Open innovation is now widely known by academics and practitioners. However, studies with large samples are important to improve its understanding. Our research aims to identify clusters of business segments with similar levels and purposes of use of open innovation using data from a large survey of Technological Innovation (PINTEC) in Brazil. We applied the clusters analysis to identify clusters of business segments for the use of open innovation. So, we used the Kruskal-Wallis test to determine if there are statistically significant differences between the groups. We identified 3 groups of business segments. Evidences suggest that open innovation is widely used by Brazilian companies. But one business segment is less prominent. At end, we present opportunities for future research.

Palavras-chave: Open Innovation, Cluster Analysis, PINTEC
1. Introduction

Open innovation is now widely known by academics and practitioners. Despite open innovation has been extensively studied since its introduction in 2003 (CHESBROUGH, 2003), it has some gaps to be explored. Quantitative studies with large samples are among these gaps (CARVALHO; SUGANO, 2016a).

The Brazilian researchers have available a large sample. The Brazilian Institute of Geography and Statistics (IBGE) has conducted a large survey of Technological Innovation (PINTEC), with the support of the Brazilian Funding Authority for Studies and Projects (FINEP) and the Ministry of Science, Technology, Innovation and Telecommunications. PINTEC has its conceptual basis based on the third edition of the Oslo Manual (IBGE, 2014). In our study, we used the latest research available in the database, from 2014, involving a total of 132,529 Brazilian companies. The data were collected between 2012 and 2014. We choose PINTEC because it is the main survey about innovation in Brazil.

Our research question is: how are the Brazilian business segments organized according to the use of open innovation? Our research aims to identify clusters of business segments with similar levels and purposes of use of open innovation. Our study will contribute with other investigations in the open innovation studies, showing main business segments applying it. Moreover, the identification of patterns of use of open innovation can help in formulating corporate strategies and public policies towards the particularities of each sector analyzed, thus favoring significant technological advances to organizations and the country.

The paper is organized as follows. After the introduction in section 2, we very briefly sketch out current theoretical and empirical developments in the open innovation studies. In section 3, we present the research methodology. Section 4 focus on presentation and discussion of results. Section 5 concludes with a discussion of potential future works and final considerations.

2. Theoretical Background

The concept of open innovation emerged at the beginning of the millennium. Despite Chesbrough (2003) be considered the seminal work of open innovation, some authors have advocated that open innovation is not a new concept. Trott and Hartmann (2009) state that open innovation is old wine in new bottles. In this sense, Huizingh (2011) states that open innovation became the umbrella that encompasses, connects, and integrates a range of already
existing activities. Hossain (2013) detailed some of the existing activities when he states that open innovation overlaps other concepts such as user generation, crowdsourcing, and distributed innovation. Dodgson and Gann (2014) tracked the origins of open innovation back to Josiah Wedgwood in 1775 when he a cooperative program with other ceramics to solve a common technical problem.

However, is there anything new or different about the open innovation paradigm? Chesbrough et al. (2006) answered yes. According to Chesbrough et al. (2006) the points of differentiation for open innovation, relative to prior theories of innovation are:

a) Equal importance given to external knowledge, in comparison to internal knowledge;
b) The centrality of the business model in converting R&D into commercial value;
c) Type I (In statistical hypothesis testing is the incorrect rejection of a true null hypothesis - a "false positive") and Type II (In statistical hypothesis testing is the incorrectly retaining a false null hypothesis - a "false negative") measurement errors (in relation to the business model) in evaluating R&D projects;
d) The purposive outbound flows of knowledge and technology;
e) The abundant underlying knowledge landscape;
f) The proactive and nuanced role of intellectual property management;
g) The rise of innovation intermediaries;
h) New metrics for assessing innovation capability and performance.

