Multidisciplinary Journal for Education,
Social and Technological Sciences

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

BIG-DATA and the Challenges for Statistical Inference and Economics Teaching and Learning

J. L. Peñaloza Figueroa*, C. Vargas Perez**

* *Department of Statistics and Operational Research II
Av. Filipinas s/n, Universidad Complutense de Madrid,
28003 Madrid, Spain.*
jluispf@ucm.es; Phone: + 34 913946746

** *Department of Applied Economics IV
Av. Universitaria s/n. Universidad Complutense de Madrid,
28040 Madrid, Spain.*
cvargas@ucm.es

Received: 2016-02-03; Accepted: 2016-06-26

Abstract

The increasing automation in data collection, either in structured or unstructured formats, as well as the development of reading, concatenation and comparison algorithms and the growing analytical skills which characterize the era of Big Data, cannot not only be considered a technological achievement, but an organizational, methodological and analytical challenge for knowledge as well, which is necessary to generate opportunities and added value.

In fact, exploiting the potential of Big-Data includes all fields of community activity; and given its ability to extract behaviour patterns, we are interested in the challenges for the field of teaching and learning, particularly in the field of statistical inference and economic theory.

Big-Data can improve the understanding of concepts, models and techniques used in both statistical inference and economic theory, and it can also generate reliable and robust short and long term predictions. These facts have led to the demand for analytical capabilities, which in turn encourages teachers and students to demand access to massive information produced by individuals, companies and public and private organizations in their transactions and inter-relationships.

Mass data (Big Data) is changing the way people access, understand and organize knowledge, which in turn is causing a shift in the approach to statistics

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

and economics teaching, considering them as a real way of thinking rather than just operational and technical disciplines. Hence, the question is how teachers can use automated collection and analytical skills to their advantage when teaching statistics and economics; and whether it will lead to a change in what is taught and how it is taught.

Keywords
New technologies, paradigm, logical reasoning, instrumental skills, scenarios, interactivity, modelling and simulation

1. Introduction

In this paper we intend to approach the subject of teaching and learning based on data and put on the table the widening gap between current and traditional ways of producing and processing knowledge (teaching and learning).

In this gap it is still true that the challenges related to the area of deductive and inductive logic, systemic thinking, ways of internalization and organization of knowledge, scalability of learning and the diversity of conceptual and analytical skills required have been neglected, but all of them are necessary for the development of new theories and practical applications in the real world.

The relevance of the issue lies in the set of challenges related to data identification, processing, analysis and interpretation, especially those cases with no clear structure, pending to be resolved. This challenges call into question the nature of current teaching methods and the design of the contents taught in different subjects in either academic or experimental fields.

Our modest contribution is drawing attention to the turning point that has led to the emergence of Big Data in teaching, particularly in the ways to teach conceptual and theoretical content in university classrooms. And also improving the diagnosis of limitations and updating the content of the subjects taught in university.

Our aim is also to make a record of the fact that teaching immersion in the piles of data and the analytics of Big-Data will only be possible if the outdated computer infrastructure of universities, which accentuates the gap and limited progress in data-based teaching, is substantially overcome in a short period.

In the second Section of this paper we develop the problem statement; then we continue with the analysis of the state of the issue. Section 4 is devoted to discussing a new way of content teaching. Section 5 outlines the problems and teaching limitations present in Big-Data environment and Section 6 introduces the software related to teaching and learning. Finally, last Section presents our conclusions.

2. Problem Statement

The unprecedented amount of data collected, stored and copied automatically from various sources, in addition to the continuous and rapid growth in volume (40% per year over the decade[1]), and also in availability and complexity, have prompted a set of challenges related to the identification, processing, analysis and interpretation of these data, especially of those who do not have a clear structure. These challenges call into question the nature of current teaching methods and the design of the various subjects taught in either academic or experimental fields.

A modern approach to Big Data[2], in the context of teaching and learning, not only appears as a stimulus for the development of new ideas on how to think and do activities, but also as an improvement in the understanding of how structured and unstructured systems behave in their different (economic, social, technological ...) dimensions which in turn leads to an increased infrastructure and stock of knowledge of the students (Peñaloza and Vargas, 2006).

So, the acquisition of analytical, comprehensive and interpretive capabilities is required, as part of an "updating process based on data and computational developments" in the different teaching and learning dimensions. This happens because the interactions and

[1] Big Data: ¿la ruta o el destino? IeFoundation. ORACLE.
[2] Fuzzy term that involves not only technological tools but also automated analytical techniques to predict future developments (trends) or identify patterns of behaviour, plus scalable and complex interpretive capabilities,

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

transactions existing in the piles of data automatically collected in the real world generate enough information about the program content of the subjects taught in the different areas of knowledge.

