

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

Sâmela Batista Arantes and Maysa S. de Magalhães

October, 2019

The Hague, Netherlands

Contents

- Introduction
- Objective
- Data source
- The $p\%$ rule
- Description of magnitude table
- Sensitivity analysis of variables
- Final remarks

Introduction

- Several National Statistical Offices – NSOs like the Brazilian Institute of Geography and Statistics - IBGE, have invested in improvements to ensure the protection of the identity of the respondents.
- Economic entities are also respondent units whose confidential variables are protected.

Introduction

- Some characteristics of economic surveys make it difficult to disseminate microdata.
 - An example of such characteristics is the asymmetry of quantitative variables.
- The tables are the main form of dissemination from IBGE economic data.

Introduction

- One of the economic surveys conducted by the Institute is the **Innovation Survey** (PINTEC).
- This survey investigates the types of innovation developed by Brazilian companies related to
 - **extractive industries, manufacturing, electricity and gas and some selected services in the survey.**

Introduction

- Investment in innovation activities is considered the most detailed quantitative variable of the Innovation Survey.
- **A magnitude table** disseminates the total of investments in innovation by economic activities.

Objective

The main objective of this study is

- to perform a **sensitivity analysis** of **investment variables** in relation to the changes in the **value of p** parameter of the **p% rule**,
- considering the **magnitude table** that contains the **types of investments** in innovation activities by economic activity.

Data source

- The Innovation Survey is a survey on innovation practices in **Brazilian companies** which **are selected by sampling**.
- The survey is conducted every three years and the **latest dissemination** refers to the triennium period **2012-2014**.

Data source

- The population of interest are the companies whose main economic activity belongs to any of the industries or services:
 - extractive industry
 - manufacturing industry
 - electricity and gas
 - some services such as telecommunications, engineering and information technology, for example,

Data source

- Innovation Survey results 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.
- Investments have high concentration in few companies.

Data source

Investments in innovation activities are classified into 8 types:

- **Inv1:** Investment in **internal Research and Development** activities
- **Inv2:** Investment in **external** acquisition of **Research and Development**
- **Inv3:** Investment in acquisition of other **external knowledge**
- **Inv4:** Investment in **software** acquisition

Data source

Investments in innovation activities are classified into 8 types:

- **Inv5:** Investment in the purchase of machinery and equipment
- **Inv6:** Investment in training
- **Inv7:** Investment in the introduction of technological innovations in the market
- **Inv8:** Investment in industrial design and other technical preparations

The p% rule

- 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 (X_1), that is:

$$X - X_2 - X_1 < (p/100) X_1$$

Description of magnitude table

- The magnitude table (Table 1) under analysis is a **hierarchical table**: the classification variable “**Economic activity**” contains classes or categories which contain subcategories.
- Each cell in Table 1 contains the **sums of the eight types of investments** in innovation activities according to the economic activities of the selected companies.

Description of magnitude table

Economic activities	Inv1	Inv2	Inv3	Inv4	Inv5	Inv6	Inv7	Inv8
1	611.399	48.066	(x)	25.346	900.740	8.191	(x)	(x)
2	17.560.176	4.535.764	2.381.504	1.263.395	22.267.198	583.353	3.304.743	3.995.625
3	776.246	164.321	76.847	182.526	4.228.850	107.646	1.392.673	177.406
.
.
.
69	3.524.329	(x)	16.939	1.785	9.265	(x)	(x)	(x)

