206.135	28.983	18.174	67.381	1.634.024	22.127	41.742	73.903	2.092.468
A.2.17	1.555.772	666.908	24.976	46.048	359.881	30.403	67.113	44.671	2.795.774
A.2.17.1	74.656	12.908	7.101	8.329	114.767	2.182	3.070	5.923	228.936
A.2.17.2	298.778	129.825	4.162	5.352	53.693	3.356	27.331	20.340	542.839
A.2.17.3	839.066	488.571	9.728	22.634	157.048	17.502	18.036	9.865	1.562.449
A.2.17.4	37.396	6.372	877	712	8.292	425	6.900	1.403	62.378
A.2.17.5	305.877	29.233	3.109	9.020	26.079	6.939	11.776	7.140	399.172
A.2.18	1.367.967	48.204	20.818	22.730	482.210	15.418	37.218	184.762	2.179.327
A.2.18.1	1.049.778	25.486	17.967	8.587	203.880	7.345	25.715	147.682	1.486.442
A.2.18.2	188.546	3.710	328	7.738	84.711	2.496	5.676	33.848	327.054
A.2.18.3	129.642	19.007	2.523	6.405	193.619	5.576	5.828	3.232	365.832
A.2.19	1.041.178	74.514	53.600	132.541	1.133.964	42.388	132.060	225.472	2.835.718
A.2.19.1	166.681	1.750	2.506	6.913	127.874	4.239	15.638	20.482	346.082
A.2.19.2	307.687	5.835	6.918	27.095	187.470	11.487	28.103	105.487	680.082
A.2.19.3	144.533	794	5.320	8.483	279.671	4.126	27.643	9.197	479.767
A.2.19.4	422.277	66.136	38.856	90.050	538.950	22.537	60.676	90.306	1.329.787
A.2.20	2.913.239	242.534	319.354	66.395	1.799.860	37.388	319.414	562.129	6.260.313
A.2.20.1	1.907.944	135.992	233.681	21.638	781.688	2.427	230.783	380.612	3.694.765
A.2.20.2	130.400	9.990	2.250	6.092	67.297	2.531	3.706	4.686	226.953
A.2.20.3	874.895	96.552	83.423	38.665	950.876	32.429	84.925	176.831	2.338.596
A.2.21	1.122.820	1.640.470	1.017.357	52.088	804.222	29.680	65.554	345.500	5.077.692
A.2.22	140.393	29.932	9.848	64.371	466.219	13.051	35.153	38.680	797.647
A.2.23	128.281	10.934	17.411	23.248	279.065	8.987	36.653	39.727	544.305
A.2.23.1	95.050	10.055	2.286	10.663	113.271	3.716	18.506	11.607	265.153

A.2.23.2	33.231	879	15.125	12.585	165.794	5.271	18.147	28.120	279.153
A.2.24	180.089	93.532	3.735	9.813	156.216	18.643	21.991	37.624	521.642
A.3	348.602	533.841	(x)	29.008	180.404	16.135	(x)	(x)	1.161.401
A.4	6.182.297	3.776.367	273.898	841.630	10.156.098	200.989	883.136	377.493	22.691.909
A.4.1	38.378	(x)	38.065	25.242	124.330	(x)	47.193	8.928	294.350
A.4.2	503.121	3.699.685	(x)	542.516	9.284.179	101.419	665.597	(x)	14.809.133
A.4.3	1.612.543	37.336	56.023	102.176	425.546	48.542	136.701	332.735	2.751.602
A.4.3.1	309.272	3.534	22.564	21.583	20.410	17.994	23.255	11.803	430.416
A.4.3.2	477.394	6.804	15.433	15.742	323.384	8.634	18.437	282.434	1.148.263
A.4.3.3	454.963	2.145	10.820	4.958	16.421	3.926	67.689	15.710	576.632
A.4.3.4	370.914	24.854	7.206	59.892	65.331	17.988	27.320	22.787	596.291
A.4.4	278.996	17.652	(x)	104.498	87.536	16.570	(x)	18.000	541.019
A.4.5	224.931	11.232	153.271	65.414	225.241	27.276	17.647	11.012	736.024
A.4.6	3.524.329	(x)	16.939	1.785	9.265	(x)	(x)	(x)	3.559.781
Total	24.702.474	8.894.039	2.743.638	2.159.379	33.504.440	808.668	4.198.500	4.480.508	81.491.645

Table 1: Investments (R\$ 1000) in innovation activities by economic activity of companies in the Innovation Survey, 2012-2014, Source: IBGE, 2016

4.2 Sensitivity analysis of investment variables

In this section we will analyse the sensitivity of investment variables with respect to changes in the value of parameter p of the $p\%$ rule. Cells considered sensitive in the performed analysis will not be highlighted for confidentiality reasons.

