

La investigación en las carreras de estadística del Ecuador: aproximación usando programas y líneas de investigación de la Escuela Superior Politécnica de Chimborazo, Ecuador (ESPOCH)

The research in Ecuadorian statistic careers: approximation using the Higher Technical Institute of Chimborazo, Ecuador (ESPOCH) research programs and lines

Rubén Pazmiño-M.*, Pablo Flores-M**, Erika Yamasque*, Marlene García*, Paulina Valverde*

Resumen

Conscientes de que los procesos de investigación no pueden mejorar si no se conoce su estado actual, proponemos el presente trabajo, el cual a partir de las líneas y programas de la carrera de Ingeniería en Estadística Informática de La Facultad de Ciencias – ESPOCH, busca una aproximación a la investigación desarrollada en las carreras de estadística del país a nivel de pregrado. Para esto se realiza un estudio descriptivo, temporal e implicative de 776 trabajos de titulación pertenecientes a las carreras de estadística de la Universidad Central del Ecuador, Escuela Politécnica del Litoral y Escuela Superior Politécnica de Chimborazo.

Los resultados muestran ciertas tendencias en el número de trabajos de titulación, detalles sobre las sublíneas más investigadas y otras características importantes de cada una de las instituciones de educación superior.

El análisis implicative muestra sublíneas de investigación similares y ciertos niveles de cohesión entre las sublíneas de investigación de la carrera perteneciente a la ESPOCH. Este trabajo muestra una metodología de análisis que puede ser aplicada en otros ámbitos de la investigación y su gestión.

Palabras clave: carreras de estadística, investigación, trabajos de titulación, líneas y programas de investigación ESPOCH.

* Docente. Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Grupo de Investigación en Ciencia de Datos CIEDD.

** Docente Escuela Superior Politécnica de Chimborazo - Correo electrónico: p_flores@epoch.edu.ec – Riobamba – Ecuador

Abstract

Aware that research processes cannot be improved if their status is not known, we propose this work, which based on the lines and programs of Engineering in Informatic Statistics career of the Sciences Faculty of the Higher Technical School of Chimborazo, Ecuador (ESPOCH), seeks an approach to the research developed in the country's statistics careers. For this, a descriptive, temporal and implicative study of 776 titling works belonging to the careers of statistics in Central University of Ecuador, Polytechnic School of Litoral and Higher Technical School of Chimborazo.

The result shows certain trends in the number of the degree studies, details about the most sub-lines researches and other important characteristics of each institutions of higher education.

The implicative analysis shows similar sub-lines and certain levels of cohesion between the sub-lines research of the career belonging to the ESPOCH. This study shows a analysis methodology that can be applied in other areas of scientific research management.

Keywords: Statistic careers, research, titling works, Ecuador, ESPOCH research programs and lines.

Introducción

La ESPOCH (Escuela Superior Politécnica de Chimborazo), nació desde el Instituto tecnológico Superior de Chimborazo, con Ley No.69,09, expedida por el Congreso Nacional, el 18 de abril de 1969. Inició sus actividades académicas el 2 de mayo de 1972. El Instituto de investigaciones es el que se encarga de fomentar, sistematizar y difundir la investigación, las líneas y programas de investigación están en vigencia hasta mayo del 2018.

La educación superior ecuatoriana actualmente cuenta con tres universidades que ofertan la carrera de Estadística: La Escuela Superior Politécnica de Chimborazo (1) con la carrera de Ingeniería en Estadística Informática, la Universidad Central del Ecuador (2) con la carrera de Ingeniería en estadística y la Escuela Superior Politécnica del Litoral (3) con la carrera de Ingeniería en Estadística Informática.

A partir de las líneas de investigación declaradas en el currículo de la carrera de Ingeniería en Estadística Informática de la ESPOCH (4)

(Análisis Multivariado, Diseño de Experimentos y Modelación estadística), realizamos una aproximación a la investigación de pregrado en las carreras de Estadística del Ecuador con un énfasis especial en la carrera tomada como base perteneciente a la ESPOCH, lo cual permite conocer la coherencia de la investigación de pregrado con las líneas y programas definidos por el Instituto de Investigaciones de la ESPOCH (IDI).

Esta investigación se basa en la gestión del conocimiento de todos los trabajos de titulación de las carreras de Estadística realizados desde el año 1999 hasta el año 2017 en el Ecuador, dando como resultado un estado situacional aproximado de la investigación en estadística en el Ecuador. El análisis se centra en comparar todos estos trabajos con las líneas de investigación de la respectiva carrera perteneciente a la ESPOCH. En la base de datos DSpace de cada universidad se registran un total de 776 trabajos de titulación realizados en las tres universidades, de las cuales a la fecha 33 pertenecen a la ESPOCH, 734 a la ESPOL y 9 a la UCE, todos ellos analizados minuciosamente y de manera individual.

Se usa además para el análisis la técnica de análisis estadístico Implicativo (AEI), la cual es una teoría estadística desarrollada hace 40 años en Francia por Regis Gras, utiliza los conceptos de similaridad, cuasi implicación y cohesión para generar relaciones asimétricas entre sujetos y variables (5). Se basa en la modelización estadística de la cuasi-implicación: cuando la variable o la conjunción de variables “a” es observada en la población, entonces generalmente la variable “b” lo es también. Los conjuntos de reglas obtenidos pueden conducir a hipótesis de causalidad (6). El sistema informático CHIC es un software desarrollado por Raphaël Couturier que automatiza la teoría del AEI (7). El AEI ha sido utilizado ampliamente en el ámbito educativo y tiene además aplicaciones en las áreas de la psicología, matemática, arte, medicina y otras (8).

