

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

ANÁLISIS BIBLIOMÉTRICO DE LA PRODUCCIÓN CIENTÍFICA EN EL
SISTEMA DE EDUCACIÓN SUPERIOR ECUATORIANO

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERA MATEMÁTICA

PROYECTO DE INVESTIGACIÓN

JESSICA GENOVEVA MOROCHO GUERRA

morocho.jessica@gmail.com

DIRECTOR: MIGUEL ALFONSO FLORES SÁNCHEZ, MSc.

miguel.flores@epn.edu.ec

CODIRECTOR: LUIS ALCIDES HORNA HUARACA, PhD.

luis.horna@epn.edu.ec

Quito, Febrero 2018

DECLARACIÓN

Yo, JESSICA GENOVEVA MOROCHO GUERRA, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

JESSICA GENOVEVA MOROCHO GUERRA

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por **JESSICA GENOVEVA MOROCHO GUERRA**, bajo mi supervisión.

MIGUEL ALFONSO FLORES SÁNCHEZ, MSC.
DIRECTOR

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por JESSICA GENOVEVA MOROCHO GUERRA, bajo mi supervisión.

LUIS ALCIDES HORNA HUARACA, PHD.
CODIRECTOR

AGRADECIMIENTOS

- *A Dios y a mi Madre del Cisne, por darme esperanza y fortaleza cuando más sola me encontré. Gracias por enseñarme a renacer de las cenizas.*
- *A mi madre, porque juntas somos el ejemplo claro de lo que significa aprender a vivir. Lidia, sin ti no sería lo que soy. Tú y yo somos el resultado del valor y del dolor. Gracias por quererme y cuidarme de la única manera que conoces. Gracias por siempre regresar.*
- *A mi mejor amigo, por ser como un hermano para mí. Cristhian, mi destino fue encontrarte, mi decisión quererte. Gracias por abrirme tu corazón. Gracias por estar a mi lado cuando me es difícil incluso estar conmigo. Gracias por darle a mi vida un toque de realidad; a donde la vida te lleve, mi corazón te acompañará.*
- *A Diana, por afrontar juntas los altos y bajos de una lucha contra el tiempo.*
- *A Miguel Flores, por siempre apoyarme. Gracias por los consejos y observaciones. Gracias por tu amistad.*
- *A Luis Horna, por sus sugerencias y comentarios. Gracias por su paciencia y tiempo.*
- *A Mauricio y Alex, amigos que se encuentran lejos físicamente pero cerca emocionalmente.*
- *A todos mis profesores, por transmitirme sus conocimientos y hacer que ame con el corazón mi carrera.*
- *A los amigos de universidad, que compartieron conmigo un pedazo de sus historias.*
- *A las personas que llegaron y se fueron, ya que cada una fue parte de este largo camino llamado vida.*

DEDICATORIA

*A Lidia. Por el simple y llano hecho de ser mi madre.
A la vida. Porque todo llega cuando tiene que llegar y por su inmensa capacidad de
poner todo, absolutamente todo, en su lugar.
No existe futuro sin pasado.*

Jessica

Índice general

Índice general	IX
Lista de figuras	XII
Lista de cuadros	XIII
Resumen	1
Abstract	3
Introducción	5
1 Antecedentes y Aspectos Generales.	8
1.1 Inicios de la Bibliometría.	8
1.2 Definiciones Generales.	10
1.3 Leyes Bibliométricas.	11
1.3.1 Reglas para el desarrollo bibliométrico.	11
1.3.1.1 Regla de Oro (Rule of Thumb) o Ley de Lotka.	11
1.3.1.2 Ley de Bradford.	12
1.3.1.3 Ley de Garfield.	12
1.3.1.4 Ley de Zipf.	12
1.3.1.5 Ley del envejecimiento u obsolescencia de la literatura cien- tífica.	13
1.3.1.6 Ley de visibilidad.	13
1.3.1.7 Principio de la ventaja acumulativa.	13
1.4 Modelos de crecimiento de producción científica.	13
1.5 Indicadores Bibliométricos.	14
1.6 Bibliometría en Latinoamérica.	14
1.7 Bibliometría en el Ecuador.	15

2	Metodología en Estudios Bibliométricos.	17
2.1	Metodología General.	17
2.2	Metodología utilizada en el estudio.	19
2.3	Metodología con particularización a una IES.	21
3	Teoría Bibliométrica.	23
3.1	Indicadores Bibliométricos.	24
3.1.1	Indicadores de Producción.	24
3.1.1.1	Número total de Publicaciones (PC).	24
3.1.2	Indicadores de Visibilidad e Impacto.	25
3.1.2.1	Número Total de Citas (CT).	25
3.1.2.2	Promedio de citas por documento científico (CD).	25
3.1.2.3	Índice-H (H).	26
3.1.3	Indicadores de Colaboración.	28
3.1.3.1	Número total de publicaciones con colaboración internacional (DCI).	28
3.1.3.2	Colaboración Internacional.	28
3.1.3.3	Indicadores en el presente estudio.	29
3.2	Leyes Bibliométricas.	29
3.2.1	Ley de Lotka.	29
3.2.1.1	Formulación Original (Clásica).	29
3.2.1.2	Formulación Generalizada.	30
3.2.1.3	Usos de la Ley de Lotka en los estudios bibliométricos.	31
3.2.1.4	Lotka en el presente estudio.	31
3.3	Modelos de Crecimiento.	32
3.3.1	Modelo de Crecimiento Exponencial.	32
3.3.1.1	Modelos de crecimiento en el presente estudio.	34
4	Extracción y Depuración de Datos.	35
4.1	Universo de Estudio.	35
4.2	Consideración para el procedimiento de extracción y depuración.	37
4.3	Procedimiento de Extracción.	37
4.4	Consolidación de la base inicial.	41
4.5	Procedimiento de Depuración.	42
4.5.1	Selección de datos que se van a considerar en el análisis.	43
4.5.2	Corrección de datos cargados en Scopus con errores.	43

4.5.3	Eliminación de datos duplicados (generados por mecanismo de consulta)	44
4.5.4	Consolidación de citas por artículo.	45
5	Resultados Obtenidos.	47
5.1	Análisis de los indicadores bibliométricos.	47
5.1.1	Número total de Publicaciones.	47
5.1.2	Número Total de Citas.	52
5.1.3	Colaboración Internacional.	55
5.1.3.1	Citación y Colaboración.	57
5.1.4	Índice H.	57
5.1.4.1	*Categoría A.	59
5.1.4.2	*Categoría B.	61
5.1.4.3	*Categoría C.	62
5.2	Áreas de Mayor Producción Científica.	63
5.3	Análisis y prueba de las Leyes Bibliométricas.	65
5.3.1	Crecimiento de la producción: Price.	65
5.3.2	Modelo de Price en las Categorías del CEAACES.	67
5.3.2.1	Modelo de Price en la Categoría A.	69
5.3.2.2	Alternativas para las IES de la Categoría A que no cumplen con Price.	71
5.3.3	Ley de Lotka. Índice de Productividad Personal.	73
5.3.3.1	*Categoría A.	73
5.3.3.2	*Categoría B.	74
5.3.3.3	*Categoría C.	74
5.3.3.4	Productividad Personal para la EPN.	75
6	Conclusiones y Recomendaciones.	77
6.1	Conclusiones.	77
6.2	Recomendaciones.	83
6.2.1	Fuente de Datos: Situaciones Observadas a la fecha de corte.	83
6.2.2	Futuros estudios.	86
A	Anexos	87
A.1	Producción Científica IES por año	87
A.2	Producción según el tipo de documento por año.	88

A.3 Citas por IES.	88
A.4 Citas y Frecuencia de documentos.	89
A.5 Programación en R.	89
Bibliografía	129

Índice de figuras

4.1	Pantalla Inicial Scopus	38
4.2	Ingreso Nombre IES	39
4.3	Resultado búsqueda IES	39
4.4	Documentos IES	39
4.5	Datos a exportar	40
4.6	Errores en país de afiliación.	43
4.7	Corrección errores país de afiliación.	44
5.1	Producción Científica 2010-2017*.	47
5.2	Producción Científica IES-NoIES 2010-2016.	48
5.3	Comportamiento Producción Científica IES-NoIES 2010-2016.	48
5.4	Número de autores dentro de la producción científica 2010-2016.	49
5.5	Producción Tipo de Documento 2010-2016.	49
5.6	Producción Científica Cat. A.	50
5.7	Producción Científica Cat. B.	51
5.8	Producción Científica Cat. C.	51
5.9	Citaciones por año 2010-2017*.	52
5.10	Citaciones por año 2010-2016.	52
5.11	Citaciones vs Frecuencia de documentos.	53
5.12	Citaciones (≤ 300) vs Frecuencia de documentos.	54
5.13	Concentración Citaciones vs Autores	54
5.14	Colaboración Nacional - Extranjera 2010-2016.	55
5.15	Concentración de Colaboraciones por año: 2010-2017*.	56
5.16	Colaboraciones 2010-2016.	56
5.17	Citación vs Colaboración.	57
5.18	H-index. Nivel: IES.	58
5.19	H-index: Impacto Sistema de Educación Superior.	58
5.20	H-index: Categoría A (subperíodos).	59

5.21 H-index: Categoría A (período de análisis).	59
5.22 H-index: Impacto Categoría A.	60
5.23 H-index: Categoría B (subperíodos).	61
5.24 H-index: Categoría B (período de análisis).	61
5.25 H-index: Impacto Categoría B.	62
5.26 H-index: Categoría C (subperíodos).	62
5.27 H-index: Categoría C (período de análisis).	63
5.28 H-index: Impacto Categoría C.	63
5.29 Documentos por áreas de producción 2010-2016.	64
5.30 Áreas de alta producción 2010-2016.	65
5.31 Datos reales PC.	66
5.32 Ajuste de la producción: Modelo de Price.	66
5.33 PC Real vs PC Ajustado.	67
5.34 Proyecciones: Modelo de Price.	67
5.35 Comportamiento Producción Científica Categoría A 2010-2016.	68
5.36 Comportamiento Producción Científica Categoría B 2010-2016.	68
5.37 Comportamiento Producción Científica Categoría C 2010-2016.	68
5.38 Comportamiento Producción Científica 2010-2016.	70
5.39 Proyecciones: Modelo de Price EPN, ESPOL, UCUENCA, ESPE.	71
5.40 Ajuste de la producción: Modelo lineal en el período de análisis.	71
5.41 Proyecciones UEES	72
5.42 Ajuste de la producción: Modelo exponencial en el subperíodo.	72
5.43 Proyecciones USFQ	73
5.44 Lotka. Categoría A.	73
5.45 Lotka. Categoría B.	74
5.46 Lotka. Categoría C.	74
5.47 Composición Productores EPN.	75
5.48 Grandes Productores EPN.	76
6.1 Detalle criterio Scopus.	83
6.2 Documento en la base B_IES_C2*.	84
6.3 Consulta en la base E_TOT.	84
6.4 Documento asociado a un autor de la EPN.	85
6.5 Consulta del documento en la IES.	85
6.6 Consulta en la base E_TOT.	85

A.1 Producción Científica IES por Año: 2010-2017*	87
A.2 Producción según el tipo de documento por año: 2010-2017*	88
A.3 Citación por IES y año: 2010-2017*	88
A.4 Citas y Frecuencia de documentos.	89

Índice de cuadros

4.1 Instituciones Acreditadas A.	36
4.2 Instituciones Acreditadas B (Actualizado 05-06-2017).	36
4.3 Instituciones Acreditadas C (Actualizado 05-06-2017).	37
4.4 IES con datos en Scopus.	41
4.5 Variables de Scopus.	42
4.6 Variables a considerarse en el estudio.	43
5.1 Resultados ajuste exponencial por categorías.	69
5.2 Resultados ajuste exponencial.	70

Resumen.

En el presente trabajo, se propone analizar el comportamiento de la Producción Científica generada por las IES ecuatorianas (Instituciones de Educación Superior), en el período 2010-2016 a través del uso de indicadores y leyes propias de la Bibliometría, con el fin de conocer el estado de la Investigación Científica en el Ecuador.

La Bibliometría, constituye de manera general, la aplicación de métodos estadísticos y matemáticos a la literatura científica (representante de la actividad investigativa); para lo cual, utiliza indicadores y leyes determinadas por el comportamiento evidenciado a lo largo del tiempo en el desarrollo de la Ciencia.

Los indicadores, conocidos como Indicadores Bibliométricos, constituyen los instrumentos de medición que permiten obtener información de la actividad científica.

Las leyes, conocidas como Leyes Bibliométricas, constituyen reglas básicas de la Bibliometría fundamentadas en el comportamiento estadístico regular evidenciado a través del tiempo.

Los indicadores bibliométricos a utilizarse serán de 3 tipos: de producción, de visibilidad e impacto y de colaboración siendo uno de los indicadores de visibilidad e impacto más relevantes el índice-H o H-index.

El H-index, constituye una función de impacto definida como el máximo de los índices asociados a los estadísticos de orden de las citaciones de todos los documentos científicos generados por el autor tal que dicho estadístico de orden es mayor o igual al índice asociado. La importancia de este indicador radica en su capacidad de abarcar tanto productividad como calidad del autor.

Las leyes bibliométricas a revisarse son la Ley de Lotka y la Ley de Price.

La principal utilidad de la Ley de Lotka dentro de este estudio, es su uso para determinar el índice de productividad personal, el cual, permite una clasificación según el nivel de la producción del autor.

La principal utilidad de la Ley de Price dentro de este estudio, es su uso para establecer proyecciones de la producción científica, lo cual, permite conocer cómo será el comporta-

miento futuro de acuerdo a su tendencia evolutiva.

Para esto, se establece una metodología de obtención, limpieza y análisis de datos que permita desarrollar una investigación bibliométrica a nivel Meso, siendo las universidades y escuelas politécnicas del Sistema de Educación Superior categorizadas por el CEAACES las unidades de estudio; es decir, se considera como autores de los documentos científicos a las IES ecuatorianas.

Los datos se obtienen de Scopus, una de las bases indexadas más reconocidas dentro de la comunidad científica.

Como uno de los principales resultados se tiene que la Producción Científica ecuatoriana se encuentra en un momento de crecimiento exponencial donde las IES se encuentran fomentando relaciones colaborativas nacionales y extranjeras.

La forma predilecta de publicación es el artículo científico y las áreas de más interés dentro del período de análisis son las Ciencias de la Computación y Física y Astronomía sin dejar de lado las áreas de Agricultura y Medicina.

Finalmente, el presente estudio cuenta con un código de programación en R que puede ser particularizado según los intereses de futuros lectores.

PALABRAS Y CONCEPTOS CLAVES: Bibliometría, Indicadores Bibliométricos, Leyes Bibliométricas, Scopus, Expresiones Regulares, Modelo de Lotka, Modelo de Price.

Abstract.

In this work, we propose to analyze the behavior of Scientific Production generated by the Ecuadorian Higher Education Institutions, in the period 2010-2016 through the use of indicators and laws specific from Bibliometrics, in order to know the status of Scientific Research in Ecuador.

Bibliometrics, constitutes in a general way, the application of statistical and mathematical methods to the scientific literature (representative of the research activity); for which, uses indicators and laws determined by the behavior evidenced throughout of time in the development of Science.

The indicators, known as Bibliometric Indicators, constitute the instruments of measurement that allow to obtain information of the scientific activity.

The laws, known as Bibliometric Laws, constitute basic rules of Bibliometrics based on the regular statistical behavior evidenced throughout of time. The bibliometric indicators to be used will be of 3 types: production, visibility and impact and collaboration being one of the most relevant indicators of visibility and impact the H-index.

The H-index, constitutes an impact function defined as the maximum of the indexes associated with the order statistics of the citations of all the scientific documents generated by the author such that said order statistic is greater than or equal to the associated index. The importance of this indicator lies in its ability to cover so much productivity as the author's quality.