The analysis of the cells in Table 1 was performed with the *sdcTable* package (version 0.28, Meindl, 2019) of the software R, which is a free software. This package contains tools for performing primary analysis of disclosure risk taking into account the sample weights of surveys realized by sampling, such as the Innovation Survey. The *primarySuppression* function was used to apply the $p\%$ rule. The sample weights have been rounded to the nearest integer because the package uses integer weights.

Figure 1 shows the percentage variation in the number of sensitive cells with respect to parameter p equal to 5%, since this value is the smallest value suggested in the literature.

When we vary the value of p , the number of sensitive cells is expected to increase when the p value increases, as it becomes more likely that $X - X_1 - X_2$ be less than $(p/100) X_1$, according to the definition of the $p\%$ rule presented in section 3.

By Figure 1, it can be observed that the largest increase in the number of sensitive cells occurred when the p value was changed from 15% to 20% when compared to $p = 5\%$. The Inv6 (Investment in training) variable presented the sharpest percentage increase in the number of sensitive cells. This variable had a 300% increase for $p = 20\%$ when

compared to $p = 5\%$, ie when we varied the value of $p = 5\%$ to $p = 20\%$, this variable had the number of sensitive cells multiplied by 4.

Inv1 (Investment in internal activities of Research and Development) was the variable with the smallest variation in the number of sensitive cells when we varied the p value compared to $p = 5\%$. When p varies from 5% to 10% or from 5% to 15% there is no increase in the number of sensitive cells. The increase in the number of sensitive cells occurred when p equals 25%, in this case the percentage increase was 25% compared to p equals 5%.

The variables Inv3 (Investment in the acquisition of other external knowledge) and Inv8 (Investment in industrial design and other technical preparations) showed similar behaviour when the p value was varied. Both are among the variables with the least variation in the number of sensitive cells when the value of parameter p is varied.

The variable Inv2 (Investment in external acquisition of research and development) presented median variations when compared to the other variables. The percentage increase in the number of sensitive cells when p equals 20% compared to p equals 5% is 100% and 125% when $p = 25\%$ compared to p equals 5%.

The three variables with the largest percentage increases when p equals 20% are Inv6 (Investment in training), Inv5 (Investment in machine and equipment acquisition), and Inv4 (Software acquisition investment), however they did not increase the number of sensitive cells when p varied from 20% to 25% always comparing with p equal to 5%.

It is noteworthy that when $p = 5\%$, at least one sensitive cell was found for all types of investments, making it possible to calculate the percentage increase in sensitive cells when the p value is greater than 5%.

It is important to highlight that the Figure 1 does not show the concentration level of the variables in the cells of Table 1. The objective was to show the percentage variation in the number of sensitive cells when p varies and thus help in the choice of parameter p if this rule were adopted.

Again, it is important to point out that cells classified as sensitive were not specified to protect them from the disclosure risk, and not even the number of sensitive cells, as this is a preliminary study that requires additional studies for further conclusions. Some authors emphasize the fact that the dissemination of the parameters used in the concentration rules facilitates the disclosure of individual information (Hundepool *et al.*, 2010, 2012).