Materiales y Métodos

Por el paradigma de investigación es una investigación de tipo cuantitativo, por el tipo de diseño utilizado es exploratoria, por el tiempo de estudio es transversal, el colectivo de estudio lo conforman todas las carreras de Estadística del Ecuador, la población es la información sobre nombre, filas, columnas, total de datos, tiempo y memoria, por la amplitud de estudio es un censo. Los resultados descriptivos y temporales fueron obtenidos utilizando el software estadístico R versión 3.4.2 (9), y a través de las funciones contenidas en el paquete ggplot2 (10) descargado directamente del R-CRAN. Los resultados implicativos fueron obtenidos utilizando el software CHIC (Cohesive hierarchical implicative classification) versión 6.0 (11). La metodología para la parte descriptiva y temporal es la Estadística. La metodología para el análisis Implicativo se basó en (12) y se detalla a continuación:

Selección:

Los 734 trabajos de titulación se utilizaron para el análisis descriptivo y temporal. Se elaboraron preguntas guía para el análisis implicativo y estas ayudaron a asociar los 33 trabajos de titulación de la carrera asociada a la ESPOCH con las líneas y programas definidos institucionalmente (13).

Preprocesamiento:

El conjunto objetivo fueron 33 tesis de la carrera de Estadística perteneciente a la ESPOCH. Mediante el software EndNote, la información de autor, año, título, carrera, universidad, grado, número de páginas, tutor, fecha, palabras clave, resumen y link entre otras fueron asociados a las tesis objetivo. Se manejó una base de datos dicotómica de un total de 33 filas (trabajos de titulación) y 132 columnas (variables).

Transformación

Con base en las preguntas guía, se ingresaron valores manualmente en el software Microsoft Excel y analizados mediante procesos de evaluación individual, por pares y aleatorios. Los datos fueron transformados a una codificación binaria equivalente a 1 = Si y 0 = No por su simplicidad de manejo.

Análisis estadístico Implicativo:

Se consideraron gráficos de implicación, árboles de similaridad y árboles de cohesión con el propósito de establecer semejanzas y dependencias entre las tesis objetivo y las líneas de investigación mediante los métodos SIE, analizados mediante el software CHIC.

Interpretación y evaluación:

La interpretación y evaluación se basan en los grafos implicativos que muestran mediante grafos dirigidos qué variables cuasi-implican a cuáles, los árboles de similaridad muestran las variables más similares utilizando el criterio de similaridad de Israel Lerman, el árbol de cohesión permite ver mediante un dendograma las diferentes relaciones asimétricas entre variables y clase de variables.

Resultados y Discusión

A continuación, se presenta un análisis temporal, descriptivo e implicativo sobre las investigaciones realizadas a nivel de trabajos de titulación por las tres universidades respecto a las carreras de estadística.

Estudio temporal

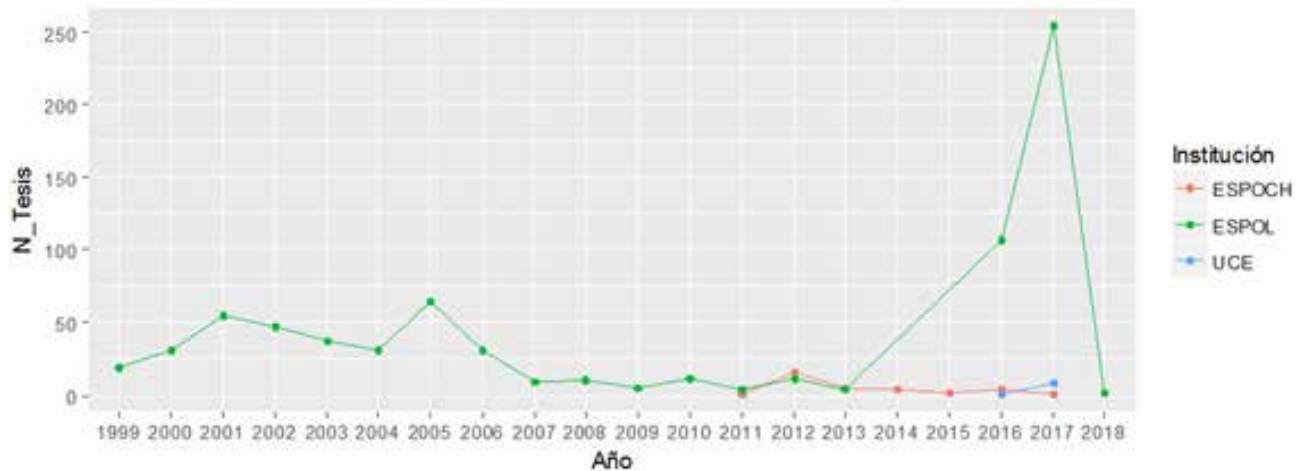


Figura 1: Tesis realizadas desde 1999 hasta 2018.

La Figura. 1 muestra el número de tesis registradas en DSpace . Se observa claramente que la gran mayoría de tesis pertenecen a la ESPOL, la cual ha producido tesis desde el año 1999. La ESPOCH y la UCE empiezan a publicar tesis en una cantidad muy reducida a partir del año 2011 y 2016 respectivamente.

las tres universidades respecto a las carreras de estadística.

Estudio descriptivo

A continuación, se presenta un análisis temporal, descriptivo e implicativo sobre las investigaciones realizadas a nivel de trabajos de titulación por

A continuación, se presenta un análisis descriptivo sobre las investigaciones realizadas a nivel de trabajos de titulación para estas tres universidades en conjunto y luego se realiza el mismo análisis para cada carrera por separado.