The bibliometric laws to be reviewed are the Law of Lotka and the Law of Price. The main utility of Law of Lotka within this study is its use to determine the personal productivity index, which allows a classification according to the level of author's production.

The main utility of the Law of Price within this study is its use to establish projections of scientific production, which allows us to know what the future behavior according to its evolutionary trend.

For this, a methodology of obtaining, cleaning and analyzing data is established that allows to develop a Meso-level bibliometric research, being the universities and polytechnic schools

of the Higher Education System categorized by CEAACES the study units; that is, they are considered as authors of the scientific documents.

The data is obtained from Scopus, one of the most recognized indexed databases within the scientific community.

As one of the main results we have that the Ecuadorian Scientific Production is in a moment of exponential growth where Ecuadorian Higher Education Institutions are fostering collaborative national and foreign relations.

The preferred form of publication is the scientific article and the areas of most interest within of the period of analysis are the Sciences of the Computation and Physics and Astronomy without leaving aside the areas of Agriculture and Medicine.

Finally, the present study has a programming code in R that can be particularized according to the interests of future readers.

KEYWORDS AND CONCEPTS: Bibliometrics, Bibliometric Indicators, Bibliometric Laws, Scopus, Regular Expressions, Lotka Model, Price Model.

Introducción.

En el año 2011, el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES) inició de manera formal sus actividades como organismo de derecho público; con la función de “ejercer la rectoría de la política pública para el aseguramiento de la calidad de la educación superior del Ecuador a través de procesos de evaluación, acreditación y categorización en las Instituciones de Educación Superior (IES) y ser un referente nacional y regional en la creación e implementación de metodologías integrales, articuladas y transparentes de evaluación, acreditación y aseguramiento de la calidad de la educación superior” [1].

Es así como, a partir de ese año, las Instituciones de Educación Superior (IES) del Ecuador, han sido sometidas a minuciosas evaluaciones académicas, en su estructura física, estructura financiera y organización, con la finalidad de evaluar su calidad de manera integral. Actualmente los procesos de evaluación del CEAACES tienen periodicidad quinquenal, de manera que, cada una de las universidades y escuelas politécnicas del Sistema de Educación Superior, debe cumplir los plazos estipulados ya sea para mantener su categoría u obtener una recategorización según lo estipulado en la Ley Orgánica de Educación Superior (LOES). Una universidad o escuela politécnica debe ser acreditada de manera obligatoria; el cumplimiento de los estándares de calidad determinados por el CEAACES, y la mejora continua, le permitirán a dicha institución mantenerse vigente en el sistema educativo.

Uno de los criterios fundamentales dentro de la evaluación institucional es la Investigación dentro de la IES, siendo la Producción Científica una de las actividades más representativas de este criterio.

De hecho, se tiene que el peso asignado al indicador **Producción Científica** del modelo utilizado por el CEAACES para la categorización, es del 9% [40].

El efecto asociado a la producción científica no es solo de importancia individual a nivel de institución; de hecho, la producción científica es una actividad que si se analiza dentro del marco de la educación superior en el país, permite conocer el status del mismo frente a sus pares regionales.

En este punto resulta evidente que si la producción científica es un factor importante al momento de determinar la calidad de una IES, es también un factor importante en el crecimiento de un país debido a la relación directa entre la educación y la creación de nuevas oportunidades para el desarrollo económico [2].

Según SCImago Research Group, en el 2016 al Ecuador le pertenecía el 0.034 % de la producción científica mundial, la cual es mil veces menos que la producción asociada a países desarrollados tales como Estados Unidos; y comparando con países de Latinoamérica se tiene que el país aún se encuentra muy por detrás de los niveles regionales de producción científica [6].

En el presente trabajo se busca analizar como es el comportamiento de la Producción Científica dentro de las universidades del Ecuador y obtener un estado real de la investigación a nivel país a través de los conceptos asociados a la Bibliometría, rama de la Cienciometría que tiene como unidad de análisis el artículo científico [3].

Debido a que las IES se hallan en la obligación de realizar un seguimiento de los resultados obtenidos de la Investigación a través de la producción científica para obtener su acreditación en el sistema educativo ecuatoriano, el uso de los métodos cuantitativos de la Bibliometría es una opción atractiva al momento de evaluar programas de investigación, la eficiencia y eficacia de su implementación, determinación del estado de cumplimiento de objetivos y recomendación en caso de ajustes [3].

Los indicadores bibliométricos a utilizarse lo constituyen las tradicionales funciones de impacto de la Bibliometría tales como productividad y citación de los autores, el H-index como una función de impacto mucho más especializada para la evaluación de la calidad de la producción y los indicadores de colaboración que permiten reflejar los intereses y las relaciones de contribución entre los autores.

Para las leyes bibliométricas a revisarse, se abordará de manera inicial, el ajuste de la producción científica a las leyes de Lotka y de Price. Para la verificación del ajuste de los datos a la Ley de Lotka se utilizará la prueba de bondad de ajuste de Kolmogórov-Smirnov entre la distribución teórica o clásica de Lotka y los datos, mientras que para la Ley de Price, que postula que el comportamiento de la producción científica sigue una distribución exponencial, se utilizará la prueba de bondad de ajuste de Kolmogórov-Smirnov entre la distribución teórica exponencial y los datos y el cálculo del coeficiente de determinación R^2 . El ajuste a la Ley de Lotka permitirá determinar el índice de productividad personal de los autores afiliados a las IES, es decir, el uso de la Ley de Lotka permite una clasificación a nivel interno de cada IES.

El ajuste a la Ley de Price permitirá determinar los valores proyectados de la producción científica de las IES para el período 2017-2020, es decir, el uso de la Ley de Price permite obtener una visualización futura de los niveles de producción.

Es importante mencionar que el presente trabajo expone los lineamientos metodológicos replicables al momento de realizar una investigación bibliométrica, debido a que los capítulos 3, 4, 5 y 6 se consideran etapas fundamentales en el desarrollo de este tipo de estudios; los cuales, a nivel mundial están tomando fuerza e interés a causa del incremento general en la producción de información.

Finalmente, el código en R presentado en el Anexo [A.5](#), permite replicar el proceso práctico de un estudio bibliométrico desde la obtención de datos hasta el desarrollo de los indicadores y las leyes a utilizarse.

El presente trabajo se halla estructurado de la siguiente manera:

- En el capítulo [1](#) se trata los antecedentes y aspectos generales de la Bibliometría,
- En el capítulo [2](#) se presenta la metodología del estudio,
- En el capítulo [3](#) se detalla la Teoría Bibliométrica a usarse en el estudio,
- En el capítulo [4](#) se describe la extracción y pre-procesamiento de los datos a utilizarse,
- En el capítulo [5](#) se analizan los resultados obtenidos,
- Y en el capítulo [6](#) se realiza las conclusiones y recomendaciones del estudio con base en los resultados obtenidos.

Capítulo 1

Antecedentes y Aspectos

Generales.

En el presente capítulo se expone una visión rápida y general de la Bibliometría como rama de la Ciencia, lo cual permitirá al lector familiarizarse con la misma y comprender de mejor manera aspectos básicos para desarrollar un estudio bibliométrico general.

En este punto, es importante mencionar que **tanto el enfoque de desarrollo, el tipo de estudio y las definiciones a utilizarse en el presente trabajo se detallarán de manera específica en el capítulo 2.**

Dicho de otra manera, se recomienda al lector considerar al presente capítulo como un capítulo introductorio y al capítulo 2 como el capítulo principal del presente trabajo.

1.1 Inicios de la Bibliometría.

A pesar de que la primera aparición oficial de los términos Cienciometría y Bibliometría ocurrió en 1969, el análisis estadístico de las publicaciones científicas ya tenía más de 40 años de desarrollo.

Se conoce que el primer estudio (posteriormente llamado de tipo bibliométrico) fue desarrollado y publicado por **Alfred J. Lotka** en 1926. Dicho estudio trataba sobre la *Distribución de frecuencia de la productividad científica determinada a partir de un índice decenal (1907- 1916) de los Abstracts de Ciencias Químicas* [4].

La conclusión de dicho estudio es lo que se conoce actualmente dentro de la Bibliometría como *Ley de Lotka*.

En 1934, **Samuel C. Bradford** publica su estudio de la distribución de frecuencias de los artículos en las revistas científicas y concluye lo hoy llamado dentro de la Bibliometría co-

mo *Ley de Bradford*.

En 1949, **George K. Zipf** formula lo conocido como *Ley de Zipf* o *Principio del menor esfuerzo*, concluida con base en sus estudios de las frecuencias de las palabras en un texto y que representa “la relación entre el rango de las palabras y la periodicidad dada en una porción de la literatura determinada” [9].

En 1963, **Derek de Solla Price** cambia radicalmente el nivel de desarrollo de la Bibliometría y “establece las bases de las técnicas modernas de evaluación de la investigación”. Su libro, *Little Science-Big Science*, presenta “el primer enfoque sistemático de la estructura de la ciencia moderna aplicada a la ciencia, con base en el sistema de comunicación científico de la época” [4].

La coyuntura de la época estaba cambiando profundamente; la comunicación científica se veía afectada de manera positiva por la globalización, el crecimiento del conocimiento fomentó la publicación, se evidenció una creciente influencia de la interdisciplinariedad y especialización, dando lugar al desarrollo de la mayoría de modelos básicos para la comunicación [4].

En 1955, **Eugene Garfield** evidenció la importancia de la citas dentro de los análisis de comunicación científica e impulsó el desarrollo de estudios basados en citación. A la par, postuló lo conocido como *Ley de Garfield* [8].

En 1964, **W. Goffman y V. A. Newill** introducen “el modelo de epidemias intelectuales como modelo de comunicación científica” [4].

En 1969, **Alan Pritchard** definió a la Bibliometría como: “La aplicación de métodos estadísticos y matemáticos a libros y otros medios de comunicación” [3].

En el mismo año, **Nalimov y Mulchsenko** definen a la Cienciometría como “La aplicación de métodos cuantitativos que se ocupan del análisis de la ciencia visto como un proceso de información” [4].

El desarrollo de la Cienciometría y la Bibliometría estuvo a cargo de investigadores que inicialmente lo consideraron un pasatiempo debido a que la recopilación de datos en los años sesenta (época de auge para la Cienciometría y la Bibliometría) era una tarea manual; luego, se integraron modelos matemáticos, físicos, sociológicos y psicológicos con diversos enfoques.

En los años ochenta se evidenció lo siguiente [4]:

- La Bibliometría evolucionó “en una disciplina científica con un perfil de investigación específico, varios subcampos y sus respectivas estructuras de comunicación científica”.

- Se publica la primera revista especializada en temas bibliométricos llamada *Scientometrics*.
- Las Bases de Datos empezaban a ser más claras en su escritura y contexto debido al desarrollo de la tecnología.
- Tanto la *Cienciometría* como *Bibliometría* empezaron a conocerse a nivel internacional dentro de la comunidad científica debido a la publicación de libros. Dentro de los autores se encontraron: “Haitun (1983), Ravichandra Rao (1983), Bujdosó (1986), Van Raan (1988), Egghe y Rousseau (1990), y Courtial (1990)”.

Ya para los años noventa, los métodos bibliométricos eran aplicados a la *Bibliometría* en sí. Como podemos observar, la *Cienciometría* y la *Bibliometría* son relativamente nuevas dentro de la Investigación Científica, hecho que no ha influido en su desarrollo vertiginoso.

1.2 Definiciones Generales.

Como se mencionó anteriormente, los términos *Cienciometría* y *Bibliometría* han ido de la mano desde el momento de su introducción en la comunidad científica, y aunque para algunos, estos términos pueden ser considerados sinónimos, dentro de la práctica, su diferencia radica en el enfoque del campo.

La **Cienciometría** constituye “el estudio cuantitativo de las actividades desarrolladas dentro de la investigación científica y técnica” mientras que la **Bibliometría** “es la rama de la *Cienciometría* enfocada en el estudio de la actividad científica tomando como base las publicaciones científicas, es decir, el artículo científico” [3].

Los **artículos científicos o publicaciones** constituyen el principal resultado y medio de comunicación de las actividades científicas desarrolladas por un país, una institución o un investigador, ya que son el resultado de un proceso minucioso de trabajo investigativo.

Se define como **Indicadores Bibliométricos** a los cálculos o medidas que permiten obtener información cuantitativa y objetiva al ser aplicados en los resultados de la actividad científica.

*De manera compacta: La **Bibliometría** “aplica métodos matemáticos y estadísticos a toda la literatura de carácter científico y a los autores que la producen, con el objetivo de estudiar y analizar la actividad científica. Para ello, se apoya en leyes bibliométricas, basadas en el comportamiento estadístico regular que a lo largo del tiempo han mostrado los diferentes elementos que forman parte de la Ciencia siendo los indicadores bibliométricos, los instrumentos utilizados para medir los aspectos de este fenómeno social” [5].*

Los **estudios bibliométricos** se basan en “el supuesto de que la mayoría de los descubrimientos científicos y resultados de investigaciones son publicados en revistas científicas donde pueden ser leídos y citados por otros investigadores” [3].

Actualmente existen tres **enfoques de la investigación bibliométrica** [3]:

- *Investigación en metodología para Bibliometría.*- Investigación básica que ha desarrollado los indicadores, modelos y metodologías a usarse en un estudio bibliométrico.
- *Investigación bibliométrica de disciplinas científicas.*- Investigación que permite la aplicación de metodologías bibliométricas al estudio de las publicaciones científicas.
- *Investigación bibliométrica para gestión y políticas.*- Investigación que evalúa la investigación expresada en publicaciones y que permite establecer los alcances de los resultados compartidos en la comunidad científica. Actualmente, este tipo de investigación tiene un fuerte interés dentro del área de Salud.

Los enfoques anteriormente mencionados permiten conocer el estado de: productividad, volumen, visibilidad a través de citación, investigadores dominantes, ramas de investigación dominantes, etc. dentro de la comunidad científica, lo que permite cumplir con el **objetivo práctico de la Bibliometría** el cual es: *permitir planificar de manera eficaz y eficiente las actividades asociadas a la Investigación y de esta manera gestionar la política científica* [7].

1.3 Leyes Bibliométricas.

1.3.1 Reglas para el desarrollo bibliométrico.

Los siguientes resultados constituyen **Leyes** establecidas a lo largo del tiempo dentro de la Bibliometría.

1.3.1.1 Regla de Oro (Rule of Thumb) o Ley de Lotka.

Llamada también **Ley de la Productividad**, fue concluida por Lotka en 1926 para predecir el crecimiento de la publicación científica; postula de manera simplificada que “muy pocos autores publican la mayoría de artículos en un tema y que para una productividad alta, el número de autores va disminuyendo” [3].

A continuación se presentan una formulación verbal:

- *“El número de autores haciendo n contribuciones es de aproximadamente $1/n^2$ de aquellos que están haciendo uno.* [4].

1.3.1.2 Ley de Bradford.

Llamada también **Distribución de Bradford, Ley de dispersión o núcleo y dispersión**, fue concluida por Bradford en 1934. A continuación se presentan 2 formulaciones verbales :

- *“Si las revistas científicas están dispuestas en orden decreciente de la productividad en un determinado tema, éstas pueden dividirse en un núcleo de revistas más particularmente dedicadas al tema y varios grupos o zonas que contengan el mismo número de artículos que el núcleo cuando el número de publicaciones periódicas en el núcleo y en las zonas siguientes sea de la forma $1 : b : b^2 \dots$ ”* [4].
- *“Si las revistas se agrupan en orden descendente a su productividad de artículos sobre un tema determinado, éstas pueden ser divididas en un núcleo de revistas que publican particularmente sobre un tema y otros grupos o zonas que contienen el mismo número de artículos que el núcleo pero que incluyen para reunir ese mismo número de artículos un mayor número de revistas”* [12].