The observations of Chesbrough are important because he presented the most known and used definitions of open innovation. The seminal definition is open innovation means that valuable ideas can come from inside or outside the company and can go to market from inside or outside the company as well (CHESBROUGH, 2003). However, after three years Chesbrough reviewed and updated the definition emphasizing the intentionality of the knowledge flows into and out of the firm. Thus, Chesbrough (2006) defined open innovation as the use of purposive inflows and outflows of knowledge to accelerate internal innovation and expand the market for the external use of innovation. Recently, Chesbrough & Bogers (2014) considered the increasing interest in non-pecuniary knowledge flows, defining open innovation as being a distributed innovation process based on purposively managed knowledge flows across organizational boundaries, using pecuniary and non-pecuniary mechanisms that are in line with the organization's business model.
However, there is a paucity of quantitative studies with large samples in open innovation mainly in Brazil (CARVALHO; SUGANO, 2016a; CARVALHO et al., 2016). For instance, Carvalho & Sugano (2016b), using data from 92 firms from database of the Brazilian Association of Startups, explored different open innovation models in new technology-based firms. But Carvalho & Sugano (2016b) have a small sample. On the other hand, Oliveira & Alves (2014) present the influence of open innovation practices in the exploration of knowledge for the creation of value in environments of high complexity under conditions of uncertainty and unpredictability, using a sample of 24 responses from high technology companies in Brazil. Therefore, the sample of Oliveira & Alves (2014) is smaller than the sample of Carvalho & Sugano (2016b). So, quantitative studies with large samples in Brazil are a gap open innovation literature.

3. Research Methodology

This study follows a quantitative approach, using a comprehensive analysis of secondary data. The data were collected from the Technology Innovation Survey (PINTEC), for the years 2012-2014. PINTEC sets out in detail the methodology, showing strength and a high degree of formalization in the production of information on innovation. PINTEC uses the Oslo Manual as conceptual reference, and its methodology follows the model applied by the Statistical Office of the European Communities (EUROSTAT), the Community Innovation Survey — CIS 2008, 2010 and 2012. For this study, we used the latest research available in the database, from 2014, involving a total of 132,529 Brazilian companies, distributed in the extractive industries, manufacturing and services.

Our research followed three principal logical steps. First, we checked possible multicollinearity problems between variables, which could affect the cases grouping results. Secondly, we applied the clusters analysis to identify clusters of business segments for the use of open innovation. Thirdly, we used the Kruskal-Wallis test to determine if there are statistically significant differences between two or more groups of an independent variable. We used IBM SPSS to perform our analysis.

The studies variables are shown in table 1.
4. Presentation and Discussion of Results

Initially, we checked possible multicollinearity problems between variables. Hair et al. (2009, p. 447) state that multicollinearity acts as a weighting process in a cluster analysis in favor of the set of variables having more items. We can diagnosis multicollinearity using VIF (Variance Inflation Factor). The VIFs are shown in Table 2.

![Table 2 - Variance Inflation Factor](image)

Source: Prepared by the authors using SPSS

The VIFs is less than 3, so we have no problems of multicollinearity.
Next, we conduct the cluster analysis. We chose to use the hierarchical clustering with Ward's method. Ketchen Jr. & Shook (1996) state that hierarchical algorithms progress through a series of steps that build a tree-like structure by either adding individual elements to (i.e., agglomerative) or deleting them from (i.e., divisive) clusters. The Ward’s method produces clusters of about same size (HAIR et al., 2009). This method merges the two clusters that have the smallest cost to merge as per the equation (1) below.

\[
\text{Ward}(S_i, S_j) = \frac{N_{S_i} N_{S_j}}{N_{S_i} + N_{S_j}} d(C_{S_i}, C_{S_j}),
\]

where:

- \( N_{S_i} \) and \( C_{S_i} \): represent the cardinality and centroid of cluster \( S_i \), respectively;
- \( N_{S_j} \) and \( C_{S_j} \) represent the same for cluster \( S_j \);
- \( d() \) is a function returning the distance between the centroids of each of the two clusters.