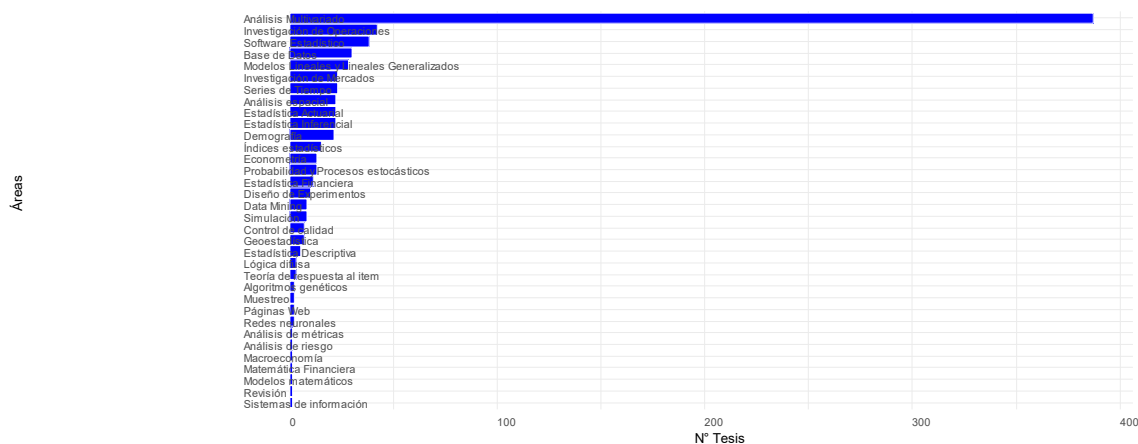


Figura 2: Tesis realizadas en las tres universidades clasificadas por áreas.

La Figura 2 muestra que, si clasificamos las tesis realizadas de las tres universidades en estudio por área de aplicación de la estadística, predominan aquellas que se realizan en el campo del análisis multivariado. Seguramente la mayoría

de los estudiantes se encuentran fuertemente influenciados por esta área, aunque posiblemente esto ocurre principalmente en la universidad con mayor número de trabajos realizados.

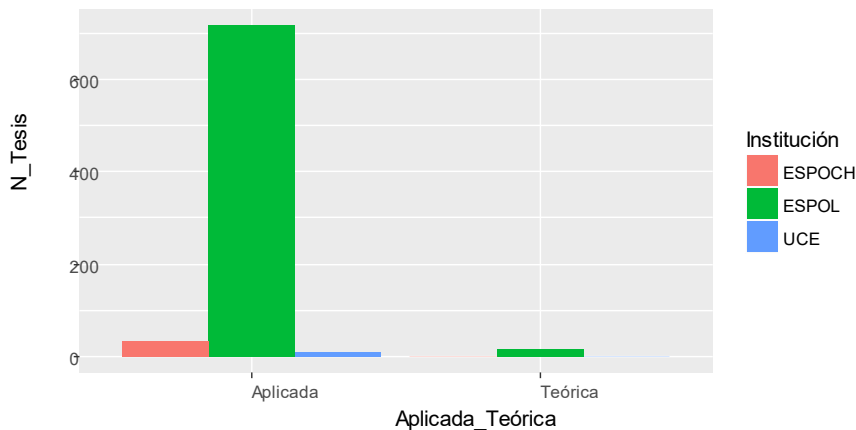


Figura 3: Tesis realizadas en las tres universidades clasificadas en el campo Aplicado o Teórico

La Figura 3 muestra que todas las tesis de tipo teóricas, es decir que aportan directamente a la estadística como ciencia han sido elaboradas en la carrera correspondiente de la ESPOL, a pesar de que la gran mayoría de tesis realizadas en esta

universidad son aplicadas a otras áreas distinta a la estadística. La UCE y ESPOCH no cuentan a la fecha con ninguna tesis de tipo teórica, todas son aplicadas.

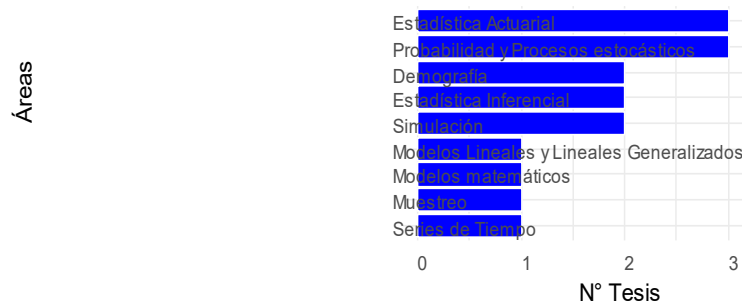


Figura 4: Tesis de tipo teórico clasificadas por área

La Figura 4 muestra las áreas de la estadística en donde se desarrollaron las tesis de tipo teórico. Ya que la gran mayoría de tesis son de tipo aplicativo, es lógico pensar que las áreas donde estas fueron desarrolladas coinciden con las que se presentaron en la Figura 2. La Figura 5, muestra el número de tesis que han sido desarrolladas en las mismas

líneas de investigación de las existentes en la carrera de Ingeniería en Estadística Informática de la ESPOCH. Se puede ver que existe un importante número de tesis que no caen dentro de ninguna de las tres líneas de investigación declaradas y que las que más coinciden son las que se realizan en el campo del análisis multivariado.