Emergent Big-Data implementation in the teaching and learning field have been designed, almost exclusively, to address issues related to the identification and evaluation of the students` learning difficulties to predict their academic progress, with the illusory idea of generalizing a personalized teaching. This bias in its implementation in addition to a narrow view of Big-Data as a "set of technologies" have hampered the progress in incorporating the philosophy of Big-Data on the contents of the various subjects included in the curricula of the different academic levels and areas of knowledge.

Big Data perceived as a "philosophy" goes beyond the technological challenges as it emphasizes factors related to the change in the way decisions are made, the scenarios are configured and analytical and interpretive processes are performed. Scalability problems caused by the exponential growth of digitized data have conditioned the definition, goals and research areas in Big-Data, reducing it to a simple technological tool (3vs or 5vs definition[3], or understanding it as the technological factors leadership[4]).

It is also true that there has been a lack of attention to those challenges related to the space of deductive and inductive logic, systemic thinking, ways of internalization and organization of knowledge, learning scalability and the diversity of conceptual and analytical skills required. They are all necessary for the development of new theories and practical implementation in the real world, which includes real-time analysis as one of the factors to consider when teaching subjects content.

Obviously, all these challenges are part of a process open to scientific knowledge related to accessibility, creation, extraction, organization and interpretation, which is based on the

[3] Volume, variety and velocity. Zikopoulos et al., 2015; Olov Schelén, Ahmed Elragal and Moutaz Haddara, 2015, Diebold, 2012; Laney, 2001.
[4] Postnote 468 July 2014 Big Data. Parliamentary Office of Science & Technology.

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

analysis and treatment of the content nested in the piles of data and the logic of the sharing resources process (new forms of communication, variety of formats and structures).

This new logic and way of generating scientific knowledge has challenged many of the established standards of scientific knowledge; as in statistical inference, to call into question the "sample-population" relationship and bring to light many of the limitations of statistics for dealing with unstructured data models; or in economic theory by questioning the validity of current economic and financial indicators used for strategic analysis and decision making in scenarios of increasing complexity and uncertainty. These problems and challenges are usually exacerbated by the scalability, complexity and speed of Big-Data (Schelen, Elragal and Haddara, 2015).

Other dimensions related to the incorporation of Big-Data on teaching content is the accessibility to non-personal information databases generated not only by the public sector but also by the different (national and international) transactions and other productive and economic activities including research projects funded by public money.

Greater accessibility and adequate distribution of resources (software, hardware and financing) would not only allow a greater use of data collected by government agencies, foundations, research institutes, business and social associations, plus businesses, etc. (i.e. accessibility policy of the CIS centre to their databases surveys –Microdata-) but it would also stimulate the full incorporation of Big-Data in the subject content taught, which would encourage the feedback from a mainly conceptual, analytical and interpretive learning, which in turn would lead to increased background knowledge of students and society in general.

3. Analysis of the state of the issue

The coexistence of teachers and students with numbers or data and data technologies is a fact that cuts across all areas of knowledge from exact science to social and behavioural science, automatically creating and generating new data types and new structures of more complex data, which are growing exponentially (Anderson, 2009; Rubin, 2007).

It is a fact that almost every aspect of daily life and of society are transformed into information or represented as numerical data (Gould, 2010; Muller et al., 2013); for example, the location of a person, the calories you consume or the web sites you visit, the searches you perform on the internet, the chats in which you participate, the e-mails you send, the services offered by a company, the customers of a supermarket, the government spending, etc. Careful observation of this type of data and what you can do with them leads to a new way of understanding the data (as resources) and to establish the role they play in the analytical process and in the generation of new knowledge (Müller, Rosenbach and Schulz, 2013).

The rediscovery of the role of data substantially changes working environments, the way of thinking and doing calculations and the access to information, things that would simply not be possible to carry out with small amounts of data. Today, models or schemes are based on the collection and analysis of large volumes of data and the interpretation of the value of those results. In the environment of Big-Data, the notion of what could be done with more than a billion individual profiles of citizens of a country, should not only be seen as a monetary value but also as a necessary social value.

The generalization of mass data in today's society has not only changed the way to understand and organize knowledge, and made possible the creation of collaboratively knowledge (Palomo et al., 2010), but has also significantly changed the context in which the content is taught and the way subjects are taught, that is, leading to data-based teaching and to shackled-to-data technologies.