Hair et al. (2009, p. 446) suggest using the Mahalanobis distance measure, which both standardizes variables and adjusts for high correlations. The problem is that statistical programs such as SPSS do not offer this measure, but we chose to standardize the variables to Z scores. Hair et al. (2009, p. 446) also recommend the use of Euclidean squared distance when Ward’s method is selected. Thus, we used Euclidean squared distance.

A variety of techniques are available to determine the number of clusters in a data set. When using hierarchical methods, the most basic procedure is to visually inspect a dendogram, a graph of the order that observations join clusters and the similarity of observations joined (KETCHEN Jr.; SHOOK, 1996). We looked for natural clusters of the data that are indicated by relatively dense 'branches'. Figure 1 shows the dendogram.
We determined the number of 3 clusters. In Table 3, which follows, the components of each group are presented.
Table 3 – Classification of Business Segments in Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Business Segments</th>
</tr>
</thead>
</table>
| **Group 1 (13 business segments)** | Textiles manufacturing  
Apparel and accessories manufacturing  
Non-metallic mineral products manufacturing  
Preparation of leather and manufacture of leather goods  
Rubber and plastic manufacturing  
Motor vehicles, trailers and bodies manufacturing  
Various products manufacturing  
Pharmaceuticals and pharmaceuticals manufacturing  
Computer equipment, electronic and optical products manufacturing  
Machinery, appliances and equipment manufacturing  
Extractive industries  
Electricity and gas  
Architectural and engineering services, testing and technical analysis |
| **Group 2 (6 business segments)** | Food products manufacturing  
Metal products manufacturing  
Machinery and equipment manufacturing  
Furniture manufacturing  
Chemical products manufacturing  
Activities of information technology services |
| **Group 3 (13 business segments)** | Wood products manufacturing  
Beverage manufacturing  
Printing and reproduction of recorded media  
Metallurgy  
Pulp, paper and paper products manufacturing  
Coke, oil products and biofuels manufacturing  
Tobacco product manufacturing  
Other transport equipment manufacturing  
Maintenance, repair and installation of machinery and equipment  
Editing and recording and music publishing  
Telecommunications  
Data processing, web hosting and other related activities  
Research & Development |

Source: Prepared by the authors using SPSS

Processing the Kruskal-Wallis test to determine if there are significant differences between the variables, all showed lower levels of significance than 0.05. We must emphasize that a new processing data with the same variables is not necessary. This test proves important
because they have been kept the clustering variables. Table 4 shows the hypothesis testing summarization.

<table>
<thead>
<tr>
<th>Variable</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAR&amp;D</td>
<td>0.000</td>
</tr>
<tr>
<td>EAR&amp;D</td>
<td>0.000</td>
</tr>
<tr>
<td>OEKA</td>
<td>0.000</td>
</tr>
<tr>
<td>CDProd</td>
<td>0.000</td>
</tr>
<tr>
<td>OGCDProd</td>
<td>0.005</td>
</tr>
<tr>
<td>CDProc</td>
<td>0.000</td>
</tr>
<tr>
<td>OGDProc</td>
<td>0.031</td>
</tr>
<tr>
<td>CDProdC</td>
<td>0.001</td>
</tr>
<tr>
<td>OGCDProc</td>
<td>0.005</td>
</tr>
<tr>
<td>CDProcC</td>
<td>0.000</td>
</tr>
<tr>
<td>OGDProc</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Source: Prepared by the authors using SPSS

By interpreting the three groups found, it is observed that the solution reasonably comprises the differences between groups, given the distances between the three groups. The chart shown in Figure 2 below presents the medians of each variable for the three groups and allows observing the differences between them.