1.3.1.3 Ley de Garfield.

También llamada **Ley de la concentración de Garfield**, es considerada una extensión de la Ley de Bradford, postula que: “para cualquier campo de la ciencia, los artículos se concentran en las mismas revistas multidisciplinarias de alto impacto o de ‘corriente principal’, ya que estas revistas ‘núcleo’ de una disciplina son parte de la ‘cola de cometa’ de otra disciplina”, lo que permitiría calcular el número de revistas necesarias para cubrir disciplinas y sus interrelaciones [8].

1.3.1.4 Ley de Zipf.

Algunos autores consideran a esta ley como una generalización de las Leyes de Lotka y Bradford [4] y postula “que la parte más importante de cualquier texto, independientemente de la lengua en que esté impreso está constituido por unas escasas palabras de uso diario; en el resto afloran decenas de miles de palabras que se manejan escasas veces” [9]. Actualmente, esta ley es de uso limitado e incluso nulo debido a [4]:

- Los métodos tradicionales de recuperación de la información ya no son suficientes.
- Las interrelaciones en los fenómenos estudiados por Zipf ya no tienen vigencia.

1.3.1.5 Ley del envejecimiento u obsolescencia de la literatura científica.

A través de sus estudios de distribución de referencias, Price constató y postuló que “la curva de envejecimiento de la literatura científica es también exponencial” y que no todas las especialidades o ciencias envejecen de igual manera [5].

Esta idea, respecto al tiempo de vigencia de los artículos científicos, permitió a Burton y Kebler, en 1960 crear el concepto de **semiperíodo**.

1.3.1.6 Ley de visibilidad.

Esta ley determina que los artículos que no son difundidos dentro de la comunidad científica (en el sentido de que estén disponibles y se conozca su existencia dentro de la comunidad) constituyen literatura gris [5].

1.3.1.7 Principio de la ventaja acumulativa.

Formulada verbalmente por Price en 1976 de la siguiente manera:

“El éxito parece generar éxito. Un documento que ha sido citado muchas veces es más probable que se cite de nuevo que uno que ha sido poco citado. Un autor de muchos artículos tiene más probabilidades de publicar de nuevo que uno que ha sido menos prolífico. Una revista que ha sido consultada con frecuencia para algún propósito es más probable que se vuelva a ser consultada que una de uso poco frecuente” [4].

1.4 Modelos de crecimiento de producción científica.

Dentro de la Bibliometría existen 3 tipos de modelos que explican cómo se comporta la producción científica:

- **Modelo Exponencial.**- También conocido como **Ley de crecimiento exponencial**, postula que el crecimiento de la literatura científica sigue una distribución exponencial. Fue enunciado en 1956 por D. J. de Solla Price y se encuentra asociado al fenómeno de crecimiento ocurrido luego de la Segunda Guerra Mundial llamado **explosión de la información** [5].
- **Modelo Logístico.**- Propuesto por Price. Es una modificación al modelo exponencial y postula que “la literatura en un determinado tema no crece hasta el infinito; el gradiente de crecimiento y producción decrece e incluso se vuelve negativo debido a que en un momento dado los recursos se vuelven escasos” [3].

- **Modelo de Epidemias Intelectuales.**- Propuesto por Goffman-Newill. En este modelo se compara la difusión de las ideas asociadas a temas específicos dentro de la comunidad científica con la diseminación del virus de la influenza causando una epidemia. De manera simplificada, “una línea de investigación novedosa o artículo científico (infección) se difunde rápidamente entre los investigadores que no habían investigado dicho tema (susceptibles), hasta que dichos investigadores (infectados) empiezan a investigar en dicha línea produciendo artículos relacionados (diseminación de la infección), llevando a un punto en que se alcanza rápidamente la mayor producción científica (epidemia)” [3]. Este modelo puede ser usado para describir y predecir: describir propagación y predecir el momento en que la infección alcanza su pico, tras lo cual disminuye [4].

1.5 Indicadores Bibliométricos.

Los indicadores bibliométricos “constituyen una de las herramientas más utilizadas para la medición del producto de la investigación científica, porque la documentación, es el vehículo más prolífico y exitoso para la transferencia del conocimiento científico, conjuntamente con su transferencia oral por medio de conferencias y comunicaciones personales” [5].

Se tiene 3 tipos de indicadores los cuales pueden ser univariantes o multivariantes:

De producción. Son indicadores que permiten determinar la magnitud de la producción según el tiempo o según otros indicadores sociales.

De visibilidad e impacto. Son los indicadores más utilizados ya que describen la influencia del artículo o investigación dentro de la comunidad científica y de la sociedad.

De colaboración. Los indicadores de colaboración permiten observar las relaciones y el grado de colaboración entre actores investigativos sean estos países, instituciones o autores.

1.6 Bibliometría en Latinoamérica.

En Latinoamérica, los estudios bibliométricos han tenido un interés reciente (desde el inicio de 1990) y creciente.

Por ejemplo, Perú en el periodo 1996-2011 generó 21 artículos que emplearon Bibliometría en el campo de la Salud, de los cuales la mayoría fueron publicados desde el 2009 [3].

Brasil ha realizado estudios en periodos cortos y en temáticas especializadas empezando

en 1984 con un estudio enfocado en la ley de Bradford. Luego, en el 2003, Vanz realizó un estudio de los artículos sobre Bibliometría publicados. A partir de ese año, se han ido incrementado el número de estudios. Uno de los más recientes es “La Bibliometría, Informetría, Cienciometría y otras Metrías en el Brasil”, artículo publicado en 2016 donde analiza la literatura bibliométrica en el periodo 1973-2012 [16].

Para el año 2011 en Venezuela, la evaluación de la investigación científica aún se encontraba en un nivel inicial, frente a lo cual, el artículo “Algunas razones para evaluar la investigación científica venezolana desde la Bibliometría” fomentaba la inserción de la Bibliometría en la práctica [18].

Los datos anteriormente mencionados, evidencian que los estudios bibliométricos atraen cada vez más debido a la versatilidad de su uso en diferentes áreas de interés.

1.7 Bibliometría en el Ecuador.

En el Ecuador, la Bibliometría ha sido utilizada como herramienta de desarrollo de varios trabajos de tesis y artículos. A continuación se detalla algunos de dichos trabajos subdivididos en los enfoques de investigación bibliométrica anteriormente mencionados:

- *Investigación bibliométrica de disciplinas científicas.*- Algunos trabajos de este tipo son los siguientes:
 - “Indicadores bibliométricos de las tesis de pregrado en el repositorio digital de la Facultad de Ciencias Humanísticas y Sociales de la Universidad Técnica de Manabí, Ecuador, Período 2015-2016”, artículo recientemente desarrollado para la revista ReHuSO de la Universidad de Manabí en 2016 donde se desarrolla el manejo de repositorios usando Bibliometría [14].
 - “Análisis bibliométrico de la literatura científica publicada en la revista La Técnica (2010-2016)”, estudio del 2016, que muestra como la metodología bibliométrica sirve para analizar la literatura publicada en una revista institucional [17].
 - “Posicionamiento de los Investigadores del Ecuador en Webometrics Vol 1-2”, cuadernos del VIPS publicados en el presente año por la Escuela Politécnica Nacional con la finalidad de establecer el impacto de las políticas recientemente adoptadas en el campo de la Investigación científica [15].
 - “La producción científica en Ecuador en el contexto latinoamericano. Un análisis bibliométrico comparativo”, desarrollado por la SENESCYT con la finalidad de establecer el estatus de la producción científica en el período 2003-2012 [21].

- *Investigación bibliométrica para gestión y políticas.*- Algunos trabajos de este tipo constituyen trabajos de titulación relacionados a Ciencias Médicas, de manera específica constituyen estudios de niveles de producción en áreas predominantes tales como clínico-quirúrgicas y epidemiología (2011) [13].

Estos estudios han tomado bases de datos especializadas en literatura Biomédica como fuentes de información. A continuación algunos de ellos:

- “Análisis bibliométrico de la producción científica de influenza A-H1N1 en América del Sur desde Abril del 2009 a Abril del 2013: Identificando vacíos de conocimiento en Salud Pública sobre A-H1N1 en la región” [19].
- “Análisis bibliométrico de la producción científica en Ciencias de la Salud en Ecuador 1999-2009: Hacia una Política de Investigación y Desarrollo en Función de las Prioridades de Salud del País” [20].

Los trabajos [15] y [21] dan una idea muy precisa de cómo se encuentra el Ecuador a nivel de producción científica. En el 2010, por cada 100.000 habitantes, el número de papers publicados era de 2,41 papers, un valor considerablemente inferior al comparar con sus pares regionales: Chile (30,20 papers) y Colombia (6,14 papers).

Para el período 2003-2012, la producción científica casi se triplicó; hubo un incremento de artículos asociados a Ciencia y Tecnología siendo las principales áreas de investigación Química e Inmunología.

De hecho, el Ranking Iberoamericano SIR, el cual analiza la producción científica y visibilidad internacional, ubicaba en el 2012 a USFQ, PUCE y EPN en las posiciones 259, 305 y 392 respectivamente.

Sin embargo, lo anteriormente mencionado no era una muestra real de mejora, ya que muchas de las publicaciones eran desarrolladas fuera del país. La realidad era que Ecuador no tenía visibilidad internacional debido a su bajo nivel de producción y a su bajo nivel de publicación en grandes bases de datos de importancia mundial como Scopus, Medline, etc. En el presente año, al analizar el ranking internacional Webometrics, se tiene que el Ecuador ha mejorado en su posicionamiento, debido a que su producción científica ha perfeccionado su presencia, impacto y excelencia (indicadores usados por Webometrics) dentro de la comunidad científica; lo que muestra un correcto direccionamiento, aunque falta mucho para poder tener una producción competitiva a nivel internacional.

Capítulo 2

Metodología en Estudios Bibliométricos.

En el presente capítulo se menciona 3 enfoques metodológicos:

1. *Metodología General.*- Se presenta la metodología a seguir en cualquier caso de estudio bibliométrico.
2. *Metodología utilizada en el estudio.*- Se presenta la metodología aplicada en este trabajo. Cabe mencionar que esta metodología tiene como propósito estudiar al Sistema de Educación Ecuatoriano.
3. *Metodología con particularización a una IES.*- Se presenta la metodología a seguir en caso de que se realice un estudio bibliométrico dentro de una IES en particular.

2.1 Metodología General.

La Metodología General presentada a continuación abarca 3 principales aspectos: *Involucramiento, Depuración y Obtención de resultados.*

El aspecto de *Involucramiento* se enfoca en establecer un esquema que permita desarrollar el estudio, a pesar de que no se haya tenido contacto con la Bibliometría, como Ciencia, de manera previa.

El aspecto de *Depuración* se enfoca en la limpieza de datos tipo texto, ya que dentro de la Bibliometría, los datos como autor, afiliación, etc. del artículo o documento científico constituyen la base a utilizarse en un estudio bibliométrico.

Finalmente, el aspecto de *Obtención de resultados* se enfoca en utilizar Teoría Bibliométrica para generar resultados y establecer conclusiones.

A continuación se enuncia el detalle de la metodología:

- Antes de llevar a cabo un estudio de este tipo, se debe realizar una revisión introductoria de la Bibliometría.
- A continuación se establece el propósito del estudio.
- Con base en el propósito, se determina las unidades de estudio. El determinar las unidades de estudio permite establecer el nivel del estudio.

Existen 3 niveles de aplicación de estudios bibliométricos [3]:

- *Nivel Macro.*- Se considera como unidades a países, regiones.
- *Nivel Meso.*- Se considera como unidades a universidades, institutos o centros de investigación, revistas científicas.
- *Nivel Micro.*- Se considera como unidades a los investigadores.
- A continuación, se establece las características a estudiarse; es decir, se determina si se desea conocer producción, citación, proyecciones, etc. En este punto es importante la visión del estudio: situación actual y situación futura, ya que la visión de situación actual utiliza indicadores bibliométricos mientras que la visión de situación futura utiliza modelos bibliométricos.
- Con el nivel y las características determinadas, se consulta la Teoría Bibliométrica asociada, es decir, se detallada los indicadores y las leyes a utilizarse.
- Se procede a seleccionar la fuente de datos. Se considera como fuentes de datos a las bases con revistas indexadas: Scopus, Latindex, Web of Science, entre otras. Debido a que algunas fuentes son pagas, se podría acceder a través de una institución que tenga dichos accesos. También es importante constatar que los datos a utilizarse se encuentren en la fuente, ya que no existe una normalización en el contenido, los formatos y los archivos de descarga de las fuentes.
- Una vez seleccionada la fuente, se procede a la obtención de una o varias bases de datos iniciales de acuerdo a las necesidades del estudio. Cabe mencionar que los datos de las unidades se extraen de acuerdo a la fuente usada.
- Se procede a realizar la depuración de la(s) base(s) inicial(es). Debido al tipo de datos que se encuentran en la base, la depuración más relevante es la corrección de palabras con errores de escritura.
- Con la base o bases depuradas se procede a la aplicación de la Teoría Bibliométrica, es decir, se utiliza las definiciones de la Bibliometría para obtener resultados sobre

dichas bases.

- Con los resultados obtenidos se establece conclusiones, para lo cual, es importante recordar algunos puntos:
 - Los niveles de producción dependen del país donde se desarrolla la Investigación Científica.
 - Los niveles de citación dependen de la temporalidad de las publicaciones y de la rama de estudio.
 - La producción en las diversas áreas de estudio depende de la cantidad de productores en la misma, es decir, existen ramas con un número reducido de investigadores, razón por la cual comparar publicaciones entre ramas debe manejarse considerando las características propias de las ramas de estudio.
 - Los modelos de crecimiento se encuentran influenciados por el momento en el que se encuentren las unidades de estudio en los sentidos económicos y normativos.
 - La Ley de Lotka permite establecer un indicador de producción a nivel de investigador, sin embargo, se debe recordar que la producción de cada investigador se halla relacionada con la rama de desempeño, la existencia de pares de revisión y sobretodo la revista de publicación ya que existirán algunas revistas que no se encuentren indexadas en la fuente y que cuenten como producción del investigador.

2.2 Metodología utilizada en el estudio.

Para el desarrollo del presente trabajo se comenzó con una revisión bibliográfica enfocada en la Bibliometría y en los documentos escritos por los científicos pioneros en esta rama de la Ciencia. Debido a que el propósito del estudio es analizar la producción científica ecuatoriana, se optó por considerar a las universidades y escuelas politécnicas como unidades de estudio. Es importante mencionar que estas instituciones son generadoras de documentos científicos en el sentido de que sus investigadores plasman sus investigaciones y estudios en documentos compartidos con la comunidad científica.

Ya que se pretende realizar un análisis completo, las características a analizarse fueron producción, citas, colaboración y áreas de mayor producción y también se incursionó en modelos para establecer proyecciones. La Teoría Bibliométrica investigada se centró en la definición y cálculo de indicadores y en la aplicación del modelo de Crecimiento de Price.

Bajo este escenario, se consideró a Scopus como la mejor opción para la obtención de los datos, debido al número de revistas indexadas y al formato de descarga. El acceso a Scopus se realizó a través de credenciales de estudiante.

Para el proceso de limpieza, consolidación de la base, el cálculo de indicadores y aplicaciones de modelos se eligió el software estadístico R debido a su versatilidad. Para la consolidación de la base se programaron funciones de carga y desglose en R. También se programó la utilización de expresiones regulares para corregir la asociación Ciudad-País que provocaba errores al momento de establecer asociación entre IES, país y ciudad. Luego de las depuraciones se programó la consolidación de una base proveniente de todas las IES categorizadas por el CEAACES y que tuvieron datos en Scopus, pudiendo así consolidar datos de 33 IES.