The configuration of the new Big-Data scenarios not only includes the economic and social dimensions but also the academic dimension, in which the current content teaching methodology (a-conceptual operating procedures, technical methods, and memorization) has no place in education systems based on large volumes of data and data technology that have emerged in the environment of Big-Data. One of the important issues in these

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

scenarios is to ask to what extent it is possible to create a new system of teaching and learning based upon the content analysis of large volumes of data collected from external sources using new computer technologies.

The Big-Data system plays a stimulating role by integrating the roles of data consumer and content producer by just pressing a key (see the wiki technology educational projects of the UCA[5]) and by generalizing the automation of data management and handling, including the development of useful operational algorithms for the development of conceptual and theoretical content of the various subjects taught in universities, for example, intelligent transport systems (ITS) (Zhang, J. et. al., 2011) or in e-commerce, based on the analysis of large volumes of complex data (Tkacz, E. and Kapczyn'ski, A., 2009).

In fact, we live immersed in the world of mobile phones, laptops, tablets, game consoles, and similar devices. And if you combine that technology with Big-Data analytics, a series of new possibilities arises, ranging from new ways to learn basic concepts and advanced theories to new ways of reading and calculation, as well as new ways of approaching the computing concepts. So, if it were possible to obtain accurate information of what is in the minds of students, perhaps we could improve academic performance and potentially help transform the classroom going from the 20th century to the 21st century.

However, experimental implementation of Big-Data in the field of teaching and learning have been focused, almost exclusively, to deal just with aspects related to the identification and evaluation of educational problems and learning difficulties of the students to predict its evolution, and focusing on the collection of data from internal sources.

This widespread practice is based on the correspondence between teaching methods, activities and /or materials developed by teachers and the preferences in learning styles of students, which essentially means to adjust teaching to the students preferred styles[6]

[5] Free Software and Open Knowledge Office, Universidad de Cádiz. http://www.uca.es/es/
[6] These methods include "individual tutoring" , "troubleshooting" , "demonstrations " , "practical exercise" and elaborative processing (Ros García et al., 2008).

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

(Hederich M., Gravini D. and Camargo U., 2011; Gravini D., Cabrera P., Avila M. and Vargas G., 2009, García Ros et al., 2008; Bertoglia R., 2005).

The pioneering experiments with Big Data developed in the field of education have focused on basic education, with the idea of connecting decision makers with students in many ways. For example, the cluster of AltSchool[7] schools in San Francisco has made experiments with automatic collection and processing of data, internally generated, and with data technologies to help teachers to improve their teaching methods and their students learning styles (Bazaco, 2014; Cukier[8] and Mayer-Schonberger, 2013; Datalogic ADC, 2012; Orellana López y Sánchez Gómez, 2006; Lafuente and Olsina, 2001).

At the university level, experience with Big Data continues with this line of implementation and its use in all activities planned and developed. For example, "Purdue" Indiana University has developed a software product using regression methods to predict the results of their students based on the characteristics of their stocks of knowledge, in order to determine the shortcomings of their students and to keep track of their academic results.

Similarly, East Carolina University, NC, University of Louisville, KY, University of South Alabama, AL, and University of South Florida, also experimented with an interactive and educational online system called "MyMathLab[9]", designed by Pearson-Education to meet the needs, of teachers and students, of exercises and formal tests of knowledge and operational skills. This practice has led to neglect the impact of Big Data not only on how to present content but also on the actual content taught in different subjects in university curricula.

The collection and analysis of vast amounts of data to capture the intrinsic value included requires the development of new technological and technical tools. The various data technological tools (selection, capture, analysis and visualization of big data), developed by computing specialists, go through the heart of a number of disciplines of knowledge such as economics, computer science, statistics, semantic language and market research,

[7] A conglomeration of 4 schools located around San Francisco , USA, that are dedicated to innovation.
[8] Kenneth Cukier is the editor of "The Economist" Journal. http://www.economist.com/sections/science-technology.
[9] These are tests which have an unlimited number of attempts in all tasks and available learning resources.

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

among others (Mujeeb and Naidu, 2015; Christmann, Badgett and Licking, 1997) as well as transversely through a series of operational and analytical skills.

This understanding of the role of data requires methods supplied by multiple disciplines aimed at discovering useful information in large semi-structured or unstructured databases (see Table 1).