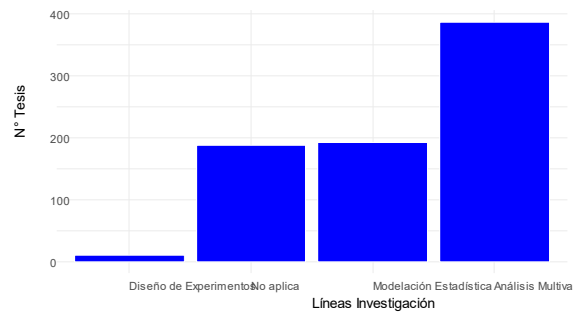


Figura 5: Tesis de las tres universidades que han sido desarrolladas en las líneas de investigación declaradas en la ESPOCH

La Figura 6, Figura 7 y Figura 8 muestran las áreas de la estadística en que se han desarrollado las tesis realizadas en las carreras correspondientes de la ESPOCH, UCE y ESPOL respectivamente. Cabe recalcar que el campo “Estadística Descriptiva”

se refiere a aquellas tesis que utilizaron un estudio descriptivo univariado para la obtención de resultados, en aquellos casos donde el análisis descriptivo fue multivariado se los categoriza dentro del área “Análisis Multivariado”.

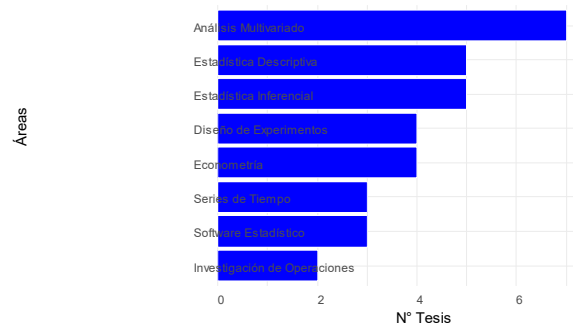


Figura 6: Tesis realizadas en la ESPOCH clasificadas por áreas.

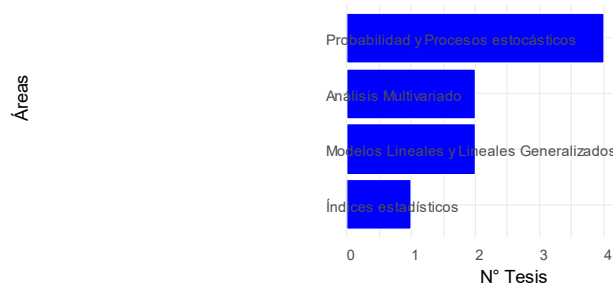


Figura 7: Tesis realizadas en la UCE clasificadas por áreas

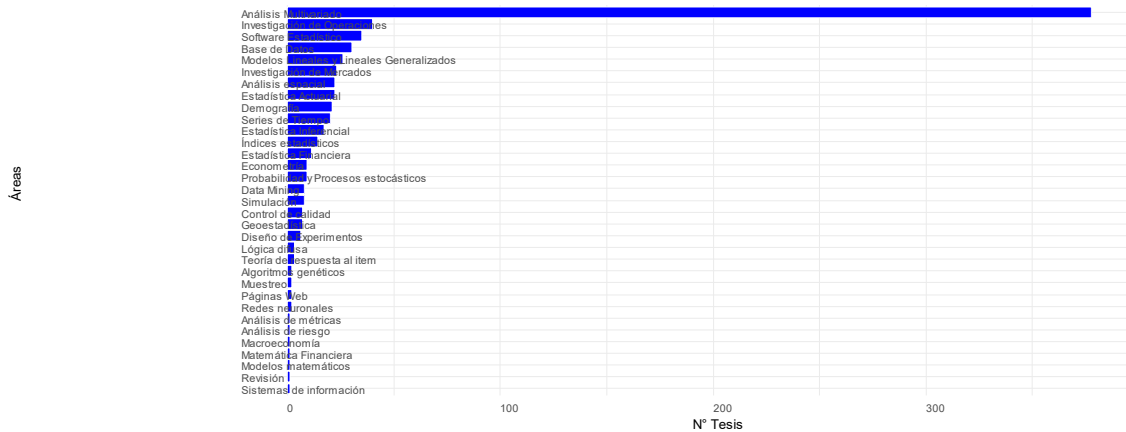


Figura 8: Tesis realizadas en la ESPOL clasificadas por áreas.

La Figura 9, Figura 10 y Figura 11, muestran el número de tesis que han sido elaboradas en la ESPOCH, UCE y ESPOL respectivamente y que pertenecen a las líneas de investigación declaradas en la carrera de Ingeniería en Estadística Informática de la ESPOCH. Se puede observar que la mayoría de tesis realizadas en la ESPOCH y la UCE son de modelación estadística

y relativamente existe un número importante de tesis que no caen dentro de ninguna de las tres líneas declaradas, mientras que las realizadas en la ESPOL concuerdan en su mayoría con la línea de investigación de Análisis Multivariado, aunque sigue existiendo un número importante de áreas que no concuerdan con ninguna línea de investigación declarada en la ESPOCH.

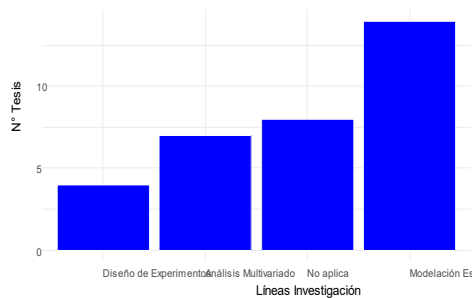


Figura 9: Tesis de la ESPOCH desarrolladas en las líneas de investigación de la ESPOCH.