Basándose en la teoría investigada, se programaron los indicadores bibliométricos y se los aplicó a la base consolidada. Para determinar las áreas con mayor difusión se utilizó la temática de las revistas en las que los documentos científicos fueron indexados. Debido a que el modelo de Price es el referenciado por la bibliografía para determinar proyecciones de crecimiento, se procedió a verificar dicho modelo. La primera ley formulada dentro de la Bibliometría (Ley de Lotka), se utilizó del propio software R y se verificó la Ley de Lotka en las IES con el objetivo de sustentar el cálculo del Índice de Productividad Personal.

Finalmente, para establecer conclusiones con base a los resultados, se realizó comparaciones entre las categorías del CEEACES.

A continuación se presenta el detalle metodológico utilizado para el caso de estudio del presente trabajo; se puede evidenciar que es una particularización de la metodología general anteriormente mencionada:

- La revisión bibliográfica realizada permitió consolidar los aspectos introductorios del Capítulo 1.
- El propósito del estudio es analizar la Producción Científica Ecuatoriana.
- Las unidades de estudio lo constituyen las Instituciones de Educación Superior (IES) y el nivel del estudio es Meso.
- Las características a estudiarse son producción, citación, colaboración nacional e internacional, áreas de producción, productividad personal y proyecciones.
- La Teoría Bibliométrica utilizada comprende los indicadores y leyes a considerarse (como mínimo) para desarrollar un estudio completo en el sentido de obtener una visión actual y obtener pronósticos. El detalle se encuentra en el Capítulo 3

- Se toma como fuente a Scopus debido a su acceso, contenidos y formatos de descarga.
- Los datos de las IES se obtienen de manera individual a través de archivos .bib descargables, consolidados con el uso de comandos y paquetes propios de R. También se obtiene una base auxiliar bajo el criterio de consulta AFFILCOUNTRY(ECUADOR). El detalle de la obtención de datos y consolidación de bases se encuentra en el Capítulo 4.
- En la depuración se utilizó expresiones regulares para corregir la asociación ciudad-país. También se determinó la temporalidad de los datos, se eliminó duplicados, se seleccionó las variables a utilizarse y se consolidó citas. El detalle se encuentra en el Capítulo 4.
- A través de la programación en R se aplica las definiciones del Capítulo 3 sobre las bases depuradas. Los indicadores utilizados son de producción, de visibilidad e impacto y de colaboración.
Mientras que se utilizó el Modelo de Crecimiento de Price para establecer proyecciones y la Ley Bibliométrica de Lotka para obtener el índice de productividad personal. Los resultados se encuentran en el Capítulo 5.
- Finalmente se establecen conclusiones con base en los resultados, los cuales se encuentran en el Capítulo 6.

2.3 Metodología con particularización a una IES.

Una particularización de interés es la aplicación de la metodología general a una determinada IES. Para este caso, los estudios solo se enfocan en el ambiente interno de la institución, por lo que los análisis van enfocados en conocer el desempeño de la IES. La metodología general mencionada es válida en este caso y se tendría lo siguiente:

1. Revisión Introductoria.
2. Definición del propósito del estudio.
3. Se tendría una sola unidad de estudio (IES) a un nivel Meso.
4. Definición de características a estudiarse.
5. Detalle de la Teoría Bibliométrica a utilizarse.
6. Elección de la fuente de datos.

7. Obtención de la base de datos inicial.
8. Depuración de la base inicial.
9. Aplicación de la Teoría Bibliométrica a la base depurada para la obtención de resultados.
10. Con base en los resultados, se establece conclusiones.

Es importante mencionar que el código programado en el presente documento permite desarrollar esta particularización, para lo cual se tendría que definir la IES, descargar de Scopus únicamente un archivo .bib y proceder a consolidarlo en R de la misma manera que se realizó para el conjunto de las 33 IES. Los indicadores, las áreas de mayor producción, el modelo de crecimiento de Price y el cálculo del índice de productividad personal se pueden calcular sin alterar el código. Cabe manifestar que la Teoría Bibliométrica del presente documento también abarca el fundamento teórico aplicado a una sola IES.

Capítulo 3

Teoría Bibliométrica.

Como ya se ha mencionado, la Bibliometría es un campo de investigación en proceso de desarrollo.

Los primeros modelos desarrollados en este campo fueron, en su momento, los adecuados para establecer aspectos observables básicos de producción y citación, siendo las conclusiones obtenidas de dichos modelos las ya mencionadas Leyes Bibliométricas (Ver 1.3), muchas de las cuales aún se consideran en los estudios, con el conocimiento de su validez limitada, causada por la no consideración de factores altamente influyentes en la producción científica, como por ejemplo la coyuntura social de la comunidad investigadora al momento de su definición [4].

Con el objetivo de tener más coherencia con la realidad investigativa, surgen los Modelos de Crecimiento de la Literatura (Ver 1.4). Dichos modelos se utilizan de acuerdo a la madurez que presenta la Ciencia, al campo de investigación y a los objetivos propios de cada estudio bibliométrico.

Finalmente, los Indicadores Bibliométricos (Ver 1.5), constituyen el aspecto más conocido al momento de desarrollar Bibliometría, ya que permiten representar aspectos propios de la producción científica de una manera absoluta y/o relativa y así obtener una visión comportamental general y/o específica de la producción.

En este capítulo se detalla las leyes, modelos e indicadores a considerarse a lo largo del presente trabajo; dicho detalle, permitirá sustentar los análisis y conclusiones resultantes.

Cabe mencionar que el presente estudio tiene **una aplicación a nivel Meso, es decir, las unidades de estudio lo constituyen las Instituciones de Educación Superior.**

3.1 Indicadores Bibliométricos.

A continuación se presenta de manera general los indicadores a utilizarse en el presente estudio; estos indicadores permiten desarrollar estudios bibliométricos y analizar producción, impacto y colaboración de la investigación científica.

Es importante mencionar que los indicadores se adaptarán a las necesidades del estudio, por ejemplo, para el análisis puede necesitarse el indicador general desagregado a nivel de año, a nivel de IES o según la categorización (Ver 4.1, 4.2, 4.3).

3.1.1 Indicadores de Producción.

3.1.1.1 Número total de Publicaciones (PC).

El *Número total de Publicaciones* es una medida básica del volumen de salida, sensible a las características propias de la comunidad investigadora [22].

De manera general, para el presente estudio, se considera todos los documentos que tienen al menos una IES ecuatoriana como afiliación:

PC = Número de documentos publicados en Scopus que tienen al menos una IES como afiliación.

A continuación se presenta la definición del presente indicador según los niveles de desagregación.

Sean:

- $j = 1, \dots, m$; donde m es el número de años del período de análisis,
- $i = 1, \dots, n$; donde n es la cantidad de IES con al menos un documento en Scopus dentro del período de análisis,
- $k = 1, \dots, 3$; donde k es la cantidad de categorías definidas por el CEAACES.

Considere d_{ijk} el número de documentos publicados en Scopus por la IES i , en el año j , en la categoría k .

Entonces:

- **Nivel: Año.-**

$$\forall j, \quad PC_{(j)} = \sum_{i=1}^n \sum_{k=1}^3 d_{ijk}. \quad (3.1)$$

- **Nivel: Categoría.-**

$$\forall k, \quad PC_{(k)} = \sum_{i=1}^n \sum_{j=1}^m d_{ijk}. \quad (3.2)$$

- **Nivel: IES.-**

$$\forall i, \quad PC_{(i)} = \sum_{j=1}^m \sum_{k=1}^3 d_{ijk}. \quad (3.3)$$

3.1.2 Indicadores de Visibilidad e Impacto.

3.1.2.1 Número Total de Citas (CT).

El *Número Total de Citas* es una medida de impacto académico sensible al tópico y tipo de documento [22], que “permite analizar la repercusión de los trabajos científicos en la comunidad científica” [27].

Para el presente estudio se considera todos las citas registradas en Scopus que tienen los documentos con al menos una IES ecuatoriana como afiliación:

CT = Número de citas pertenecientes a todos los documentos producidos por las IES.

A continuación se presenta la definición del presente indicador según los niveles de desagregación:

Sean:

- $j = 1, \dots, m$; donde m es el número de años del período de análisis,
- $i = 1, \dots, n$; donde n es la cantidad de IES con al menos un documento en Scopus dentro del período de análisis,
- $k = 1, \dots, 3$; donde k es la cantidad de categorías definidas por el CEAACES.

Considere c_{ijk} el número de citas asociadas a todos los documentos en Scopus producidos por la IES i , en el año j , en la categoría k .

Entonces:

- **Nivel: Año.-**

$$\forall j, \quad CT_{(j)} = \sum_{i=1}^n \sum_{k=1}^3 c_{ijk}. \quad (3.4)$$

- **Nivel: IES.-**

$$\forall i, \quad CT_{(i)} = \sum_{j=1}^m \sum_{k=1}^3 c_{ijk}. \quad (3.5)$$

3.1.2.2 Promedio de citas por documento científico (CD).

El *Promedio de citas por documento científico* es una medida de impacto académico sensible al tópico y tipo de documento [22] calculado como el “promedio de las citas recibidas por el total de la producción” [27].

De manera general, el cálculo considera los indicadores PC (Ver 3.1.1.1) y CT (Ver 3.1.2.1) previamente mencionados:

$$CD = \frac{CT}{PC}.$$

A continuación se presenta la definición del presente indicador según los niveles de desagregación:

Sea:

- $j = 1, \dots, m$; donde m es el número de años del período de análisis,
- $i = 1, \dots, n$; donde n es la cantidad de IES con al menos un documento en Scopus dentro del período de análisis,
- $k = 1, \dots, 3$; donde k es la cantidad de categorías definidas por el CEAACES.

Considere:

- d_{ijk} el número de documentos publicados en Scopus por la IES i , en el año j , en la categoría k ;
- c_{ijk} el número de citas asociadas a todos los documentos en Scopus producidos por la IES i , en el año j , en la categoría k .

Entonces:

- **Nivel: Año.-**

$$\forall j, \quad CD_{(j)} = \frac{\sum_{i=1}^n \sum_{k=1}^3 c_{ijk}}{\sum_{i=1}^n \sum_{k=1}^3 d_{ijk}} \quad (3.6)$$

3.1.2.3 Índice-H (H).

El Índice-H o H-Index es un indicador que permite evaluar de manera simultánea la productividad y el impacto, ya que permite “equiparar las publicaciones de un autor y las citas que estas publicaciones han recibido” [39].

El Índice-H es una función de impacto y a continuación se presenta su formulación [28]:

Función de Impacto:

Sea \mathbb{N}_0 el conjunto de todos los enteros no negativos; $\mathbb{N}_0^{1,2,\dots} = \bigcup_{n=1}^{\infty} \mathbb{N}_0^n$ el conjunto de todas las secuencias de longitud arbitraria de elementos en \mathbb{N}_0 .

Dentro del contexto de la Bibliometría, $\mathbf{x} \in \mathbb{N}_0^{1,2,\dots}$ representa el registro de una citación.

Una función de impacto en $\mathbb{N}_0^{1,2,\dots}$ es un operador de agregación $J : \mathbb{N}_0^{1,2,\dots} \rightarrow \mathbb{R}$ que cumple que:

1. Es no decreciente en cada variable:

$$(\forall n)(\forall \mathbf{x}, \mathbf{y} \in \mathbb{N}_0^n), \mathbf{x} \leq \mathbf{y} \Rightarrow J(\mathbf{x}) \leq J(\mathbf{y}) \quad (3.7)$$

2. Es arity-monótona, es :

$$(\forall n, m)(\forall \mathbf{x} \in \mathbb{N}_0^n)(\forall \mathbf{z} \in \mathbb{N}_0^m), J(\mathbf{x}) \leq J(\mathbf{x}, \mathbf{z}) \quad (3.8)$$

3. Es simétrica:

$$(\forall n)(\forall \mathbf{x}, \mathbf{y} \in \mathbb{N}_0^n), \mathbf{x} \cong \mathbf{y} \Rightarrow J(\mathbf{x}) = J(\mathbf{y}) \quad (3.9)$$

4. Satisface la condición de límite inferior débil:

$$\inf_{\mathbf{x} \in \mathbb{N}_0^{1,2,\dots}} J(\mathbf{x}) = 0 \quad (3.10)$$

5. Satisface la condición de límite superior débil:

$$\sup_{\mathbf{x} \in \mathbb{N}_0^{1,2,\dots}} J(\mathbf{x}) = \infty \quad (3.11)$$

donde para $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$ y $\mathbf{z} = (z_1, \dots, z_m)$, se tiene que:

- $\mathbf{x} \leq \mathbf{y}$, ssi $x_1 \leq y_1, x_2 \leq y_2, \dots, x_n \leq y_n$.
- $(\mathbf{x}, \mathbf{z}) = (x_1, \dots, x_n, z_1, \dots, z_m) \in \mathbb{N}_0^{n+m}$.
- $\mathbf{x} \cong \mathbf{y}$ ssi existe una permutación σ tal que $\mathbf{x} = (y_{\sigma(1)}, \dots, y_{\sigma(n)})$.

Definición del Índice-H.

El H-index o Índice-H es una función de impacto tal que:

$$\forall \mathbf{x} \in \mathbb{N}_0^n, \quad Hindex(\mathbf{x}) = \max\{i = 1, \dots, n : x_{(n+1-i)} \geq i\} \quad (3.12)$$

donde $x_{(i)}$ corresponde al i-ésimo estadístico de orden, es decir, es el i-ésimo valor más pequeño de \mathbf{x} .

En el presente trabajo, en el cálculo del Índice-H se realiza considerando a cada IES como un autor.

3.1.3 Indicadores de Colaboración.

3.1.3.1 Número total de publicaciones con colaboración internacional (DCI).

El *Número total de publicaciones con colaboración internacional* es una medida de la producción generada en colaboración [29].

De manera general, para el presente estudio, se considera todos los documentos que tienen al menos una IES extranjera o internacional como afiliación:

$$DCI = \text{Número de documentos publicados en Scopus que tienen al menos una IES extranjera como afiliación.}$$

A continuación se presenta la definición del presente indicador según el nivel de desagregación:

Sean:

- $j = 1, \dots, m$; donde m es el número de años del período de análisis,
- $i = 1, \dots, n$; donde n es la cantidad de IES con al menos un documento en Scopus dentro del período de análisis,
- $k = 1, \dots, 3$; donde k es la cantidad de categorías definidas por el CEAACES.

Considere dc_{ijk} el número de documentos publicados en Scopus por la IES i , en el año j , en la categoría k , con al menos una afiliación extranjera.

Entonces:

- **Nivel: Año.-**

$$\forall j, \quad DCI_{(j)} = \sum_{i=1}^n \sum_{k=1}^3 dc_{ijk} \quad (3.13)$$

3.1.3.2 Colaboración Internacional.

El indicador de *Colaboración Internacional* constituye un porcentaje que refleja el nivel de participación internacional en la producción de documentos científicos.

De manera general, el cálculo considera los indicadores PC (Ver 3.1.1.1) y DCI (Ver 3.1.3.1) anteriormente mencionados [29].

$$\%CI = \frac{DCI * 100}{PC}.$$

A continuación se presenta la definición del presente indicador según el nivel de desagregación:

Sean:

- $j = 1, \dots, m$; donde m es el número de años del período de análisis,
- $i = 1, \dots, n$; donde n es la cantidad de IES con al menos un documento en Scopus dentro del período de análisis,
- $k = 1, \dots, 3$; donde k es la cantidad de categorías definidas por el CEAACES.

Considere:

- d_{ijk} el número de documentos publicados en Scopus por la IES i , en el año j , en la categoría k ;
- dc_{ijk} el número de documentos publicados en Scopus por la IES i , en el año j , en la categoría k , con al menos una afiliación extranjera.