Table 1: Data Types in the Big-Data paradigm

Structured data	**Semi-structured data**	**Unstructured Data**
Customer file	Emails	Person to person
Birthdate	**Structured part:**	Communications in social networks
Name	Addressee	Person to Machine
Address	Receivers	Medical devices
Transactions in the month	Topic	e-commerce
Point of purchase	**Unstructured Part:**	Computers / Mobile Phones
Expenditure in a month	Message Body	Machine to Machine
	Atttach	Sensors / Security Cameras
		GPS devices

Source: Maté Jiménez, 2014

In these environments, based on data and data technologies (Akerkar, 2014), in which a wide variety of methodologies and techniques provided by different (Multidisciplinary) areas of knowledge are required, the areas of statistics and economics play an important role as they mix and/or merge directly or indirectly with almost all other areas of knowledge. For example, the development of models and statistical techniques has been stimulated mainly by demand from other scientific disciplines (Peña, 1995 and 2014) and that is the origin of our interest to emphasize the impact of Big Data on these two areas of knowledge.

3.1 Knowledge of inferential statistics and Big Data

Recall that our performance in the context of Big-Data occurs at the intersection between the present and the future, giving rise to a new paradigm in data analysis consisting of the implementation of intelligence in the analysis, on structured, semi-structured and unstructured data, most often in real time, to extract value from them.

The effects of massive amounts of data on real-world analysis, which correspond to the concept of Big-Data, puts into question the validity of analyses based on small samples and should be considered not as a mere trend but as a turning point in the way to build knowledge. In fact, it is causing profound changes not only in statistical inference teaching or general statistics but also in the field of quantitative and qualitative research (Peña, 2014).

The need for methods to understand the information contained in existing data and to extract value from it, has stimulated the development of statistics as a science that analyses the data and builds statistical models in order to explain reality. Although current data used to be a small, homogenous and structured sample of the phenomenon under study (biometric, economic, social or experimental data), its updating in real time is actually non-existent.

However, the complexity of new databanks (heterogeneity of sources, measurements, formats, etc.) which are a mix of different populations and formats including images, graphics and/or text, highlights the limitations of current inferential statistics, which was created to deal with samples and specifically with small samples and not with a "sample" close to population size. This situation demands innovative and efficient processing solutions for improving knowledge and decision making (Peña, 2014; Zicari, 2012; Mitchell et al., 2012.).

The key points that have an impact on the conceptual and theoretical content and the way to teach them are paradigms of Big-Data, regarding scalability, speed and variety of data, which appear as a result of the proliferation of data technologies and automation (collection and analysis), since they cannot be studied using traditional statistical techniques invented to operate on structured, homogeneous and small-scale data (samples) (Duboc, Rosenblum and Wicks, 2006; Bondi, 2000; Hill, 1990).

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

Scalability is a difficult issue to incorporate in the field of statistics especially because of the frequent mistake of thinking that the management and troubleshooting at a small scale is enough to extrapolate traditional methods when the size of the problem increases, just by adjusting the same techniques to a larger problem, which leads to a dilemma between working with a sample or the entire population (or its more realistic approach: Big-Data).

The speed and real-time updating of databases are factors that have a strong impact on the nature of the conceptual and theoretical content, since many of the tools and statistical models lose their effectiveness and generate biases and much noise, due to the existence of real time solutions.

Finally, the variety of sources and formats also affects the handling and the use of statistical tools, as we are dealing with unstructured or semi-structured databases with different types of formats (numbers, texts, graphics, videos, comments, etc.), so that statistics teaching should go beyond numbers and mechanical reproduction of probabilities, by mapping or following the traces left by the data, taking into account the conditions under which each observation arises (Gould, 2010; Deming, 1940).

In general, modern data analysis is based on the belief that it is enough to gather large volumes of data of a phenomenon to answer any question about the phenomenon under study, provided that appropriate methods are applied and that we have a powerful operational and analytical computational algorithm. For example, multi-scale methods used in image processing; or the analysis of functional data searching statistical regularities; or the application of clustering techniques by similarities or metric measurements, etc.

3.2. Economic knowledge and Big Data

The challenges Big-Data proposes in the teaching and learning of Economic Theory are based on the use of large volumes of economic, financial, business, public management, government and leisure, as well as social data, among many others. The exploitation of this

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

large collection of data requires academics (professors and researchers) greater involvement in processing and analysis, including the interpretation of results.

The forward leap to make Big-Data immersion in economics teaching come true requires a mentality shift and also experimental logic by teachers to tell the conceptual, theoretical and practical content of economics and business, based on data, in addition to the support of computers, automated calculation systems and computer applications.

Interest in massive data in the area of economics, particularly in areas such as government and industry, has been increasing in the last five years, even surpassing known issues such as data mining and structured data analysis.