![Figure 2 – Diagram of the medians of group profiles](image)

Source: Prepared by the authors using Microsoft Excel

In general, the median of groups 1 and 2 are higher than the median of all companies, while the median of groups 3 is smaller than the median of all companies. Evidences show that the companies of the business segments in group 3 developed few innovative products and
process. They also invested little in internal R&D activities, external R&D activities, and acquisition of other external knowledge. We must highlight that among the companies in this group are telecommunications, research and development, and Coke, oil products and biofuels manufacturing business segments. Here, telecommunications refers to service, not manufacturing. However, telecommunications is a technology-intensive sector and should at least develop innovative processes or make use of innovative products. Its presence in group 3 is at least odd. The same goes for coke, oil products and biofuels manufacturing business segment. It is a capital and technology-intensive industry. The concessionaires are required to invest an amount equal to 1 percent of the field’s gross production income in R&D projects. This obligation is also established by the National Agency of Petroleum, Natural Gas and Biofuels (ANP) in Resolution No. 33/2005 (ANP Resolution), which provides detailed guidance on the performance of the R&D expenditures, including models of standard reports to be used by the concessionaires when evidencing their R&D investments to ANP. At last, we have the case of research and development business segment. The core activity of these companies is innovation. Could it be a case of Narcissus without a mirror?

The evidences suggest a distinct behavior of groups in 4 main aspects: internal R&D activities, the company developed an innovative product, the company developed an innovative process, and other companies or institutes have developed an innovative process. The median of group 2 is higher than the others. The group 2 consists of companies from food products manufacturing, metal products manufacturing, machinery and equipment manufacturing, furniture manufacturing, chemical products manufacturing, and activities of information technology services business segments. Although the high investment in internal R&D activities, the companies in group 2 use other companies or institutes to develop an innovative process. This is the main difference in the aspect of open innovation.

Despite the statistical significance, the investment in external R&D activities is not very different from other groups, but the acquisition of other external knowledge has a slight difference from the others. We can deduce that the companies in group 2 use R&D outsourcing as open innovation mode, but they prefer acquisition of other external knowledge as open innovation mode, considering the high investment in internal R&D activities.

Machado et al. (2015) aimed to identify clusters of business segments for the use of cooperative agents for the development of innovations in the manufacturing industry of Brazil. Despite Machado et al. (2015) used data from PINTEC 2008, we can cross some
information. For instance, the food products manufacturing segment was in the group with the worst average index of cooperation (18.3%), standing out only in cooperation with suppliers and professional training centers and technical assistance. On the other hand, metal products manufacturing, machinery and equipment manufacturing and furniture manufacturing were in the second best group regarding the average index of cooperation (23.5%), standing out from others in cooperation with suppliers, customers or consumers, competitors, professional training centers and technical assistance and participating in fairs and exhibitions. The chemical products manufacturing segment was in the group with the best average index of cooperation (26.0%), especially in relations with R&D department, universities, research institutes, test institutions, stakeholder conferences and networks. The activities of information technology services business segment was not selected by Machado et al. (2015).

In summary, Table 4, below, provides an overview of open innovation characteristics for each of the groups.

<table>
<thead>
<tr>
<th>Group 1: Textiles manufacturing; Apparel and accessories manufacturing; Non-metallic mineral products manufacturing; Preparation of leather and manufacture of leather goods; Rubber and plastic manufacturing; Motor vehicles, trailers and bodies manufacturing; Various products manufacturing; Pharmachemicals and pharmaceuticals manufacturing; Computer equipment, electronic and optical products manufacturing; Machinery, appliances and equipment manufacturing; Extractive industries; Electricity and gas; Architectural and engineering services, testing and technical analysis</th>
<th>Group 1 is an intermediary group. Its companies are above the general median in all aspects. The dendogram allows us to divide this group in 3 subgroups. One subgroup contains capital and technology-intensive industries, such as Pharmachemicals and pharmaceuticals manufacturing; Machinery, appliances and equipment manufacturing; Electricity and gas; Motor vehicles, trailers and bodies manufacturing; and Computer equipment, electronic and optical products manufacturing. This subgroup deserves to be scrutinized. Another subgroup that deserves attention is composed by business segments from agribusiness, such as Textiles manufacturing; Apparel and accessories manufacturing; and extractive industries. The other subgroup has the following business segments: non-metallic mineral products manufacturing, rubber and plastic manufacturing, and architectural and engineering services, testing and technical analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2: Food products manufacturing; Metal products manufacturing; Machinery and equipment manufacturing; Furniture manufacturing</td>
<td>The median of group 2 is higher than the others in all aspects, but 4 aspects stand out: internal R&amp;D activities, the company developed an innovative product, the company developed an innovative process, and other</td>
</tr>
</tbody>
</table>
Chemical products manufacturing; Activities of information technology services