Entonces:

- **Nivel: Año.-**

$$\forall j, \%CI_{(j)} = \frac{(\sum_{i=1}^n \sum_{k=1}^3 dc_{ijk}) * 100}{\sum_{i=1}^n \sum_{k=1}^3 d_{ijk}} \quad (3.14)$$

3.1.3.3 Indicadores en el presente estudio.

En este trabajo, se utilizarán todos los indicadores mencionados para lo siguiente:

1. Analizar el estado de la producción científica ecuatoriana en el período de análisis 2010-2016.
2. Utilizar el Índice-H para determinar el impacto de cada IES dentro de la categoría de pertenencia.

Lo mencionado, se puede visualizar en el capítulo 5 de manera detallada.

3.2 Leyes Bibliométricas.

3.2.1 Ley de Lotka.

3.2.1.1 Formulación Original (Clásica).

El 19 de Junio de 1926, Alfred J. Lotka publicó en el Journal of the Washington Academy of Science su artículo denominado “The frequency distribution of scientific productivity”, en el que, según sus propias palabras, Lotka se encontraba “interesado en determinar (si era posible) la contribución de hombres de diferente calibre, al progreso de la Ciencia” [24].

Para su estudio, Lotka utilizó 2 muestras; la primera formada por los autores con nombre de letras A y B publicados en el Decennial Index of Chemical Abstracts 1907-1916 y la segunda formada por todos los autores indexados hasta 1900 en el Auerbach's Geschichtstafeln der Physik [24].

Como resultado, se obtuvo la siguiente fórmula general (llamada "Ley" de manera no apropiada debido a que Lotka no la menciona como tal en su documento) para la relación que existe entre la frecuencia absoluta y de personas que hacen x contribuciones:

$$x^n y = c \Leftrightarrow y = \frac{c}{x^n}; \quad c : cte \quad (3.15)$$

a partir de la cual, y con un proceso de estimación de los parámetros n y c , Lotka enuncia lo siguiente:

El número de autores produciendo n contribuciones es el $\frac{1}{n^2}$ de aquellos produciendo 1, y la proporción de aquellos que hacen una sola contribución entre todos los autores es aproximadamente del 60 por ciento [24].

De otra manera:

Sea:

- A_n el número de autores desarrollando (o aportando) n artículos y A_1 el número de autores desarrollando (o aportando) 1 artículo,

Entonces:

$$A_n = \frac{A_1}{n^2}. \quad (3.16)$$

3.2.1.2 Formulación Generalizada.

En 1985, Miranda Lee Pao, realiza "una de las aportaciones metodológicas más significativa y esclarecedora para la comprensión del Modelo Matemático de Lotka" en su artículo "Lotka's Law: a testing procedure"; en dicho documento, Lee Pao analiza los inconvenientes encontrados en diversos estudios por el uso exacto de la potencia cuadrática y señala que la relación inversa de Lotka debe mantenerse con la variante de que la potencia se calcule a partir del propio conjunto de datos [25].

Se propone la siguiente generalización [26]:

Sea:

- A_n el número de autores desarrollando (o aportando) n artículos y A_1 el número de autores desarrollando (o aportando) 1 artículo,

Entonces:

$$A_n = \frac{A_1}{n^m}. \quad (3.17)$$

con $m \in \mathbb{R}^+$.

3.2.1.3 Usos de la Ley de Lotka en los estudios bibliométricos.

La Ley de Lotka Clásica, de manera simplificada, postula que “gran proporción del progreso de la Ciencia se debe a pocos científicos muy productivos” generando una **productividad sesgada** [34].

Es así como, el uso más importante de la Ley de Lotka Clásica (Ver 3.15), es el **Índice de Productividad Personal**, ya que, de ser el caso que se cumpla la ley, este indicador permite distribuir a los autores en 3 niveles de productividad.

Se define el índice de la siguiente manera [35]:

Sea:

- N el número de artículos generados por el autor,

Entonces:

$$IP = \log(N) \quad (3.18)$$

con $N \in \mathbb{Z}^+$.

A continuación, se muestra la clasificación de los autores según este índice [35]:

N	IP	Tipo de Productor
≥ 10	$IP \geq 1$	Grande
2 – 9	$0 < IP < 1$	Intermedio
1	$IP = 0$	Transitorio

3.2.1.4 Lotka en el presente estudio.

En este trabajo, se utilizará la Ley de Lotka para lo siguiente:

1. Verificar si la producción científica dentro de las IES ecuatorianas satisfacen la Ley de Lotka Clásica.
2. Ejemplarizar la clasificación de los autores según el Índice de Productividad Personal.

Lo mencionado, se puede visualizar en el capítulo 5 de manera detallada.

3.3 Modelos de Crecimiento.

3.3.1 Modelo de Crecimiento Exponencial.

En 1956, D. J. de Solla Price, postuló que el crecimiento de la información científica presenta un comportamiento similar a fenómenos naturales y no a otros fenómenos sociales y postula que el crecimiento de la Ciencia tiene un comportamiento exponencial [8].

La justificación de la misma, se basa en múltiples observaciones referidas al crecimiento de autores, al crecimiento de revistas científicas, etc. [37], pero sobre todo, al hecho de que “la Ciencia es producto de la actividad humana, cuya población crece exponencialmente” [38]. Dentro de las formas de enunciar la “Ley del crecimiento exponencial de Price”, se tiene las siguientes:

- “La Ciencia crece a interés compuesto, multiplicándose por una cantidad determinada en períodos iguales de tiempo” [37].
- “La tasa de crecimiento es proporcional al tamaño de la población o magnitud total adquirida” [37].
- “Cuanto más grande es la Ciencia, más deprisa crece” [37].

A continuación se describe el modelo de crecimiento exponencial de Price, el cual es la consolidación de los anteriores enunciados [36]:

Si $P(t)$ indica la producción científica en un instante de tiempo t , entonces, $P(t + \Delta t)$ sería la producción científica en el instante $t + \Delta t$.

También, se tiene que el crecimiento neto del número de artículos en un intervalo de tiempo $[t, t + \Delta t]$, es igual a $P(t + \Delta t) - P(t)$.

Se desea calcular el número de publicaciones generadas por unidad de tiempo en un intervalo $[t, t + \Delta t]$.

La respuesta parte de la siguiente expresión:

$$\frac{P(t + \Delta t) - P(t)}{\Delta t} \quad (3.19)$$

Si se utiliza el concepto de derivada (el cual explica el desarrollo dinámico de la mayoría de los fenómenos naturales) sobre la expresión 3.19, se tiene que, cuando $\Delta t \rightarrow 0$:

$$\lim_{\Delta t \rightarrow 0} \left(\frac{P(t + \Delta t) - P(t)}{\Delta t} \right) = P'(t). \quad (3.20)$$

Es así como $P'(t)$, es el número de publicaciones generadas por unidad de tiempo.

Ahora, considerando un intervalo de tiempo sumamente pequeño, se tiene que:

$$P(t + \Delta t) - P(t) \cong P'(t) \cdot \Delta t \quad (3.21)$$

siendo $P'(t) \cdot \Delta t$, el número de artículos generados en el intervalo.

Por otro lado, la tasa de crecimiento media relativa, será la variación neta del número de artículos, por unidad de tiempo (por ejemplo un año), dividida por el número de artículos que se tenía al comienzo de ese año; se tiene la siguiente expresión:

$$\frac{\frac{P(t+\Delta t)-P(t)}{\Delta t}}{P(t)} = \frac{P(t + \Delta t) - P(t)}{P(t) \cdot \Delta t} \quad (3.22)$$

y al calcular el límite sobre la expresión 3.22 cuando $\Delta t \rightarrow 0$, se obtiene la siguiente expresión para la tasa de crecimiento instantánea:

$$\frac{1}{P(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t) - P(t)}{\Delta t} = \frac{P'(t)}{P(t)}. \quad (3.23)$$

Considerando una tasa de crecimiento constante, se tiene la siguiente ecuación diferencial ordinaria:

$$\frac{P'(t)}{P(t)} = a \Rightarrow P'(t) = a \cdot P(t) \quad (3.24)$$

Al resolver la ecuación diferencial 3.24, se tiene que:

$$\ln(P(t)) = at + c ; c \text{ cte.}$$

$$P(t) = C \cdot e^{at} ; C \text{ cte}$$

y debido a que en:

$$t = 0, P(0) = C = P_0$$

se tiene la siguiente solución:

$$P(t) = P_0 \cdot e^{at}$$

En conclusión, se tiene que la curva de crecimiento de la producción científica, se representa por:

$$P(t) = P_0 \cdot e^{at} \quad (3.25)$$

con $a \in \mathbb{R}^+$ y $P(0) = P_0$ como condición inicial.

3.3.1.1 Modelos de crecimiento en el presente estudio.

En este trabajo, se utilizará los modelos de crecimiento para lo siguiente:

1. Verificar si la producción científica de las IES ecuatorianas satisfacen el modelo de crecimiento de Price.
2. Establecer proyecciones al año 2020 utilizando el modelo ajustado.

Lo mencionado, se puede visualizar en el capítulo 5 de manera detallada.

Capítulo 4

Extracción y Depuración de Datos.

Los datos que se utilizan en el presente trabajo provienen de Scopus-Elsevier, la cual, es la mayor base indexada de referencias bibliográficas de literatura científica mundial revisada por pares, con más de 21.500 títulos de más de 5.000 editoriales internacionales.

“Scopus permite una visión multidisciplinaria de la ciencia e integra todas las fuentes relevantes para la investigación básica, aplicada e innovación tecnológica a través de patentes, fuentes de Internet de contenido científico, revistas científicas de acceso abierto, memorias de congresos y conferencias. Es actualizado diariamente y contiene los Articles in Press de más de 5.000 revistas” [10].

Para desarrollar el estudio, se utilizará el software libre R, un conjunto integrado de librerías y un lenguaje de programación que nos permite realizar cálculos fundamentalmente estadísticos (R Core Team, 2012).

4.1 Universo de Estudio.

El universo de estudio está conformado por todas las Universidades y Escuelas Politécnicas de pregrado y posgrado que se encuentran registradas en el CEAACES independientemente de su categoría (55 IES en total), a julio del 2017 y la Universidad de Investigación de Tecnología Experimental Yachay (Yachay Tech), la cual se considera como IES sin categorización.

A continuación el detalle de las IES categorizadas [11]:

Categoría A	Escuela Politécnica Nacional
	Escuela Superior Politécnica del Litoral
	Universidad San Francisco de Quito
	Universidad de Cuenca
	Universidad de las Fuerzas Armadas (ESPE)
	Universidad de Especialidades Espíritu Santo
	Universidades de Postgrado
	Facultad Latinoamericana de Ciencias Sociales
	Universidad Andina Simón Bolívar

Cuadro 4.1: Instituciones Acreditadas A.

Categoría B	Escuela Superior Politécnica de Chimborazo
	Pontificia Universidad Católica del Ecuador
	Universidad Casa Grande
	Universidad Católica de Santiago de Guayaquil
	Universidad Central del Ecuador
	Universidad del Azuay
	Universidad Estatal de Milagro
	Universidad Nacional de Loja
	Universidad Particular Internacional SEK
	Universidad Politécnica Salesiana
	Universidad Técnica de Ambato
	Universidad Técnica del Norte
	Universidad Técnica Estatal de Quevedo
	Universidad Técnica Particular de Loja
	Universidad Tecnológica Empresarial de Guayaquil
	Universidad Tecnológica Equinoccial
	Universidad Tecnológica Indoamérica
	Universidad de los Hemisferios
	Universidad Estatal Amazónica
	Universidad Politécnica del Carchi
	Universidad Iberoamericana
	Universidad Técnica de Manabí
	Universidad de las Américas
	Universidad Internacional del Ecuador
	Universidad de Guayaquil
	Universidad Técnica de Machala
	Universidad Católica de Cuenca
	Universidades de Postgrado
Instituto de Altos Estudios Nacionales	

Cuadro 4.2: Instituciones Acreditadas B (Actualizado 05-06-2017).

Categoría C	Escuela Superior Politécnica Agropecuaria de Manabí
	Universidad de Especialidades Turísticas
	Universidad del Pacífico Escuela de Negocios
	Universidad Estatal de Bolívar
	Universidad Laica Vicente Rocafuerte de Guayaquil
	Universidad Metropolitana
	Universidad Nacional del Chimborazo
	Universidad Regional Autónoma de los Andes
	Universidad Técnica de Babahoyo
	Universidad Tecnológica Israel
	Universidad Estatal Península de Santa Elena
	Universidad Particular San Gregorio de Portoviejo
	Universidad Tecnológica ECOTEC
	Universidad Técnica de Cotopaxi
	Universidad Estatal del Sur de Manabí
	Universidad de Otavalo
	Universidad Agraria del Ecuador
Universidad Laica Eloy Alfaro de Manabí	
Universidad Técnica Luis Vargas Torres de Esmeraldas	

Cuadro 4.3: Instituciones Acreditadas C (Actualizado 05-06-2017).

4.2 Consideración para el procedimiento de extracción y depuración.

Es importante mencionar que las secciones 4.3 y 4.5 fueron desarrolladas de acuerdo al estado de la plataforma de Scopus a la fecha de corte: 31 de julio del 2017; es decir, en un futuro a partir de la fecha de corte, se puede tener actualizaciones de la plataforma que implementen mejoras o corrijan fallas.

De todas maneras, los cambios serán de forma y no de fondo, siendo posible adaptar los criterios de las secciones 4.3 y 4.5 a posibles futuras actualizaciones.

Cabe mencionar que en el caso de la sección 4.5, pueden surgir otras necesidades de depuración no observadas en el presente estudio, causadas por posibles actualizaciones, siendo así, las mencionadas en el presente documento, criterios mínimos a verificarse en futuros estudios.

4.3 Procedimiento de Extracción.

Nota: Se considera como documentos científicos a todos los posibles tipos de representación científica: artículos, libros, conferencias, etc.

Antes de proceder a extraer los datos de la base de Scopus es importante mencionar lo

siguiente:

- Algunas de las IES anteriormente indicadas no contarán con información dentro de las bases internas de Scopus (Número de documentos = 0), lo cual no implica que la IES no desarrolle actividades investigativas; las razones de la ausencia pueden ser variadas:
 - La IES no publica en las revistas que considera Scopus para su repositorio.
 - Los documentos de la IES pueden estar en proceso de revisión.
 - Los documentos de la IES pueden no cumplir con los criterios determinados por Scopus para su ingreso en la base de datos.

A continuación muestra un ejemplo que detalla el proceso de extracción :

1. **Ingreso a la Base de Datos.**- Para poder acceder a la página de Scopus <https://www.scopus.com/home.uri> es necesario contar con un acceso privado o uno dado usualmente por un campus universitario.

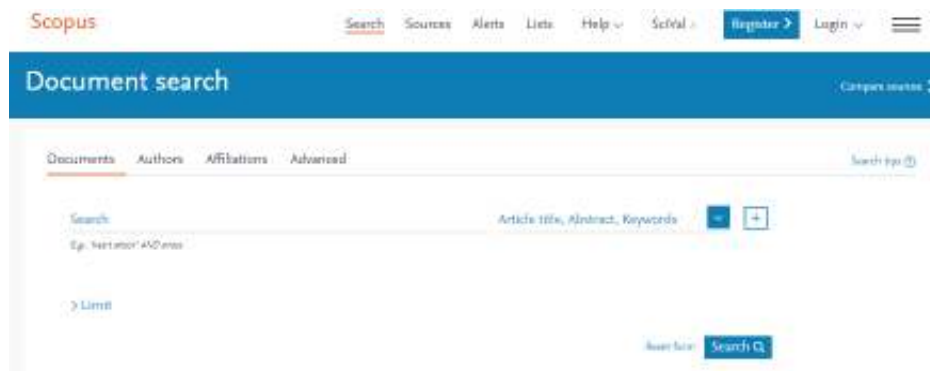


Figura 4.1: Pantalla Inicial Scopus.

2. **Búsqueda por IES.**- Ingresando el nombre de la IES en la opción *Search* de la pestaña *Affiliations*, se puede acceder a la información a recolectarse. La búsqueda puede refinarse usando la opción *Advanced* (a través del uso de operadores y códigos propios de Scopus) o usando *Search for documents by affiliation*.