The rise of Big Data is causing a turning point both at the computing and statistical level, but above all, on making markets work more effectively and efficiently, leading the global economy to a substantial change in management models and the way business is done, as well as the ways to approach clients and customers (Che, Safran and Peng, 2013).

In many branches of economic activity, economic agents target these challenges and make use of Big-data towards organizing and managing their own actions, tasks and developments. For example, governments use Big Data to map the effects of the implementation of its pro-cyclical or counter-cyclical policies or to monitor an economic shock (i.e. financial, oil or housing bubble crisis) and their effects on disposable income and/or consumption expenditure of individuals, families or companies.

Since economic processes are the direct and indirect result of the economic agents behaviour in the different dimensions of society (motives, attitudes and expectations versus spending, saving and investment), mapping techniques and automatic tracking of Big-Data system helps to identify the nature of economic crisis, like the present one, and its concrete manifestations in the real world, such as the failure of current government and business management models, the lack of ability to manage the expectations of investors and consumers and the absence of strategic management by businessmen and consumers (economic myopia) (Warren, 2008).

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

Big-Data changes the perception of uncertainty –the core of content teaching in economic science- in the sense that we have greater information and knowledge and better analytical tools, which are used to narrow uncertainty and transform it in a risk indicator. Hence, the theoretical and empirical analysis of the attitudes and expectations of economic and social agents requires new (other than current) economic variables and new analytical data tools combined with psychological variables in the context of probabilistic analysis.

The scenarios set by Big-Data in the predictive dimension are one of the strengths of economics teaching and learning; however there are certain difficulties for the incorporation of the the results of the analysis of large masses of data, collected in real time, into the programs of the subjects, as a way of approaching prediction of the future. For example, a predictive Big-Data analytics which gets, in a digital way, all the information in the market would have reduced the size of the property bubble and minimized its consequences.

Digitalization and automation of millions of jobs emerge not only as an inevitable trend, but as a deep change in labour relations, the way people work and the type of work they have. These processes affect on the one hand, the configuration and dynamics of the labour market and raises questions like: Will there be enough jobs in the future to relocate and retrain people displaced by automation? Or Would it be possible to establish a wealth distribution system to regulate the labour market? And, on the other, they affect the management and functioning of those organizations or systems which were not able to change and align with the guidelines of the new Big-Data scenarios (Gould, 2013 and Cobb, 2007).

In the business dimension, companies use Big Data to develop specific implementation in general work areas. For example, credit card companies (VISA, MasterCard, …) are interested in the rapid or real-time recognition of unusual patterns of behaviour in the use of credit cards (sudden charges of large amounts of money on a credit card at an unusual place for the customer) in order to warn the cardholder about that anomaly (Müller, et. al, 2013).

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

Overall, Big-Data can be applied to economic policy management, to microeconomic and macroeconomic theory, to companies like Google, Facebook, Coca-Cola, Nestle, etc., and/or to a small business which is dynamic and flexible not only in its operations but also in its decision-making. A clear example is found in the statement by Fabinger et al. (2015) who tell us that the use of large databases allows us to a better approach to the correct form of the demand curves, while acknowledging that the larger is the data set used, the more complex is the construction of the curves of empirical demand and the calculations involved.

In other words, data is becoming the new currency on which both short and long term business and consumer decisions are based, and it is a clear improvement in overall productivity.

4. Options for a new way of content teaching

Our basic proposal is founded on the idea that conceptualization and theorizing is the best way to provide the necessary basis and language, to develop analytical and interpretive methods, based on large bases of semi-structured and unstructured data, to solve problems and challenges of the real world.

The collecting and analytical potential of Big-Data will be useful if it is able to stimulate the shift of focus from operational activity-based teaching to one based on data and data technologies, so that it is possible to make a forward leap from sample- based analytics to multiscale analytics and parallel effectiveness, which makes possible the immersion of theoretical concepts and models in Big-Data environments.

The figures from Gartner Research (2015) support these ideas. In 2015, 4.4 million technology-related jobs will be created worldwide to support the management of large volumes of data (Big Data). However, limited access to large volumes of data by universities currently contributes to the generation of a skills gap in the workforce of countries or among a group of countries.

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

From the statistics point of view the potential of Big-data manifests itself as a system that uses information technology to collect and organize data and (statistical) analytical methods to study and interpret them. The aim is to generate numerical descriptions of relationships between data, automatically and in real time.

Exploiting the "causality" and correlation relationships among different objects usually involves the use of efficient approximation algorithms such as monotonic multivariate regression on a large scale[10]. Nevertheless, its implementation has certain limitations related to the dynamics of updating the database in real time.