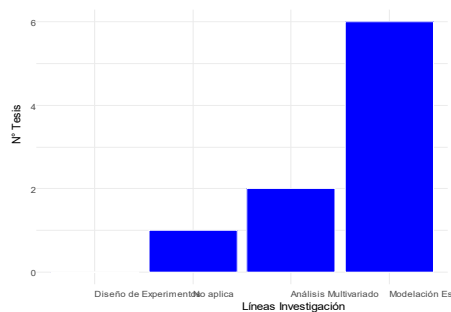


Figura 10: Tesis de la UCE desarrolladas en las líneas de investigación de la ESPOCH

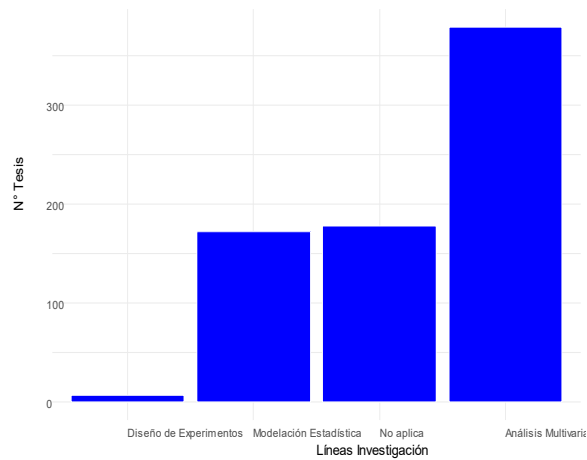


Figura 11: Tesis de la ESPOL desarrolladas en las líneas de investigación de la ESPOCH.

Preocupa de gran manera la gran cantidad de tesis desarrolladas en áreas que no concuerdan con las líneas de investigación declaradas por la correspondiente carrera de la ESPOCH. La Figura 12 muestra las áreas en las que se desarrollaron las tesis de la ESPOCH y UCE que no se encuentran

dentro de ninguna de las líneas de investigación mencionadas. Se puede observar que apenas existen tres áreas que difieren a las líneas de investigación de interés, esto seguramente debido a la poca cantidad de tesis desarrolladas en estas dos universidades analizadas.

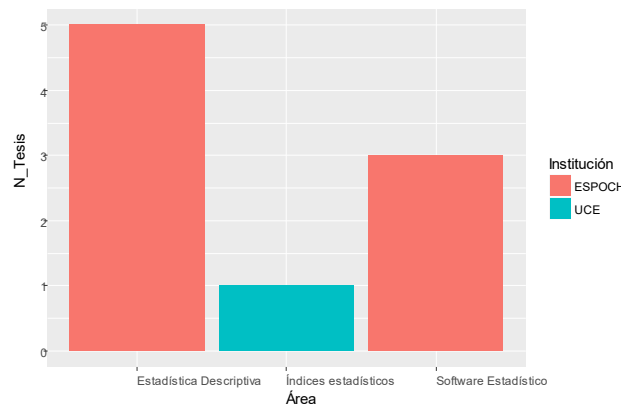


Figura 12: Tesis de la ESPOCH y UCE desarrolladas áreas distintas a las líneas de investigación de la ESPOCH.

Resulta más interesante realizar el mismo análisis, pero con las tesis elaboradas por la correspondiente carrera de la ESPOL, que es donde se registra el mayor número de trabajos. La Figura 13 muestra estas áreas, siendo las más

comunes aquellas desarrolladas en los campos de Software estadístico, Base de datos, Demografía, Análisis espacial, Índices estadísticos, estadística financiera, Data Mining, etc.

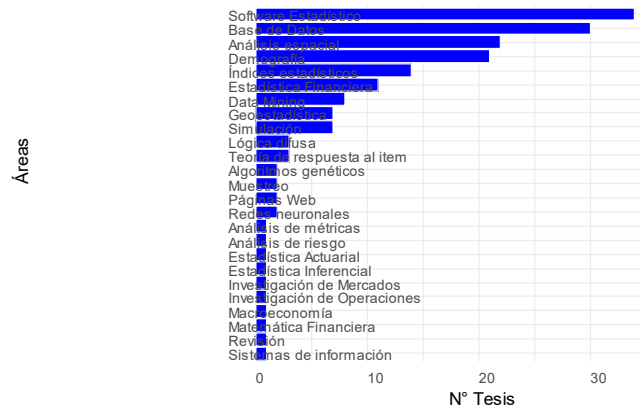


Figura 13: Tesis de la ESPOL desarrolladas áreas distintas a las líneas de investigación de la ESPOCH.

Estudio Implicativo

A continuación, se muestran ejemplos sobre la utilización de tres de las herramientas de análisis proporcionadas por el Análisis Estadístico Implicativo (AEI) y automatizadas con el software CHIC.

Semejanzas con árboles de similitud

El árbol de similitud de la Figura 14 muestra la formación de una agrupación de dos variables y otra independiente. Esta última no presenta alguna correspondencia o semejanza

directa con las otras sublíneas de investigación de la carrera de Estadística Informática, mientras que la agrupación formada por las variables correspondientes a las sublíneas de investigación de Diseño de Experimentos y Análisis Multivariado se unen formando un nodo significativo, el cual muestra la calidad de agrupación de las variables y una clasificación compatible lo mejor posible con los valores, este enlace se manifiesta debido a que ambas variables guardan un alto grado de similitud en la aplicación de las sublíneas de investigación de la carrera de Estadística correspondiente a la ESPOCH.