Figura 4.2: Ingreso Nombre IES.

Affiliation name	Documents	City	Country/Territory
Escuela Politécnica Nacional Escuela Politécnica Nacional Escuela Politécnica Nacional	754	Quito	Ecuador

Figura 4.3: Resultado búsqueda IES.

3. **Descarga de Datos.**- La búsqueda arrojará los documentos que posean como afiliación a la IES considerada. Es importante recordar que muchos de los artículos son realizados en cooperación con otras IES (nacionales y/o extranjeras). La selección de los documentos puede ser refinada por años.

Document title	Authors	Year	Source	Cited by
1. Synbiotic fermentation... with very low ethanol tolerance for coupled domain wall motion	Lopodatu, S., Samborsk, H., Dabchin, R., G., Tatra, G., Marova, C.H.	2017	Scientific Reports, 7(111648)	3
2. Control of stability... caused by Alternaria alternata in yellow platynine. Sciencewise implementation through hot-water dips	Vilapiana, R., Ros, D., Valero-Chambers, S.	2017	LWT - Food Science and Technology, 82, pp. 162-169	8

Figura 4.4: Documentos IES.

Utilizando la opción *Export*, y de acuerdo con la finalidad, se puede seleccionar los datos a ser descargados. Se tienen 6 tipos de formatos de salida.

Para los propósitos del presente trabajo, los datos se descargarán en formato .bib, y se considerará únicamente los datos asociados a *Citation information* y *Bibliographical information*.

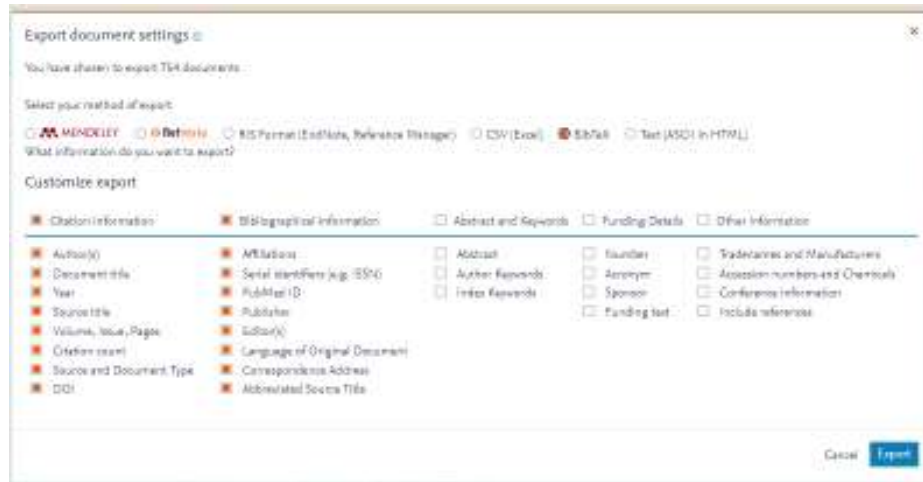


Figura 4.5: Datos a exportar.

Al final, se tendrá un archivo tipo BibTex por cada IES que posea información en Scopus. En total se obtuvieron 33 archivos con fecha de corte 31-07-2017.

En la tabla 4.4 se muestra el detalle de las IES que presentaron información y que conformarán la base cruda a utilizarse (Ver 38):

Nro.	IES	SIGLA	Nro. Documentos
1	Escuela Politécnica Nacional	EPN	796
2	Escuela Superior Politécnica del Litoral	ESPOL	747
3	Universidad San Francisco de Quito	USFQ	1334
4	Universidad de Cuenca	UCUENCA	540
5	Universidad de las Fuerzas Armadas	ESPE	669
6	Universidad de Especialidades Espíritu Santo	UEES	184
7	Facultad Latinoamericana de Ciencias Sociales	FLACSO	1
8	Universidad Andina Simón Bolívar	UASB	3
9	Escuela Superior Politécnica de Chimborazo	ESPOCH	182
10	Pontificia Universidad Católica del Ecuador	PUCE	786
11	Universidad Católica de Santiago de Guayaquil	UCSG	205
12	Universidad Central del Ecuador	UCE	482
13	Universidad del Azuay	UDA	78
14	Universidad Nacional de Loja	UNL	131
15	Universidad Particular Internacional SEK	SEK	34
16	Universidad Politécnica Salesiana	UPS	380
17	Universidad Técnica de Ambato	UTA	130
18	Universidad Técnica del Norte	UTN	64
19	Universidad Técnica Particular de Loja	UTPL	714
20	Universidad Tecnológica Equinoccial	UTE	82
21	Universidad Tecnológica Indoamérica	UTI	1
22	Universidad Técnica de Manabí	UTM	44
23	Universidad de las Américas	UDLA	198
24	Universidad Internacional del Ecuador	UIDE	1
25	Universidad de Guayaquil	UG	214
26	Universidad Técnica de Machala	UTMACH	75
27	Instituto de Altos Estudios Nacionales	IAEN	1
28	Universidad Nacional del Chimborazo	UNACH	112
29	Universidad Regional Autónoma de los Andes	UNIANDES	1
30	Universidad Estatal Península de Santa Elena	UPSE	44
31	Universidad Agraria del Ecuador	UAE	25
32	Universidad Laica Eloy Alfaro de Manabí	ULEAM	83
33	Universidad de Investigación de Tecnología Experimental Yachay	YACHAY	156

Cuadro 4.4: IES con datos en Scopus.

Utilizando R, estos archivos se consolidarán en un solo objeto (base), el mismo que será depurado y luego utilizado en el desarrollo del presente trabajo.

4.4 Consolidación de la base inicial.

El paquete Bibliometrix permite consolidar los archivos .bib descargados en un solo objeto del tipo dataframe.

Esto se realiza con las siguientes funciones:

- *scopus2df*.- Función que permite leer y extraer el formato de escritura de los datos (caracteres especiales) de manera directa de los archivos .bib generados por Scopus.
- *mergeDbSources*.- Función que permite construir un dataframe a partir de archivos .bib descargados.

En la tabla 4.5, se muestra los campos que se pueden extraer de los archivos descargados:

Campo	Descripción
AU	Autor(es)
TI	Título del Documento
SO	Revista de Publicación
JI	Revista de Publicación (ISO Abreviación)
DT	Tipo de Documento
DE	Palabras Clave del Autor
ID	Palabras Clave asociadas por Scopus
AB	Abstract
C1	Dirección del Autor
RP	Dirección de Reimpresión
CR	Referencias Citadas
TC	Número de Citaciones
PY	Año de Publicación
UT	Identificador del Primer Autor
DB	Base de Datos Bibliográfica

Cuadro 4.5: Variables de Scopus.

Para tener claridad respecto a la IES de procedencia del registro, se genera un campo con la sigla de la IES. Al final se tiene un dataframe denominado B_IES_C1 que consta de 8.497 registros y 16 campos.

Nota.- Para los casos en que las IES poseen un solo documento, la función *mergeDbSources* puede generar errores y no proceder a cargar. Como solución, se edita el archivo .bib duplicando el mencionado registro y luego de que se tiene la base consolidada, usando la función *distinct* del paquete dplyr (paquete que permite un manejo práctico de dataframes) se procede a retirar los registros auxiliares generados.

4.5 Procedimiento de Depuración.

La limpieza de datos constituye uno de los pasos más importantes para reducir el riesgo que se genera al tomar decisiones sustentadas en la información proveniente de una fuente externa; este proceso asegura que los datos a utilizarse en los análisis tengan un nivel de calidad que permita sustentar los futuros resultados.

Debido a la estructura de los archivos .bib, los aspectos de depuración a considerarse son los siguientes:

- Selección de datos que se van a considerar en el análisis.
- Corrección de datos cargados en Scopus de manera incorrecta.
- Eliminación de datos duplicados (generados por mecanismo de consulta).
- Consolidación de citaciones.

4.5.1 Selección de datos que se van a considerar en el análisis.

Se considera que todos los artículos registrados en Scopus son de alta calidad metodológica debido a los estándares internos de Scopus asociados a las revistas fuentes; por lo cual, en este paso, únicamente se limita los datos y los campos a utilizarse en el estudio.

Los datos a conservarse de la base inicial para el presente estudio **corresponden a la temporalidad 2010-2017, siendo el período de análisis 2010-2016 y los datos asociados al 2017 datos informativos.**

Al observar la base inicial, se tiene artículos con fechas anteriores y posteriores a la temporalidad; de hecho, existen documentos con fecha 2018, los cuales se consideran como valores atípicos ya que su publicación en alguna revista científica, tiene una fecha tentativa.

Se toman en cuenta todos los posibles tipos de documentos: Artículo, Paper de Conferencia, Artículo in Press, Libro, Revisión, Nota, Editorial, Carta y Errata.

Finalmente, en la tabla 4.6, se muestra los campos o variables a conservarse de la tabla 4.5 y a considerarse en el estudio:

Campo	Descripción
AU	Autor(es)
TI	Título del Documento
SO	Revista de Publicación
DT	Tipo de Documento
C1	Dirección del Autor
TC	Número de Citaciones
PY	Año de Publicación
UT	Identificador del Primer Autor
IES	Campo generado para identificación. (Ver 4.4)

Cuadro 4.6: Variables a considerarse en el estudio.

Este conjunto de variables constituyen los campos necesarios para el desarrollo de los indicadores del estudio.

Para estas acciones se utilizará las funciones *filter* y *select*, pertenecientes al paquete dplyr. Al final se tiene un dataframe corregido denominado B_IES_C2 que consta de 6.836 registros y 9 campos.

4.5.2 Corrección de datos cargados en Scopus con errores.

Existen registros de la base que poseen el país de afiliación incorrecto dentro del campo C1 (Ver 4.6):

[1] "ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, LAUSANNE, SWITZERLAND; ESCUELA POLITÉCNICA NACIONAL, P.O. BOX 2759, QUITO, SWITZERLAND"
[2] "DEPARTMENT OF MECHANICAL ENGINEERING, ESCUELA POLITÉCNICA NACIONAL, SPAIN; DEPARTMENT OF MECHANICAL ENGINEERING, PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO, RIO DE JANEIRO, BRAZIL"

Figura 4.6: Errores en país de afiliación.

Para realizar las correcciones se utilizará **expresiones regulares** [31].

Las expresiones regulares son patrones de caracteres utilizados para encontrar una determinada combinación dentro de una cadena de texto, que se pueden utilizar en cualquier lenguaje de programación.

Dentro del R, las expresiones regulares se aplican a un dataframe usando las funciones *apply* y *gsub*.

Ejemplo:

Como se observa en el primer ejemplo de la figura 4.6, a la ciudad de Quito se le asocia el país Suiza, lo cual es un error.

Para la corrección, se utiliza lo siguiente:

$$patron < -"QUITO, \\ s \\ w+;"$$

$$M < -apply(M, 2, function(M)gsub(patron, "QUITO, ECUADOR;" , M, ignore.case = T))$$

Donde *M* es el dataframe con el registro incorrecto, *patron* es la cadena de caracteres a buscarse en el dataframe, y "QUITO, ECUADOR;" la cadena de sustitución en caso de hallarse *patron*. Estas sentencias se utilizan de acuerdo a las necesidades de corrección del dataframe.

Para el presente estudio, se necesitaron 30 sentencias diferentes de corrección.

Luego de la corrección, los registros ejemplo de la figura 4.6, tienen la siguiente forma:

```
[1] "ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, LAUSANNE, SWITZERLAND; ESCUELA POLITÉCNICA NACIONAL, P.O. BOX 2759, QUITO
, ECUADOR"
[2] "DEPARTMENT OF MECHANICAL ENGINEERING, ESCUELA POLITÉCNICA NACIONAL, QUITO, ECUADOR; DEPARTMENT OF MECHANICAL ENGINE
ERING, PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO, RIO DE JANEIRO, BRAZIL"
```

Figura 4.7: Corrección errores país de afiliación.

Al final se tiene un dataframe corregido denominado **B_IES** que consta de **6.836 registros y 9 campos** (base a utilizarse en construcción de indicadores que necesiten datos por IES).

4.5.3 Eliminación de datos duplicados (generados por mecanismo de consulta).

Debido a que cada uno de los archivos .bib no poseen duplicados a nivel de registro total, todos los registros de la base B_IES son únicos considerando los 9 campos de la tabla 4.6.

En este punto cabe mencionar que si alguno de los campos de la tabla 4.5 es diferente para dos documentos que tengan los mismos campos de autor y título, Scopus considera esta situación como 2 registros diferentes.

En el presente estudio se conserva esta estructura, es decir, si alguno de los campos de la tabla 4.6 es diferente para dos documentos que tengan los mismos campos de autor y título, se considera como 2 registros diferentes.

Sin embargo, debido a la construcción mencionada en la sección 4.4, dentro de la base B_IES, se genera lo denominado **duplicación por colaboración**, ya que un documento puede ser generado en colaboración y tener varias IES ecuatorianas como autores; este documento se presentará en cada archivo .bib asociado a las IES involucradas, generando así duplicados a nivel de documento, es decir, se presenta duplicados al omitir el campo que guarda la IES de procedencia y considerar los campos restantes de la tabla 4.6 en la base B_IES.

Para la eliminación de este tipo de duplicados se utiliza la función *distinct* del paquete dplyr sobre la base generada únicamente con los campos: AU, TI, SO, DT, C1, TC, PY y UT de la base B_IES. Se obtiene un dataframe denominado B_T_C1 que consta de 6.293 registros y 8 campos.

4.5.4 Consolidación de citas por artículo.

Debido a la estructura en la que Scopus indexa documentos a su base (Ver 4.5.3) se puede generar problemas en la contabilización de citas.

Por ejemplo, al tener 2 registros diferentes para un mismo artículo debido a que el artículo se encuentra en la versión física de la revista y también en la versión online; en este caso, se contabiliza como 2 documentos diferentes dentro de Scopus y debido a esto, al momento de ingresar las citas puede tenerse en el primer registro un valor de cita diferente al segundo, o incluso puede que uno tenga valor y el otro no, cuando en realidad se trata del mismo artículo y la cita debería ser la misma en ambos registros.

Scopus no consolida las citas a nivel de artículo sino a nivel de registro total, por lo cual, si se quiere tener la cita real de un documento, debe sumarse los valores del campo TC para los documentos que presenten esta característica.

Para esto, se crea el campo CODIGO, que tendrá función de clave primaria y que se construye de la concatenación de los campos AU y TI, ya que si se tiene **los mismos autores y el mismo título son los mismos documentos**.

También se crean las variables:

- **SUMA**.- Cita consolidada para un documento.
- **MAXI**.- Valor de cita máxima encontrada para un documento.
- **MINI**.- Valor de cita mínima encontrada para un documento.

- **n.**- Número de registros encontrados para un documento.

Se obtiene un dataframe denominado B_T_C2 que consta de 6.293 registros y 13 campos. Como B_T_C2 aún conserva el campo original TC, se puede generar otro dataframe que solo contenga el campo SUMA como variable de citación.

Se obtiene un dataframe denominado B_T que consta de 6.293 registros y 12 campos (base a utilizarse en construcción de indicadores que necesiten datos de manera global considerando todas las IES).

Nota (Base Auxiliar).- Adicionalmente, se realiza una consulta particular bajo el criterio *AFFILCOUNTRY(ECUADOR)* en los años 2010-2016 en Scopus para la obtención de una base auxiliar a ser utilizada en uno de los indicadores bibliométricos; para esta base, y debido a los fines del indicador, únicamente se adaptará el procedimiento de consolidación (Ver 4.4) a los archivos .bib descargados. Se obtiene un dataframe denominado E_TOT que consta de 7.342 registros.

Capítulo 5

Resultados Obtenidos.

Nota: Se considera como documentos científicos a todos los posibles tipos de presentación de la investigación científica: artículos, libros, conferencias, etc.