New techniques to exploit the potential of Big Data, in structured or semi-structured architectures, are characterized by the study of the problems related to the presence of outliers and the configuration of clusters (cluster analysis, survival analysis, statistical learning, data mining, meta-analysis, and KKD and latent analysis, among others). In addition new multivariate time series methods are used to analyse dynamic data and independent principal components to analyse structured functional data.

These challenges require a comprehensive review of the existing data management systems covering the scalability metrics as partitions, replicas, concurrency control and consistency of large semi-structured and unstructured databases. This happens due to the physical principle that stablishes that the exponential increase in the speed of data collection and information extraction is closer to real time, making the traditional principles and methods obsolete and ineffective for analysis and verification.

From the point of view of economics, due to the ubiquity of data it is possible that, as part of an International Trade course, we could automatically examine and map the effects of imperfect competition on trade and terms of exchange among countries, or similarly, we could track the impact and spread of economic shocks in the economy or group of economies, as part of the content of a Macroeconomics course.

[10] Econometric models work by statistical causality relations, while the Big-Data system works by statistical correlations and associations, and not by causalities, since information is instantaneous.

For example, a macroeconomic problem of great interest is measuring inflation. Big-Data could provide a new way to gauge the growth of prices. Inflation is a simple concept, but continuous price increases are surprisingly difficult to measure. Big data could easily carry out the process of collecting and monitoring products prices and consumption baskets over time.

In theory, online prices and baskets could be tracked digitally. ADOBE, a data technology company, is trying to calculate the US CPI figure. The company collects sales data with an anonymous design of the websites that use their software. The amount of data available is huge: according to the company, it includes three-quarters of online spending made in the top 500 retail businesses in the United States. The company is using this ocean of information to compile or create a "digital price index" (DPI) to compete with the official measures of inflation.

In the business dimension the potential of Big Data requires a new team of analysts who can bridge the gap between information technology and businesses, making use of the conceptual evolution and the real time decision-making approach. This requires a different use of resources configuration, based on data and software.

Among the limitations of the immersion of the teaching of economics in the piles of data are, in the first place, the speed increase which modifies the economic relations that occur among economic agents when moving from one level to other (macro -7 micro), due to the new forces acting at that level. In the second place, the semi-structured and unstructured nature of the data collected and stored which make many of the analytical tools of economic theory obsolete, both at the micro level and the macro level.

In summary, the challenges in teaching and learning Statistics Inference and Economic Theory, in the process of immersion in the use of large volumes of data of any kind, require the full exploitation of the data. For this purpose, special skills are needed to lead the challenges involved, which can be exacerbated by the scalability, complexity or speed of Big-Data.

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

5. Problems and Limitations

So far teachers and scientists have developed various technologies and algorithms to collect, capture, analyse and visualize data from large databases. These technologies and algorithms cross a number of disciplines such as economics, computer science, and statistics, and they also require certain analytical and operational skills. This means that multidisciplinary approaches are necessary to discover useful information from large volumes of data.

Generally, categorization of Big-Data based on data heterogeneity, results in certain semi-structured and unstructured architectures. Management and analysis of these architectures requires a dynamic and flexible computing infrastructure that universities do not currently own (most of them have an outdated computer infrastructure).

Storing Big-Data is a critical task that requires advanced technology such as Hadoop and HDFS from Apache project, which are beyond the power of universities and even worse for teachers, to implement the new way of conceptual and theoretical content teaching based in large volumes of data.

6. Software for teaching and learning

In recent years, there have been several solutions using advanced analytics in the software market for higher education, such as: Big Data for Management Companies, Big-Data for Business and Trade and Big-Data for Teaching and Scientific Research, among others, all of them focused on automatic and analytical learning.

This is the case of MapReduce which is an efficient analytical tool of large volumes of data consisting of a visual and graphical search of behaviour patterns using clustering and classification techniques, combined with some other conventional techniques such as DBMS and indexing methods.

The wide variety of applications, with potential uses in conceptual and theoretical content teaching and learning, can be classified as follows:

Big Data tools for batch processing:

Karmasphere Studio and Analyst	Tableau
Jasper soft BI	Talend Open Studio
Sky tree Server	Apache Hadoop
Pentaho Business Analytics	MapReduce: AIDE y A2DE
Apache Mahout	Dryad

Big Data tools for stream processing:

Storm	Splunk
S4	Apache Kafka
SQLstream s-Server	SAP Hana

Big Data Techniques such as:

Statistical techniques	Machine Learning Techniques
Optimization Methods	Classification and Clustering techniques
Data mining techniques	Regression Analysis techniques

7. Conclusions

Activities related to data collection and data analysis are not new as they have been done for thousands of years. What is new is the great number of advances in data technologies during the last twenty years, combined with the convergence of phenomena such as the Internet, smartphones, tablets, mobiles, Wi-Fi, apps, social networks, the cloud, automatic learning, cognitive computing, etc., so we can assume that data technology will continue improving, resulting in a digital gap among individuals and among organizations in the framework of a competition based on data analytics.