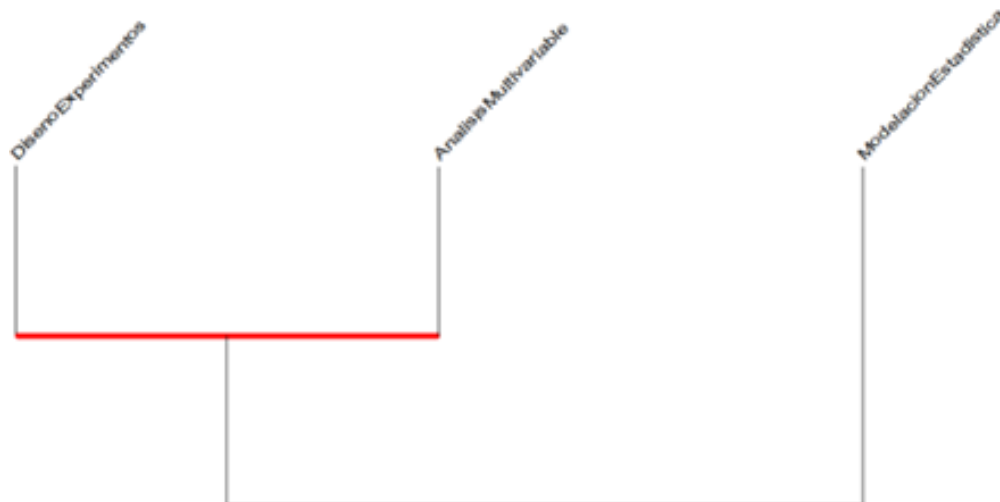


Figura 14: Similitud entre las sublíneas de investigación de la carrera de Estadística Informática (ESPOCH)

Dependencias con árboles de cohesión

En la Figura 15, se puede observar mediante el árbol cohesitivo la formación de dos agrupaciones de variables. La primera formada por la bicondicionalidad de las variables de Programa de Análisis Estadístico Implicativo y Computacional y Diseño de Experimentos, representada por T10SIA_low □ Diseño de Experimentos, lo cual nos indica que si se aplica la línea de investigación institucional T10SIA se aplica a la vez la línea de investigación de la

Carrera correspondiente a la ESPOCH - Diseño de Experimentos y viceversa. Esta relación de cohesión muestra el índice cohesitivo **más alto de formación con un 100%, además de ser un nodo significativo al nivel uno de jerarquía. Esta relación de bicondicionalidad es resultado de la condicionalidad de implicación consecutiva del Nivel de Cumplimiento de las líneas de la Facultad (Rtinst) en con un 98.4% y Análisis Multivariable con 14.2% de cohesión y la formación de un nodo significativo al nivel quinto de jerarquía.**

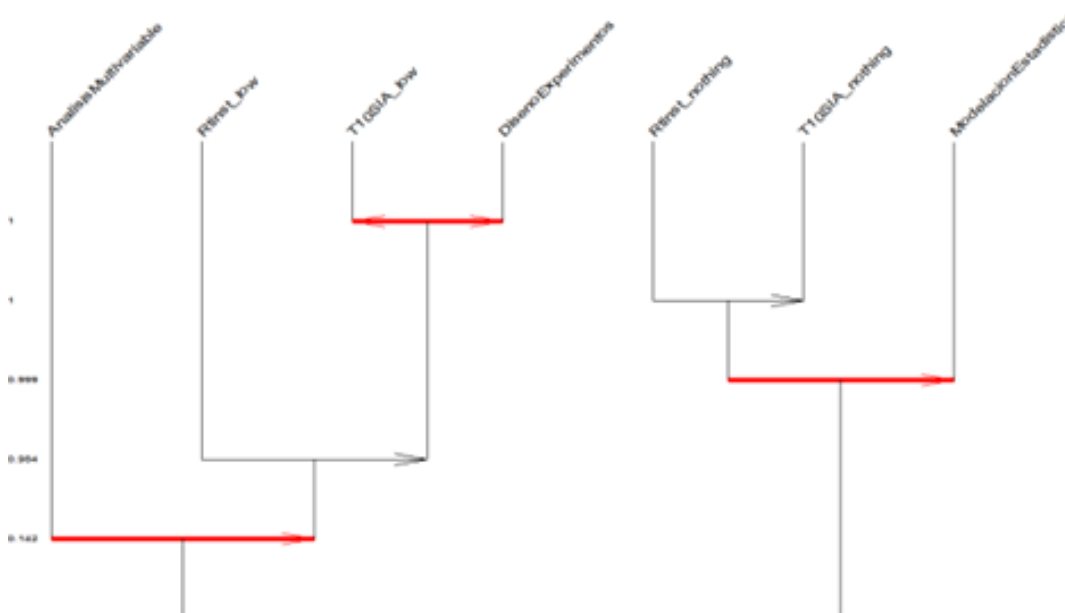


Figura 15: Cohesión entre sublíneas de investigación de la carrera de Estadística Informática, programas institucionales y nivel de cumplimiento

La segunda agrupación muestra la implicación del Nivel de Cumplimiento de las líneas de la Facultad y el Programa de Análisis Estadístico Implicativo y Computacional con un índice cohesitivo de 100% en el segundo nivel de jerarquía, y estos implican a su vez la utilización de Modelación Multivariable con un 99.9% y además la formación de un nodo significativo en el tercer nivel de jerarquía.