En el presente capítulo se analiza los resultados obtenidos utilizando como base la metodología del capítulo 2.

El análisis se divide en dos partes: Análisis de Indicadores Bibliométricos y Usos de Leyes Bibliométricas en el estudio.

Es importante mencionar que debido a la fecha de corte, los datos pertenecientes al año 2017 se notan con 2017* (información parcial) y no se considerarán para el análisis en sí. El período de análisis corresponde a los años 2010-2016.

5.1 Análisis de los indicadores bibliométricos.

5.1.1 Número total de Publicaciones.

En la Figura 5.1, se tiene el número de publicaciones generado por las IES en los años 2010-2017*:

	A	PC
1	2010	231
2	2011	293
3	2012	383
4	2013	454
5	2014	720
6	2015	1199
7	2016	1833
8	2017	1180

A: Año; PC: Número de documentos (Producción Científica).

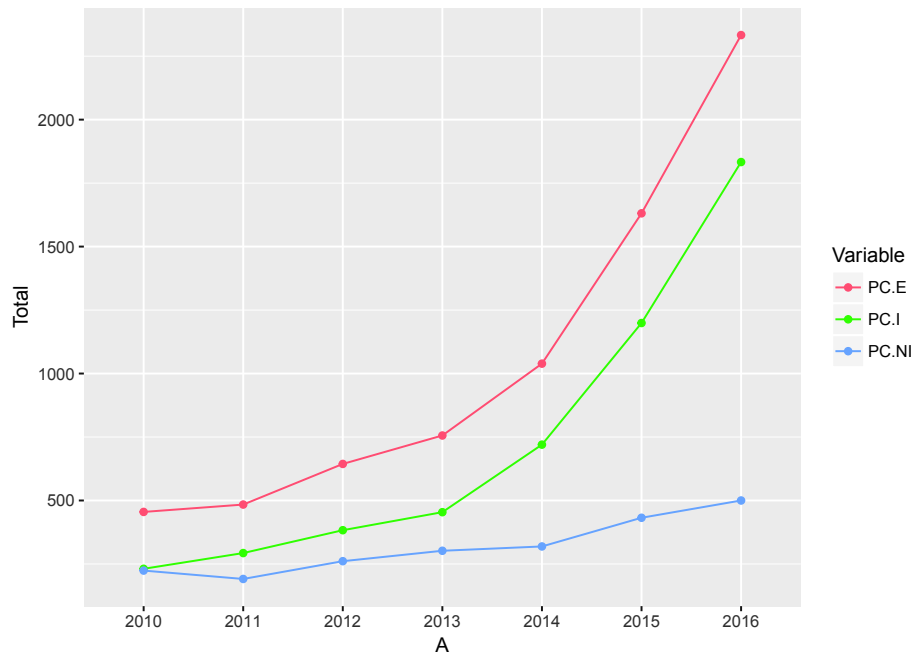
Figura 5.1: Producción Científica 2010-2017*.

En la Figura 5.2 se tiene las composiciones de publicación generada por IES, y publicación generada por no IES, dentro del período de análisis:

	A	PC.E	PC.I	PC.NI	%E.I	%E.NI
1	2010	455	231	224	50.77	49.23
2	2011	484	293	191	60.54	39.46
3	2012	644	383	261	59.47	40.53
4	2013	756	454	302	60.05	39.95
5	2014	1039	720	319	69.30	30.70
6	2015	1631	1199	432	73.51	26.49
7	2016	2333	1833	500	78.57	21.43

A: Año; PC.E: Producción Científica Ecuador; PC.I: Producción Científica IES; PC.NI: Producción Científica Instituciones no IES; %E.I: Porcentaje de PC de IES respecto a Ecuador; %E.NI: Porcentaje de PC de no IES respecto a Ecuador.

Figura 5.2: Producción Científica IES-NoIES 2010-2016.



PC.E: Producción Científica Ecuador; PC.I: Producción Científica IES; PC.NI: Producción Científica Instituciones no IES.

Figura 5.3: Comportamiento Producción Científica IES-NoIES 2010-2016.

Con base en la Figura 5.1, en el período de análisis, se generó 5.113 documentos en total. La Figura 5.3 permite observar una tendencia evolutiva creciente para la producción de IES y NoIES, siendo las IES los productores dominantes dentro de la producción total con una tasa de crecimiento medio anual del 41,23%.

Para el 2016, la producción es 7,93 veces la producción año 2010 (231 documentos), siendo el 2016 (1.833 documentos) el año más productivo, constituyendo junto al 2015 (1.199 documentos), el 59,3% (3.032 documentos) de la producción total.

En la Figura 5.4 se muestra la relación entre el número de autores y el número de documentos generados. Se observa que dentro del período de análisis, el 90,5 % de la producción científica lo constituyen documentos con máximo 11 autores.

	Nro.autores	Total	Doc.acu	Porc.doc	Porc.acum.doc
1	3	916	916	17.9	17.9
2	4	835	1751	16.3	34.2
3	5	669	2420	13.1	47.3
4	2	620	3040	12.1	59.5
5	6	513	3553	10.0	69.5
6	7	314	3867	6.1	75.6
7	1	249	4116	4.9	80.5
8	8	203	4319	4.0	84.5
9	9	138	4457	2.7	87.2
10	10	86	4543	1.7	88.9
11	11	86	4629	1.7	90.5

Nro.autores: Número de autores; Total: Número de documentos con Nro.autores; Doc.acu: Número de documentos acumulados; Porc.doc: Total en porcentaje; Porc.acum.doc: Doc.acu en porcentaje.

Figura 5.4: Número de autores dentro de la producción científica 2010-2016.

En la Figura 5.5 se presenta la composición de la producción por tipo de documento dentro del período de análisis:

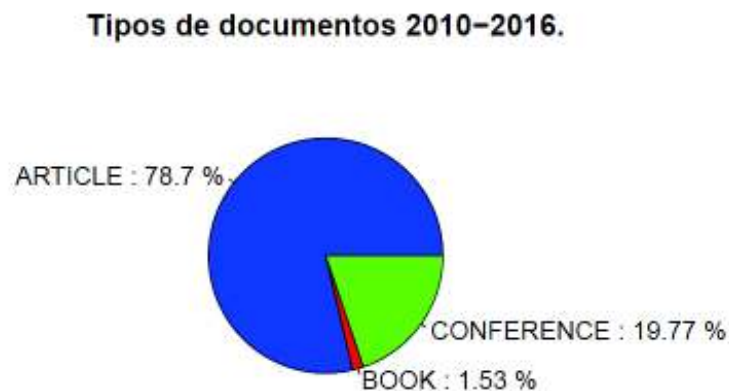


Figura 5.5: Producción Tipo de Documento 2010-2016.

Se puede observar que la mayor proporción de documentos son artículos con un 78,7 % (4.024 documentos), seguido por conferencias con un 19,77 % (1.011 documentos) y finalmente por libros o capítulos de libros con un 1,53 % (78 documentos) de la producción total. En el anexo A.1 se tiene un desglose de la producción científica por tipo de documento del 2010 al 2017*.

Debido a que las IES se encuentran distribuidas en categorías (Ver 4.2), en las Figuras 5.6,

5.7, 5.8 se tiene los resultados de producción, tomando como orden el número de documentos producidos del 2010 al 2017*:

***Categoría A.**

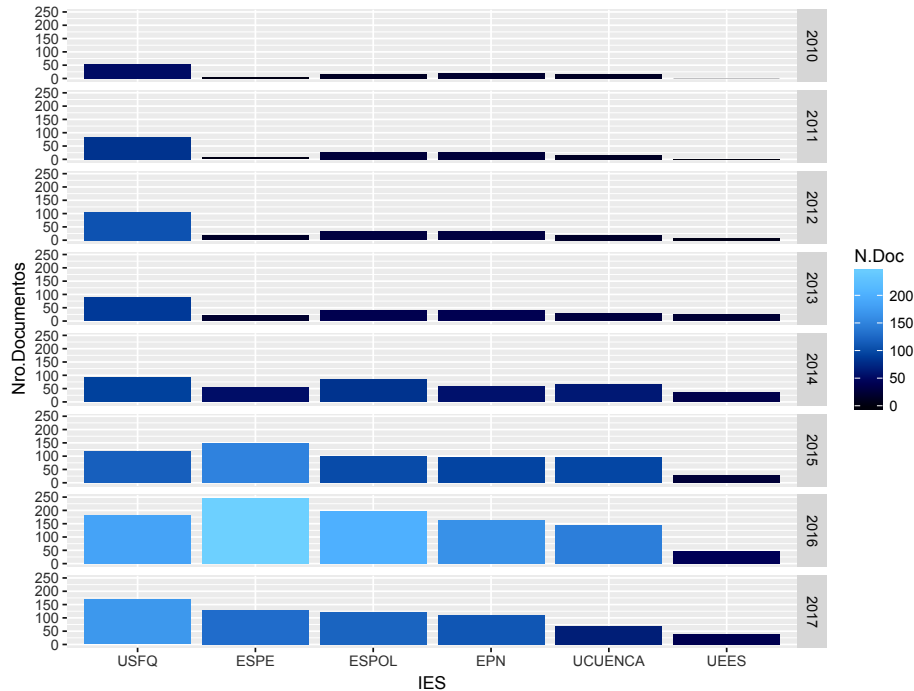


Figura 5.6: Producción Científica Cat. A.

Como se puede observar en la Figura 5.6, las IES han tenido un crecimiento positivo a lo largo del período de análisis, siendo la USFQ la IES con mayor número de documentos. Las IES: FLACSO y UASB, registran 1 y 2 aportaciones; FLACSO en 2011 y UASB en 2011 y 2012 respectivamente.

Todas las IES, exceptuando FLACSO y UASB, han tenido crecimientos de producción variados a nivel de cantidad, siendo la USFQ, la IES predominante en los años 2010-2014 y la ESPE en los años 2015-2016 dentro de esta categoría.

*Categoría B.

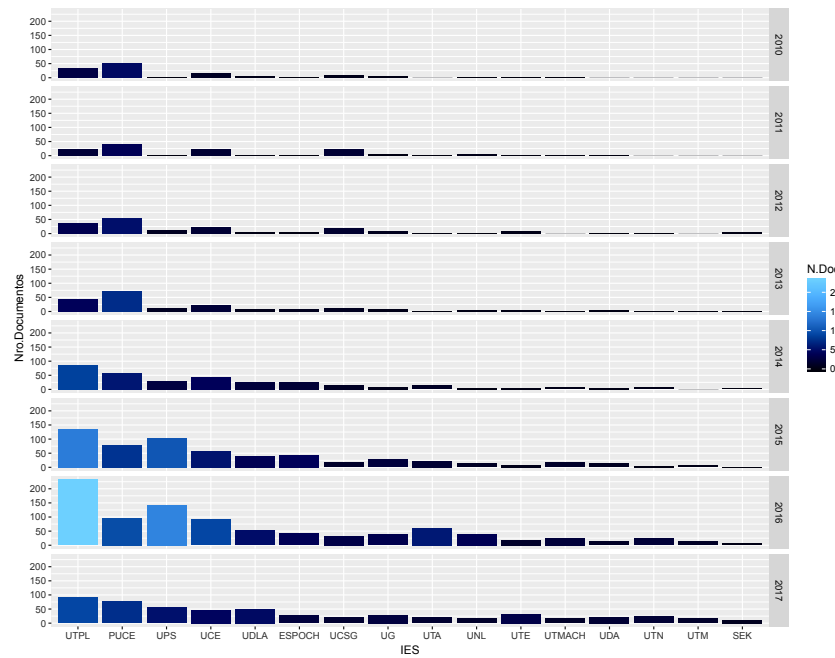


Figura 5.7: Producción Científica Cat. B.

Como se puede observar en la Figura 5.7, dentro de esta categoría, las IES: UIDE, UTI e IAEN, registran una sola aportación en los años 2012, 2012 y 2013 respectivamente. Dentro del período de análisis, las IES han tenido crecimientos de producción variados a nivel de cantidad, siendo la PUCE, la IES predominante en los años 2010-2013 y la UTP en el 2014-2016. Dentro de esta categoría, UTP es la IES con mayor número de documentos.

*Categoría C.

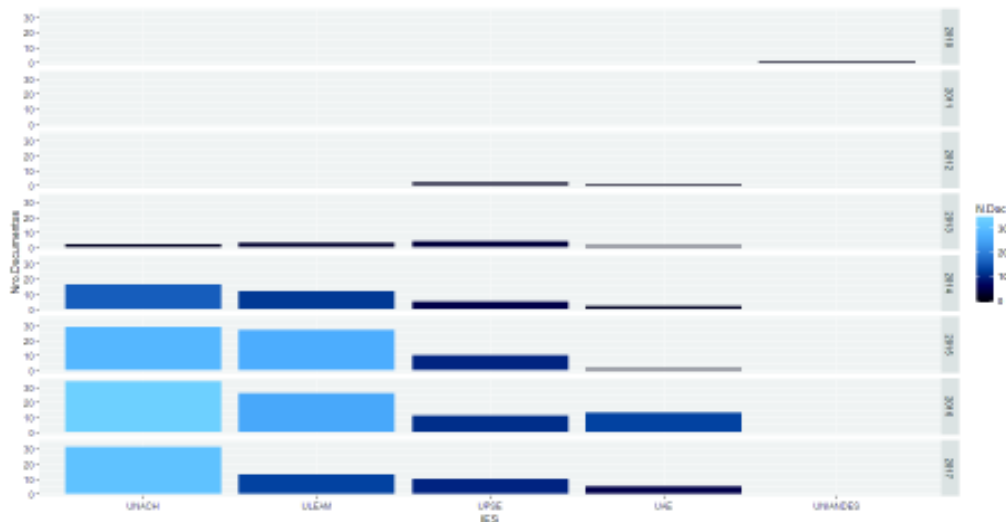


Figura 5.8: Producción Científica Cat. C.

Como se puede observar en la Figura 5.8, dentro de esta categoría, UNIANDES registra una sola aportación en el año 2010. Se evidencia que en el año 2011, esta categoría no generó producción científica en Scopus.

Las IES han tenido un crecimiento positivo a lo largo del período de análisis, siendo la UNACH, la IES con mayor número de documentos dentro del período de análisis.

De acuerdo a lo observado, se tiene las IES (exceptuando aquellas que registran 1 o 2 aportaciones en total), generaron su mayor cantidad de producción en los años 2015 y 2016. En el anexo A.2 se tiene un desglose de la producción científica a nivel de IES del 2010 al 2017*.

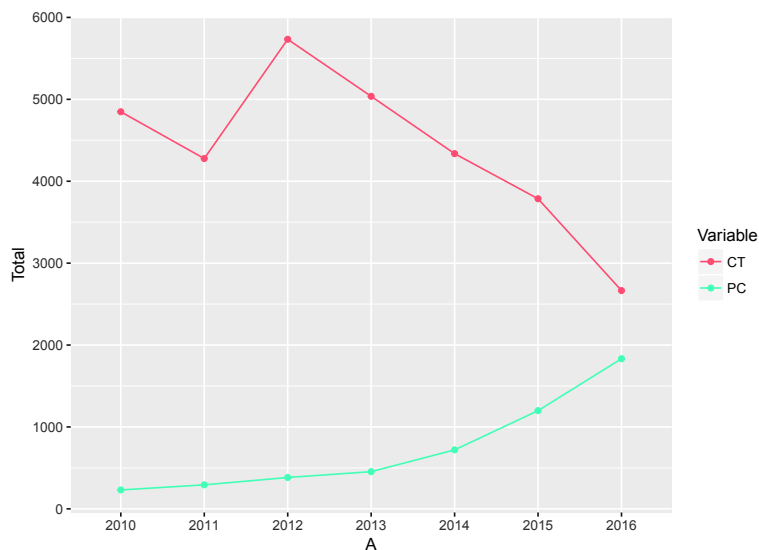
5.1.2 Número Total de Citas.