companies or institutes have developed an innovative process. Previous studies highlights cooperation with suppliers and professional training centers and technical assistance.

Group 3: Wood products manufacturing; Beverage manufacturing; Printing and reproduction of recorded media; Metallurgy; Pulp, paper and paper products manufacturing; Coke, oil products and biofuels manufacturing; Tobacco product manufacturing; Other transport equipment manufacturing; Maintenance, repair and installation of machinery and equipment; Editing and recording and music publishing; Telecommunications; Data processing, web hosting and other related activities; Research & Development

Evidences show that the companies of the business segments in group 3 developed few innovative products and process. They also invested little in internal R&D activities, external R&D activities, and acquisition of other external knowledge.

Source: Prepared by the authors.

5. Final Considerations

Our research aimed to identify clusters of business segments with similar levels and purposes of use of open innovation. We identified 3 homogeneous and distinct groups using cluster analysis. In general, the median of groups 1 and 2 are higher than the median of the entire sample, while the median of groups 3 is smaller than the median of the entire sample.

Group 1 is an intermediary group. In this groups is evident the presence of capital and technology-intensive industries, and segments of agribusiness. Group 2 is which most use open innovation. Previous studies highlights cooperation with suppliers and professional training centers and technical assistance. On the other hand, group 3 developed few innovative products and process. They also invested little in internal R&D activities, external R&D activities, and acquisition of other external knowledge. We must stress in the group the presence of telecommunications; research and development; and Coke, oil products and biofuels manufacturing business segments. Telecommunications and Coke, oil products and biofuels manufacturing are technology-intensive sector, while innovation is the core activity of research and development segment. Most telecommunication companies’ suppliers are foreign, and probably it explains the little investment.
Managers should not incorporate general revenues for the management of innovation in business. Instead, recognizing the particularities of each industry’s innovation process, starting with the identification of its main patterns, is expected to enable the building of a model that meets innovation demands of different sectors, also enabling an expansion in the current innovation scope of a given company with the inclusion of modes that have not been previously considered in their strategies. The companies in research and development business segment need to adopt a coherent behavior and lead by example, using open innovation strategies.

Public managers must pay attention to strategic sectors like as telecommunications and Coke, oil products and biofuels manufacturing business segments. In Brazil, the role of public administration is to create stimulus policies to strengthen the innovative capacity of national companies. New strategies and policies to foment open innovation are important. For instance, the law 9.991 established the Brazilian Electric Sector R&D Program that states that 1% of the income of the electric companies must be invested in R&D (research and development) projects, 40% of this overall budget stay under the electric company discretion (the Program is coordinated by the Brazilian electric sector regulatory agency, ANEEL).

We must highlight that evidences suggest that open innovation is widely used by Brazilian companies. But some business segments are less prominent. Why do telecommunications and Coke, oil products and biofuels manufacturing business segments make little use of open innovation? Future (qualitative and/or quantitative) studies focusing open innovation in companies from business segments of group 2 is another opportunity. Agribusiness segments show heterogeneous behavior. Thus, studies in agribusiness are interesting.

6. Acknowledgements

The authors would like to thank FAPEMIG, CNPQ, CAPES, CEFET-MG and UFLA for their support.

REFERÊNCIAS


BRASIL. Lei nº 9.991, de 24 de julho de 2000. Dispõe sobre realização de investimentos em pesquisa e desenvolvimento e em eficiência energética por parte das empresas concessionárias, permissionárias e