En la figura 16 se muestran dos agrupaciones, la primera representada por el nivel de cohesión más alto existente entre la línea de investigación institucional y la sublínea de investigación de la

carrera de estadística informática, formado por la implicación del Programa de Análisis Estadístico Implicativo y Computacional (T10SIA_nothing) y Modelación Estadística, con un índice cohesitivo de 99.9% y un nodo significativo en el primer nivel de jerarquía, esto se debe a que el análisis estadístico Implicativo hace un estudio exploratorio pero no está dentro del campo de la modelación estadística, debido a que no genera explícitamente el modelo. La segunda agrupación, formada por cuatro variables (indicadores): tres variables correspondientes a las líneas de investigación de la facultad de ciencias y la cuarta variable correspondiente a nivel de cumplimiento,

mostrando la relación cohesiva existente entre el Nivel de Cumplimiento de las líneas de la Facultad (Rtinst) y Diseño de experimentos con un 97.5%, resultante de la implicación con

la variable Ingeniería de procesos con un 85% además de ser un nodo significativo. Estas tres variables a su vez implican la aplicación de la variable análisis multivariable con un 14.2%.

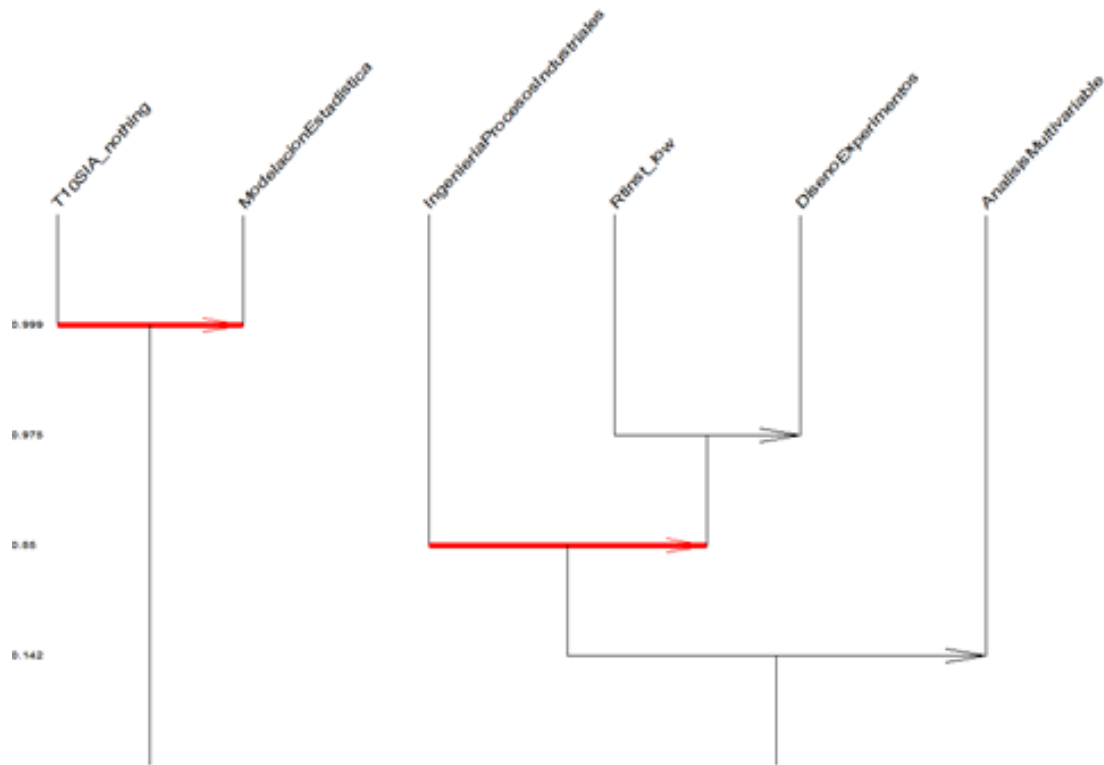


Figura 16: Cohesión entre sublíneas de investigación de la carrera de Estadística Informática, programas institucionales y nivel de cumplimiento

Dependencias mediante grafos implicativos

En el esquema izquierdo de la Figura 17 se puede observar que un trabajo que encaje con la línea de Modelación Estadística corresponde a un nivel de cumplimiento con las líneas de investigación bajo (esto es debido a que de los trabajos de titulación son pocos los que generan nuevos modelos más

bien solo los utilizan). Mientras que en el gráfico de la derecha se muestra que, si un trabajo en el cual se emplee la línea de investigación de la facultad - Ingeniería de Procesos industriales corresponde en un 100% a la aplicación de Diseño de Experimentos, con un nivel de cumplimiento alto con un nivel de confianza del 87.5%.



Figura 17: Implicación entre sublíneas de investigación de la carrera de Estadística Informática y nivel de cumplimiento

El esquema izquierdo de la Figura 18 muestra que la aplicación de la línea de investigación institucional-programa de Análisis Estadístico Implicativo y Computacional (T10SIA) corresponde en un 100% al cumplimiento de

las líneas de investigación institucionales y a la utilización de la línea de investigación de la carrera de Estadística Informática - Modelación Estadística.