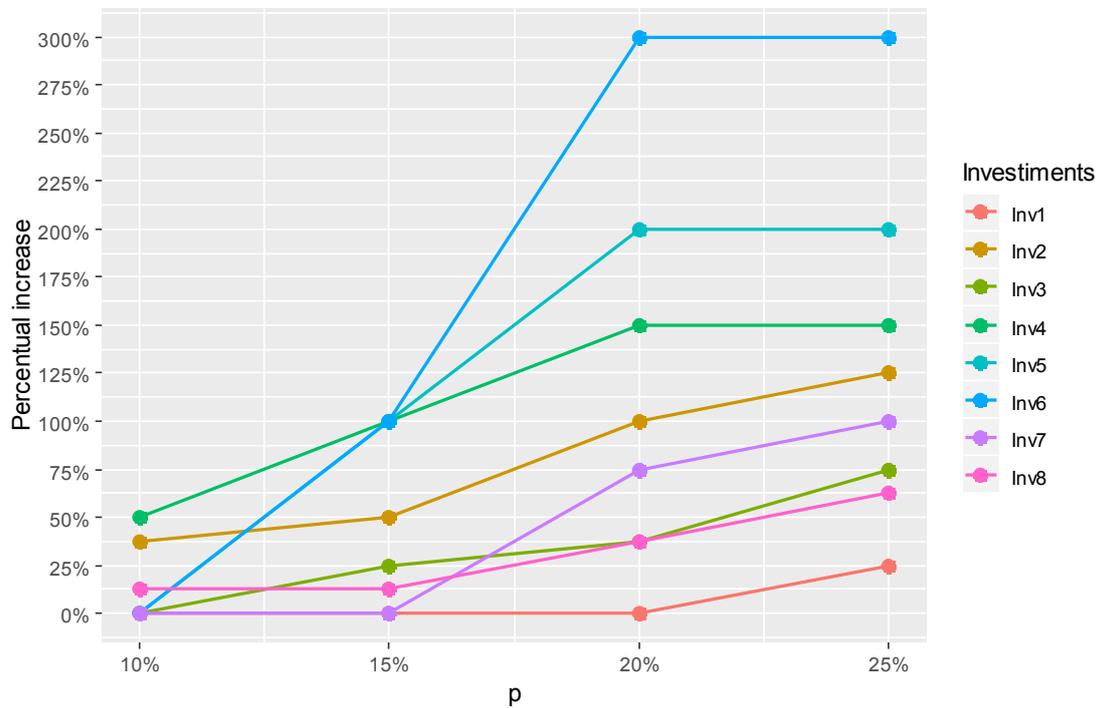


Figure 1: Percentual increases in the number of primary sensitive cells obtained with the use of p% rule, considering p = 5% as a comparison basis for Table 1

5 Final remarks

Tables are the main products of economic surveys in the Brazilian Institute of Geography and Statistics. Some of these tables are magnitude tables such as Table 1. Primary analysis of the disclosure risk of confidential information in magnitude tables can be made by rules available in the literature, such as the so-called concentration rules. The p% rule is one of these procedures of risk assessment and has been the most recommended in the literature.

Analysing the percentage increase of the number of sensitive cells in the magnitude table (Table 1), which compares these percentage increases in the eight types of investments considered, using the p% rule, based on p equal to 5%, it is observed that the variables presented different levels of percentage increase in the number of sensitive cells when the parameter p was varied. These variations between variables result from the inherent differences between the types of investments that companies in different economic activities apply to innovation activities.

Considering the characteristics of economic surveys, the primary analysis of the disclosure risk in the magnitude table cells is an important phase for the assessment of disclosure risk of confidential company information, such as investments and revenues.

It is also important to point out some characteristics of the chosen variables that refer to investments. These variables represent values that can change significantly over time as investments depend on factors such as economic scenarios, for example. The choice of these variables for the analysis performed here was because they are the most disaggregated quantitative variables of the Innovation Survey.

Further studies should be done before choosing a more definitive risk analysis method to disseminate the magnitude tables of the economic surveys. In addition, future studies will include further analysis of magnitude table cells, adding other variables, secondary risk analysis as well as the study of more advanced methods for protecting magnitude tables.

It is noteworthy that the criteria for analysis of the disclosure risk should be chosen by each NSO, considering the particularities of each country and other aspects such as cultural, for example.

References

- Cox, L. H. (2001). *Disclosure risk for tabular economic data*. In: Doyle, P., Lane, J., Theeuwes, J., Zayatz, L. (eds.) *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, ch. 8, pp. 167–183. Elsevier, New York.
- De Wolf, P. P., Hundepool, A. (2012). *P% should dominate*. Statistics Netherlands.
- Hundepool, A., De Wolf, P. P. (2012) *Statistical disclosure control*. Method Series. Statistics Netherlands.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S.; Lenz, R., Naylor, J.; Nordholt, E.S., Seri, G.; DE Wolf, P.P. (2010). *Handbook on statistical disclosure control*. Version 1.2.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., et al. *Statistical disclosure control*. (2012). Wiley Series in Survey Methodology: Wiley. ISBN 9781118348222.
- IBGE (2016), *Pesquisa Industrial de Inovação Tecnológica 2014 – PINTEC*.
- Jewett R. (1993) *Disclosure analysis for the 1992 economic census*. Technical report. Economic Statistical Methods and Programming Division, Bureau of the Census, Washington, DC.
- Kim, K. (2009). *Examples of application of linear sensitivity rules in Korean statistical data*. United Nations Statistical Commission and Economic Commission for Europe Conference of European Statisticians, Eurostat work session on Statistical Data Confidentiality, Spain.