The automatic interpretation of the results obtained from Big-Data and the specification of their semantics and the preservation of the context in which they occurred, are some of the challenges to be faced today and in the near future, when we try to immerse teaching in the ocean of data and results (Big-data). In this scenario various concepts of data classification and grouping arise, including both the context and the standardization of formats (warehouse) to carry out the extraction of relevant information and its transformation into knowledge in the Big Data environment.

Big Data environment is an opportunity to reposition the teaching of Statistics in the centre of the acquisition of knowledge, integrating concepts and creating new methodologies in line with the advances in data technologies. For that purpose professors and researchers need a mental shift and start working with massive data, doing Big Data-based studies in their areas of research.

Within the framework of Economics teaching, Big Data is the new frontier for innovation, competition and productivity which cuts across all industries, and all production and public activities. By leveraging the large volumes of data it is possible to transform the knowledge we have about economic policy, the corporate image of a company, a business partner or the requirement of making intelligent decisions by public and private operators.

With Big-Data we run the risk of carrying out data-based problem analytics without a prior assessment of the need to adjust the data tools applied to the phenomenon under study, as a result of the "conviction" of computer gurus (Microsoft, Java, Paython, etc.) that direct

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

implementation of these analytical techniques to data, whether economic or not, is a sufficient condition to turn data into knowledge.

Universities are organizations where it is feasible to apply Big Data and data technologies to increase efficiency and quality of training processes.

In essence, it is necessary to opt for a conceptual and interpretive teaching, which does not imply a detailed description of the facts or phenomena but an analysis in terms of cause and origin of the facts or phenomena of interest and its subsequent explanation. Therefore, regardless of the complexity of the problem or phenomenon it is possible to find a fundamental explanation so that the underlying theory or models can be explained, investigated or replicated.

The large number of software solutions around Apache Hadoop project is a sign of progress in resolving the various challenges that arise, as immersion generalizes and deepens in teaching and knowledge development. However, democratizing access to large databases remains a pending task in universities, without which it is not possible to meet the large demand for skilled analysts in the various areas of knowledge.

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

References

Akerkar R. (Ed.). (2014). Big Data Computing. CRC Press.

Anderson, C. (2009). "Living by Numbers". Wired Magazine. July 2009.New York: Conde Nast Publications.

Cobb, George, W. (2007). The Introductory Statistics Course: A Ptolemaic Curriculum. Journal of Technology Innovation in Statistics Education Vol. 1 No. 1. UCLA. (available at: http://escholarship.org/uc/item/6hb3k0nz).

Cukier, Kenneth and Mayer-Schönberger, Viktor (2013). Big Data: A Revolution that Will Transform How We Live, Work and Think. John Murray Publishers. London, UK.

Dean, Jefry and Ghemwat, Sanjay (2010). Map Reduce: A Flexible Data Processing Tool. Communications of the ACM, Volume 53, Issuse.1, January 2010, pp 72-77.

Dean J, Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. Communications of the ACM, 51(1), 107-113.

Diebold, F. X. (2012). A Personal Perspective on the Origin(s) and Development of "Big Data: The Phenomenon, the Term, and the Discipline". Draft paper dated November 26, 2012.

Duboc, Leticia; Rosenblum, David S.; Wicks, Tony (2006). A Framework for Modelling and Analysis of Software Systems Scalability. Proceeding of the 28th international conference on Software engineering - ICSE '06. p. 949. ISBN 1595933751.

Fabinger, Michal and E. Glen Weyl, E. Glen (2015). A Tractable Approach to Pass-Through Patterns. Link: file:///C:/Users/JL/Downloads/SSRN-id2575471.pdf (access 07/04/2016).

García Ros, R., Pérez González, F. & Talaya González, I. (2008). Preferencias Respecto a Métodos Instruccionales de los Estudiantes Universitarios de Nuevo Acceso y su Relación con Estilos de Aprendizaje y Estrategias Motivacionales. Electronic Journal of Research in Educational Psychology, 6(16), 547-570.

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

Gartner, Eric Savitz (2012). "10 Critical Tech Trends for the Next Five Years". http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five years/.