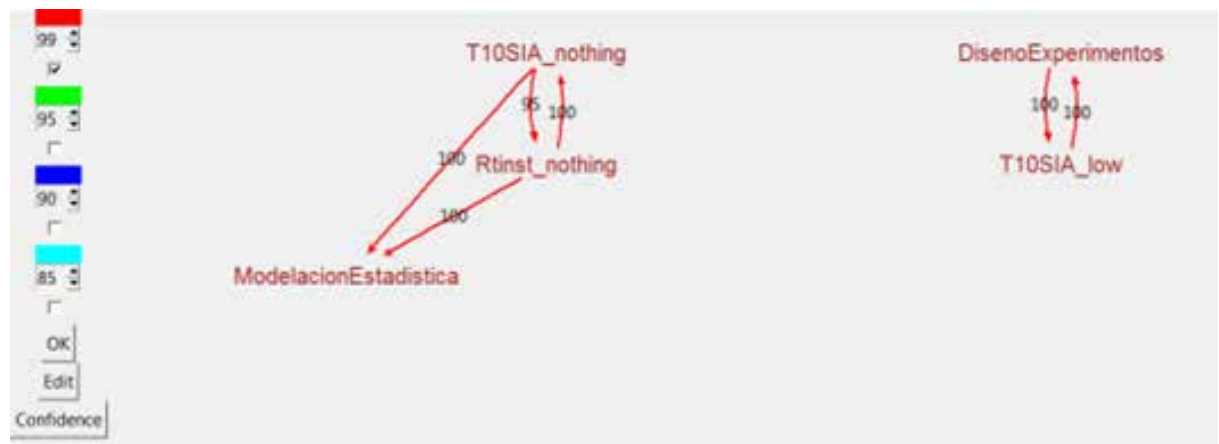


Figura 18: Implicación entre sublíneas de investigación de la carrera de Estadística Informática y nivel de cumplimiento

Se muestra además que, si un trabajo pertenece a la línea de diseño de experimentos, este empleará

un Programa de Análisis Estadístico Implicativo y Computacional (T10SIA) en un 100%.

Conclusiones

La gran mayoría de tesis son de la ESPOL pues inició a publicarlas 12 años antes que la ESPOCH y 17 años antes que la UCE, llama la atención que en 2017 la ESPOL publicó aproximadamente 250 tesis, el 33% de su producción total. Más de la mitad de los trabajos de titulación son en Análisis Multivariado, seguido de Investigación de operaciones y software estadístico. De los 776 trabajos de titulación, 760 de estas son trabajos estadísticos aplicados a otras áreas del conocimiento y apenas 16 son trabajos teóricos que contribuyen a la estadística como ciencia y se encuentran en las áreas de Estadística actuarial, Probabilidad y procesos estocásticos. Las tesis aplicativas tanto en la ESPOCH como en la ESPOL en su mayoría pertenecen al área de Análisis Multivariado y en la UCE al área de probabilidad y procesos estocásticos. La modelación estadística es la línea de investigación que más frecuencia tienen tanto la ESPOCH como la UCE, en cambio en la ESPOL es el análisis multivariado. Los trabajos de titulación que caen fuera de las líneas de investigación en la ESPOCH están contenidos dentro del área de Estadística descriptiva y Software Estadístico. Las líneas de investigación en diseño experimental y análisis multivariado se muestran similares debido a que ambas se utilizan como complementarias y el diseño experimental utilizado es multifactorial. Las líneas de análisis estadístico Implicativo y diseño experimental ambas están cohesionadas debido a que buscan relaciones de causalidad con diferentes teorías estadísticas. Las líneas de análisis estadístico

Implicativo no tiene cohesión con la modelación estadística pues su objetivo no es buscar un modelo, sino aplicar el modelo de cuasi-implicación. En la ESPOCH, si los trabajos de titulación están poco relacionados con las líneas de investigación de la facultad, también están lo están en el área de diseño experimental. El Análisis estadístico implicativo permite tratar fácilmente los trabajos de titulación y determinar su acercamiento a los diferentes programas y líneas de investigación presentados por el Instituto de Investigaciones de la Escuela Superior Politécnica de Chimborazo. Los investigadores pueden manejar la ventana de variables del software CHIC para una exploración más personalizada, detallada y profunda de tal forma que se puede encontrar relaciones de similaridad, cohesión e implicación que son de su mayor interés.

Agradecimientos

A la Escuela Superior Politécnica de Chimborazo que mediante el apoyo al proyecto “Análisis e innovación de las investigaciones en la facultad de ciencias como estrategia para incrementar la producción científica”, permite que se analice, mejore e incremente la investigación en la Facultad de Ciencias y en la ESPOCH. A los investigadores del grupo de investigación en Ciencia de Datos (Pazmiño-Maji and CIDED-ESPOCH), en especial a Rafaela Pacurucu, Paola Chiluzza y Paola Arguello por su apoyo en esta investigación. A los ayudantes de investigación que colaboraron en las primeras etapas de este proyecto.

Referencias bibliográficas

1. ESPOCH. Escuela Superior Politécnica de Chimborazo [Internet]. 2018. Available from: www.espoch.edu.ec
2. UCE. Universidad Central del Ecuador [Internet]. Available from: www.uce.edu.ec
3. ESPOL. Escuela Superior Politécnica del Litoral [Internet]. Available from: www.espol.edu.ec
4. ESPOCH. Actualización Curricular de la Carrera de Ingeniería en Estadística. 2016;
5. Raphael C, Rubén P. Use of Statistical Implicative Analysis in Complement of Item Analysis. *Int J Inf Educ Technol.* 2016;6(1):39–43.
6. Gras Regis. Analyse statistique implicative. Une méthode d'analyse de données à la recherche de causalités. 2009.
7. Couturier R. Cohesive hierarchical implicative classification. *Stat implicative Anal.* 2008;41–53.
8. Pazmiño R. Aproximación al Análisis Estadístico Implicativo desde sus Aplicaciones Educativas. 2014; Available from: <http://sired.udenar.edu.co/48/>
9. Team RC. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <https://www.r-project.org/>
10. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2009. Available from: <http://ggplot2.org>
11. Couturier R, Gras R. CHIC : traitement de données avec l'analyse implicative. *EGC.* 2005;679–84.
12. Pazmiño R. Análisis de indicadores de investigación. ESPOCH. 2017;
13. IDI. Plan de Investigación. 2014.

Correspondencia:**Autor:** Pablo Javier Flores Muñoz**Teléfono:** (593)958958295**Correo electrónico:** p_flores@espoch.edu.ec