- Loeve A. (2001). *Notes on sensitivity measures and protection levels*. Technical report, Research paper, Statistics Netherlands.
- Meindl, B. (2019). *Package 'sdcTable'*.
- O'Keefe, C.M., Shlomo, N. (2012). *Comparison of remote analysis with statistical disclosure control for protecting the confidentiality of business data*. Transactions on Data Privacy, Vol. 5.
- O'Malley, M.; Ernst, L. (2007). *Practical considerations in applying the pq-rule for primary disclosure suppressions*. Bureau of Labor Statistics.
- Zayatz, L. (2007). *Disclosure avoidance practices and research at the U. S. Census Bureau: an update*. Journal of Official Statistics 23 Vol. 2.

Appendix: Economic activities that correspond to the lines in Table 1

	Economic Activity
A.1	Extractive Industries
A.2	Manufacturing Industries
A2.1	Manufacture of food products
A2.2	Beverage production
A2.3	Manufacture of smoking products
A2.4	Manufacture of textile products
A2.5	Manufacture of clothing items and accessories
A2.6	Preparation of leather and manufacture of leather goods, travel goods and footwear
A2.7	Manufacture of wood products
A2.8	Manufacture of pulp, paper and paper products
A2.8.1	Manufacture of pulp and other pastes
A2.8.2	Papermaking, packaging and paper artefacts
A.2.9	Printing and reproduction of recorded media
A.2.10	Manufacture of coke, oil products and biofuels
A.2.10.1	Manufacture of coke and biofuels (ethanol and others)
A.2.10.2	Petroleum refining
A.2.11	Manufacture of chemical products
A.2.11.1	Manufacture of inorganic chemical products
A.2.11.2	Manufacture of organic chemical products
A.2.11.3	Manufacture of resins and elastomers, artificial and synthetic fibers, pesticides and household cleaning disinfectants
A.2.11.4	Manufacture of soaps, detergents, cleaning products, cosmetics, perfumery and personal care
A.2.11.5	Manufacture of paints, varnishes, enamels, lacquers, related products and miscellaneous items
A.2.12	Manufacture of pharmaceutical chemicals and pharmaceuticals
A.2.12.1	Manufacturing of pharmaceutical chemical products
A.2.12.2	Manufacture of pharmaceutical products
A.2.13	Manufacture of rubber goods and plastic
A.2.14	Manufacture of non-metallic mineral products
A.2.15	Metallurgy
A.2.15.1	Steel products
A.2.15.2	Metallurgy of non-ferrous metals and metal casting
A.2.16	Manufacture of metal products
A.2.17	Manufacture of computer equipment, electronic and optical products

A.2.17.1	Manufacture of electronic components
A.2.17.2	Manufacture of computer equipment and peripherals
A.2.17.3	Manufacture of communication equipment
A.2.17.4	Manufacture of electromedical and electrotherapeutic apparatus and irradiation equipment
A.2.17.5	Manufacture of other electronic and optical products
A.2.18	Manufacture of machinery, devices and electrical material
A.2.18.1	Manufacture of generators, transformers and equipment for power distribution
A.2.18.2	Manufacture of household appliances
A.2.18.3	Manufacture of batteries, lamps and other electrical appliances
A.2.19	Manufacture of machinery and equipment
A.2.19.1	Motors, pumps, compressors and transmission equipment
A.2.19.2	Machinery and equipment for agriculture
A.2.19.3	Machinery for mining and construction
A.2.19.4	Other machinery and equipment
A.2.20	Manufacture of motor vehicles, trailers and bodyworks
A.2.20.1	Manufacture of automobiles, pickup trucks and utilities, trucks and buses
A.2.20.2	Manufacture of cabins, bodies, trailers and engine reconditioning
A.2.20.3	Manufacture of parts and accessories for motor vehicles
A.2.21	Manufacture of other transport equipment
A.2.22	Manufacture of furniture
A.2.23	Manufacture of various products
A.2.23.1	Manufacture of instruments and materials for medical and dental use and optical articles
A.2.23.2	Other miscellaneous products
A.2.24	Maintenance, repair and installation of machinery and equipment
A.3	Electricity and gas
A.4	Services
A.4.1	Editing and recording and music editing
A.4.2	Telecommunications
A.4.3	Activities of information technology services
A.4.3.1	Development of custom software
A.4.3.2	Development of customizable software
A.4.3.3	Development of non-customizable software
A.4.3.4	Other information technology services
A.4.4	Data processing, hosting on the internet and other related activities
A.4.5	Architecture and engineering services, testing and technical analysis
A.4.6	Research and development