Gould, Robert (2010). Statistics and the Modern Student. International Statistical Review, 78, 2, pp. 297-315. UK.

Hederich Martínez, C., Gravini Donado, M. & Camargo Uribe, A. (2011). El Estilo y la Enseñanza: Un Debate Sobre Cómo Enfrentar las Diferencias Individuales en el Aula de Clase. In R. Roig Vila y C. Laneve, C. (Eds.), La pratica educativa nella società dell'informazione. L'innovazione attraverso la ricerca (pp. 213-222). Alcoy-Brescia, Italia: Marfil y La Scuola Editrice.

Kambatla K, Kollias G, Kumar V, Grama A. (2014). Trends in Big Data Analytics. Journal of Parallel Distributed Computation, 74, pp. 2561-2573.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Stamford, CT: META Group Research Note, February 6, 2001.

Leedy, P. & Ormrod, J. (2001). Practical Research: Planning and Design. 7th Editon. Upper Saddle River, NJ: Merrill Prentice Hall. Thousand Oaks: SAGE Publications.

Maté Jiménez, Carlos (2014). Big Data: Un Nuevo Paradigma de Análisis de Datos. Revista Anales De Mecánica Y Electricidad. Nov-Dec. http://www.revista-anales.es/web/n_29/pdf/10-16.pdf.

Meyer-Schonberger, Viktor and Cukier, Kenneth (2013). Big Data: A Revolution that Will Transform How We Live, Work and Think. John Murray Publishers. London. UK Company.

Mitchell, Ian and Wilson, Mark (2012). Linked Data. Connecting and Exploiting Big-Data. White-Paper, march. Fujitsu UK. Link: https://www.fujitsu.com/uk/ Images/Linked-data-connecting-and-exploiting-big-data-(v1.0).pdf.

Mujeeb, S. MD. And Naidu, L. Kasi (2015). A Relative Study on Big Data Applications and Techniques. International Journal of Engineering and Innovative Technology (IJEIT).Volume 4, Issue 10, April. ISSN: 2277-3754.

Müller, Martin U., Rosenbach, Marcel and Schulz, Thomas (2013). Living by Numbers: Big-Data Knows What your Future Holds. DER SPIEGEL No. 20. Germany (Translated from German by Christopher Sultan).

Peña, D., Prieto, J. and Viladomat, J. (2010) "Eigenvectors of a Kurtosis Matrix as Interesting Directions to Reveal Cluster Structure", Journal of Multivariate Analysis 9, 1995 -2007, 2010.

Peñaloza F., Juan Luis y Vargas P., Carmen G. (2015). "Construction and Evaluation of Scenarios as a Learning Strategy through Modelling-Simulation". In: Multidisciplinary Journal for Education, Social, and Technological Science. Vol. 14, Num. 1. pp. 40 – 62. DOI: http://dx.doi.org/10.4995/muse.2015.2245.

Peñaloza F., Juan Luis y Vargas P., Carmen G. (2006). ¿Qué debe Cambiar en el Aprendizaje de la Estadística en las Ciencias del Comportamiento?. XIV Jornadas de ASEPUMA y II Encuentro Internacional, 21 y 22 de septiembre de 2006. Facultad de Ciencias Económicas y Empresariales, Badajoz, España.

Schelén, Olov, Elragal, Ahmed and Haddara, Moutaz (2015). A Roadmap for Big-Data Research and Education. Technical Report. Luleá University of Technology. ISBN 978-91-7583-275-3.

Tkacz, Ewaryst and Kapczyn´ski, Adrían (2009). Internet: Technical Development and Applications, Springer, 2009.

Warren, Kim (2008). Strategic Management Dynamics. John Wiley & Sons, LTD. England.

Williams, Carrie (2007). Research Methods. Journal of Business & Economic Research. Volume 5, Number 3 – March.

http://dx.doi.org/10.4995/muse.2017.6350
EISSN: 2341-2593

Zhang, Junping, Wang, Fei-Yue, Wang, Kunfeng, Lin, Wei-Hua, Xu, Xin, Chen, Cheng (2011). Data-Driven Intelligent Transportation Systems: A Survey. IEEE Trans. Intell. Trans. Syst. 12 (4), pp. 1624–1639.

Zicari, R. V. (2014). Big Data: Challenges and Opportunities. In Akerkar R. (Ed). Big data computing. CRC Press.

Zikopoulos, P., deRoos, D., Bienko, C., Buglio, R., & Andrews, M. (2015). Big Data Beyond the Hype: A Guide to Conversations for Today's Data Center. New York, NY: McGraw-Hill. Education.