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

Sensitivity analysis of variables

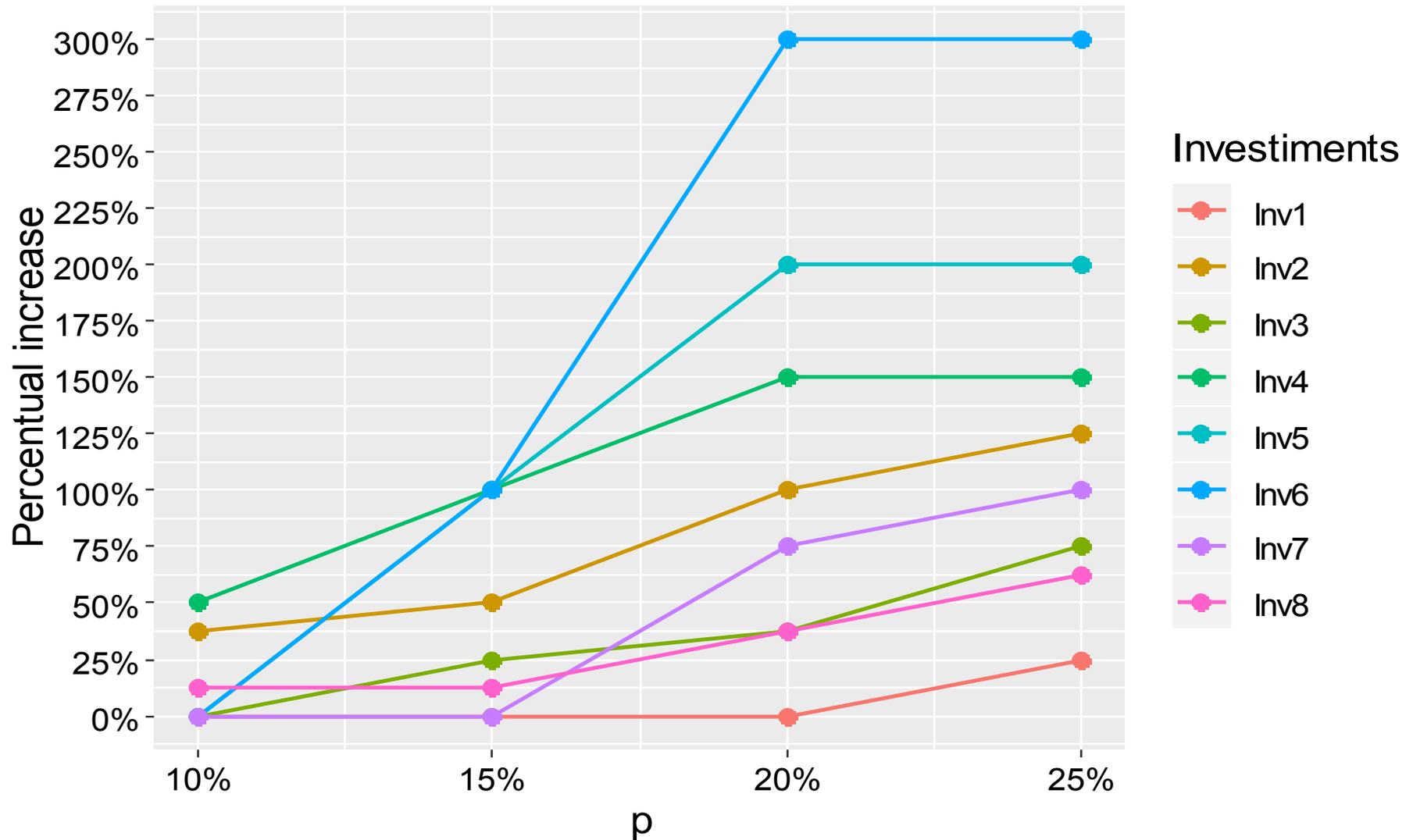


Figure 1: Percentual increases in the number of primary sensitive cells obtained with the use of p% rule, considering $p = 5\%$ as a comparison basis for Table 1

Sensitivity analysis of variables

- 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.

Final remarks

- 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.

Final remarks

It is important to highlight 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.

Final remarks

- Further studies should be done before choosing a more definitive risk analysis method for analysing the magnitude tables of the economic surveys.

Final remarks

Future studies will include further analysis of magnitude table cells:

- Adding other variables
- Secondary risk analysis
- Study of more advanced methods for protecting magnitude tables of the economic surveys.

Thank you for your attention!

samela.arantes@ibge.gov.br