En las Figuras 5.9, 5.10 se presentan los resultados de las citas obtenidas:

	A	CT	PC	C.D
1	2010	4848	231	21.0
2	2011	4277	293	14.6
3	2012	5733	383	15.0
4	2013	5037	454	11.1
5	2014	4337	720	6.0
6	2015	3786	1199	3.2
7	2016	2665	1833	1.5
8	2017	322	1180	0.3

A: Año; CT: Número de Citas (Citación Total); PC: Producción Científica; C.D: Citas por documento en promedio.

Figura 5.9: Citaciones por año 2010-2017*.



CT: Citación Total; PC: Producción Científica.

Figura 5.10: Citaciones por año 2010-2016.

La Figura 5.10 evidencia que a partir del año 2012, la producción científica y las citaciones tienen una tendencia inversa de comportamiento, ya que a pesar de tener un incremento productivo, las citaciones disminuyeron.

La Figura 5.9 permite establecer que el año de mayor nivel de citación neta es el 2012 (5.733 citas); sin embargo, en promedio, un documento recibió más citaciones en el 2010 que en el 2012 (un artículo tuvo 21 citas en promedio en el año 2010).

La citación total dentro del período de análisis es de 30.683 citas.

En el Anexo A.3, se encuentra el detalle de las citaciones por IES y año en el período 2010-2017*.

A continuación se analizará el comportamiento de la citación dentro de la producción científica.

En el Anexo A.4, se encuentra una tabla cuyos campos x y y contienen los siguientes valores:

- x .- Número de citas.
- y .- Frecuencia absoluta de los artículos que poseen x citaciones.

ordenados de manera creciente respecto al campo x .

En la Figura 5.11 se grafica la tabla en su totalidad, mientras que en la Figura 5.12 se grafica la tabla considerando únicamente un valor x menor o igual a 300:

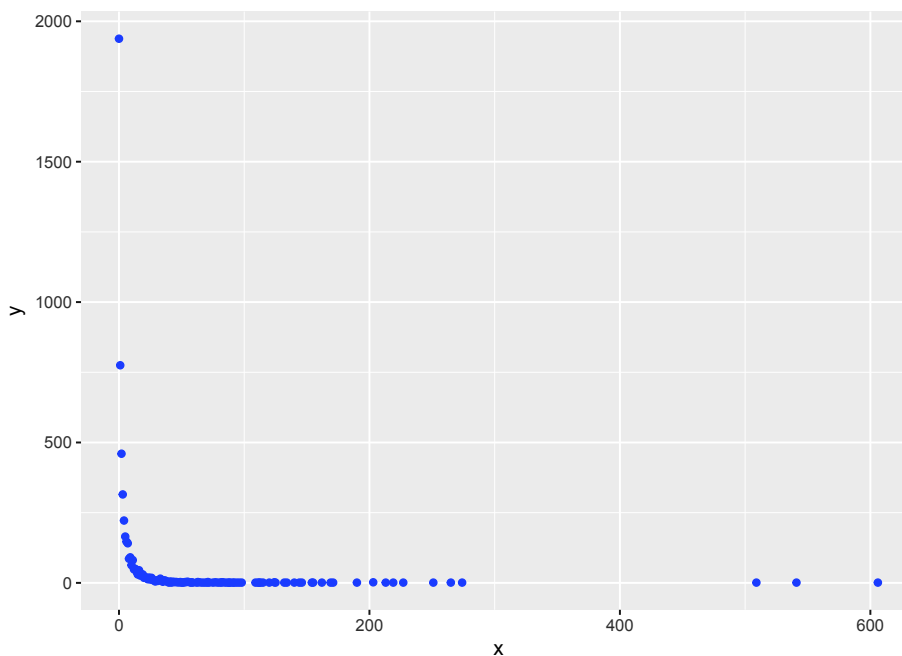


Figura 5.11: Citaciones vs Frecuencia de documentos.

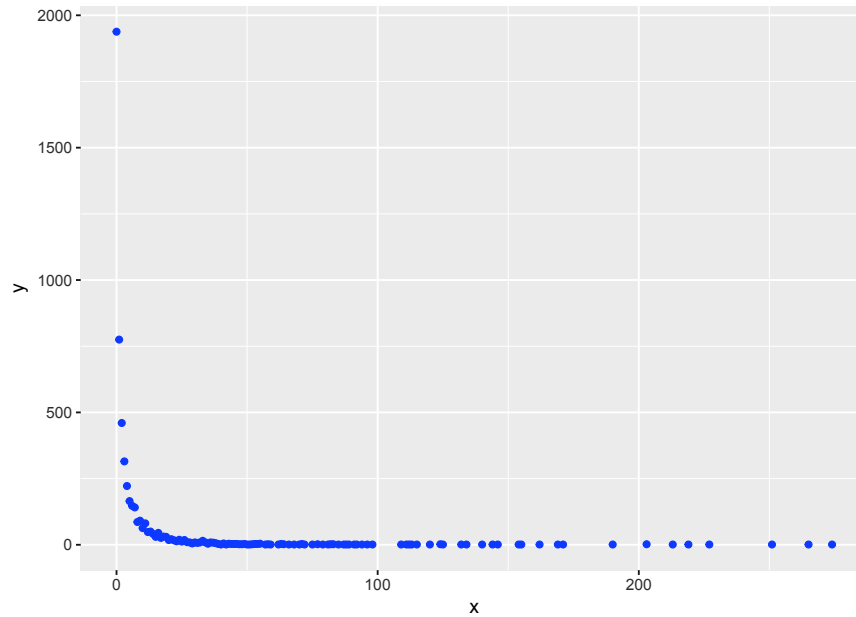


Figura 5.12: Citaciones (≤ 300) vs Frecuencia de documentos.

Como se puede observar, la proporción mayor de documentos tiene un número de citas menor a 100, siendo 1.938 el número de documentos con 0 citas; también se tiene 3 documentos con más de 500 citas.

En la Figura 5.13 se presenta la concentración de citas según el número de autores:

N.AU	N.L	%C	%C.A
1	0	2223	11.27
2	4	2020	9.78
3	11	1821	8.94
4	6	2023	7.87
5	7	1878	8.12
6	2	1888	8.07
7	8	1278	4.88
8	8	888	3.77
9	10	788	3.88
10	26	687	3.27
11	17	644	3.10
12	11	604	2.97
13	36	555	2.81
14	47	528	2.72
15	174	509	2.66
16	13	449	2.37
17	14	410	2.28
18	18	360	2.27
19	448	339	1.63
20	423	335	1.59
21	144	274	0.88
22	58	258	0.83
23	1	228	0.78

*N.AU: Número de autores; N.C: Citación total considerando los documentos con N.AU;
%C: N.C representado en porcentaje respecto a la citación del período de análisis; %C.A: %C acumulado.*

Figura 5.13: Concentración Citaciones vs Autores

Se evidencia que los documentos con un número de autores entre 2 y 10, representan el 57,92 % (17.774 citas) de la citación total. Cabe mencionar que los documentos con 1 autor representan únicamente el 0,78 % (238 citas) de la citación.

Se puede concluir que la producción científica que genera mayor citación se encuentra concentrada en los artículos con un número de autores mayor a 1 y menor o igual a 10. Una de las posibles causas del comportamiento de los 1.938 documentos (0 citaciones) es que los artículos recientemente producidos aún no han tenido tiempo de obtener citaciones.

5.1.3 Colaboración Internacional.

En la Figura 5.14 se representa la composición de los tipos de colaboración evidenciada en los documentos, dentro del período de análisis:

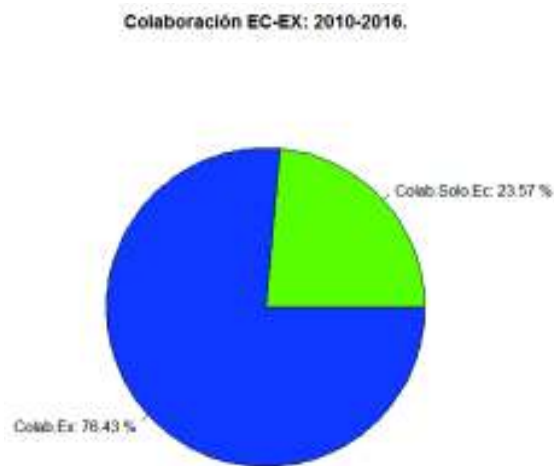


Figura 5.14: Colaboración Nacional - Extranjera 2010-2016.

Se observa que de la producción total, el 23,57 % (1.205 documentos) corresponde a producción ecuatoriana pura (artículos sin colaboración extranjera) y el 76,43 % (3.908 documentos) corresponde a producción mixta (artículos con colaboración extranjera).

En la Figura 5.15 se presenta este indicador desagregado a nivel anual:

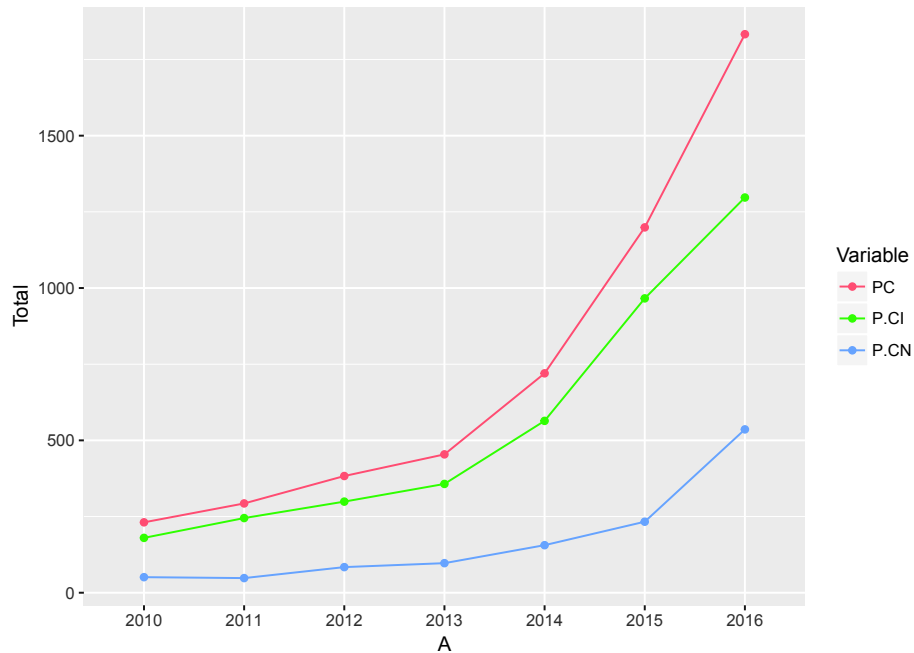
	A	PC	P.CI	P.CN	%CI	%CN
1	2010	231	180	51	77.9	22.1
2	2011	293	245	48	83.6	16.4
3	2012	383	299	84	78.1	21.9
4	2013	454	357	97	78.6	21.4
5	2014	720	564	156	78.3	21.7
6	2015	1199	966	233	80.6	19.4
7	2016	1833	1297	536	70.8	29.2
8	2017	1180	905	275	76.7	23.3

A: Año; PC: Producción Científica; P.CI: Número de documentos con colaboración extranjera;

P.CN: Número de documentos con solo colaboración nacional;

%CI: P.CI expresado en porcentaje respecto a PC; %CN: P.CN expresado en porcentaje respecto a PC.

Figura 5.15: Concentración de Colaboraciones por año: 2010-2017*.



A: Año; PC: Producción Científica; P.CI: Número de documentos con colaboración extranjera; P.CN: Número de documentos con solo colaboración nacional.

Figura 5.16: Colaboraciones 2010-2016.

Como podemos observar en la Figura 5.15, del 2010 al 2017*, más del 70 % de la producción anual son documentos generados en colaboración internacional (las IES tienden a generar más documentos con este tipo de producción).

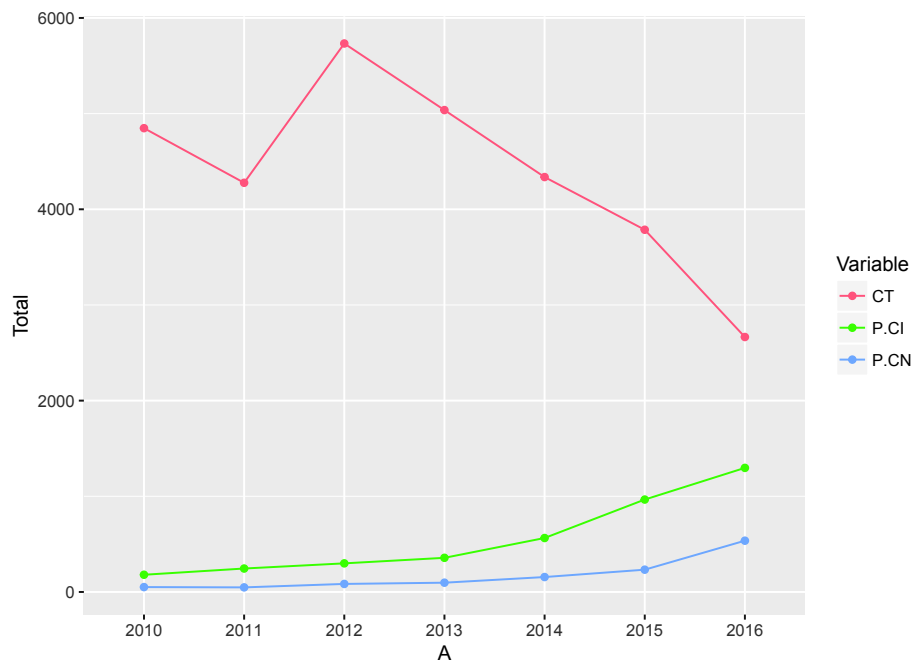
Dentro del período de análisis, el año en el que se tuvo un mayor nivel de colaboración internacional es el 2011 (83,6 %), seguido por el 2015 (80,6 %), mientras que el año con el menor nivel es el 2016 (70,8 %).

Por otro lado, en el 2016 se tiene el mayor número de artículos con afiliación únicamente ecuatoriana (536 documentos).

La tasa de crecimiento media anual de colaboración internacional y de colaboración nacional es del 38,79 % y 47,83 % respectivamente, por lo cual se evidencia que las IES ecuatorianas han buscado incrementar su colaboración interna.

5.1.3.1 Citación y Colaboración.

En la Figura 5.17, se observa el comportamiento de la citación y los dos tipos de colaboración (nacional y extranjera):



CT: Citación total; P.CI: Número de documentos con colaboración extranjera; P.CN: Número de documentos con solo colaboración nacional.

Figura 5.17: Citación vs Colaboración.

Se evidencia que debido a que los dos tipos de colaboración poseen tendencia de comportamiento creciente, las citaciones recibidas por la producción generada por las IES, es un resultado de los dos métodos de producción (nacional y colaborativa).

5.1.4 Índice H.

En la Figura 5.18, se presenta el H-index asociado a cada IES del Sistema de Educación Superior considerando los años 2010 al 2017*:

	IES	H		IES	H
1	USFQ	43	16	UTA	9
2	PUCE	31	17	UTMACH	9
3	UCUENCA	22	18	UPS	8
4	EPN	20	19	ULEAM	7
5	UCSG	19	20	UNL	7
6	UTPL	19	21	UTE	7
7	UCE	17	22	UDA	6
8	ESPE	16	23	UTN	5
9	ESPOL	15	24	UPSE	4
10	UDLA	12	25	SEK	3
11	UEES	12	26	UAE	3
12	UNACH	10	27	UTM	3
13	YACHAY	10	28	UASB	1
14	ESPOCH	9	29	FLACSO	1
15	UG	9	30	UTI	1

Figura 5.18: H-index. Nivel: IES.

Para las IES: IAEN, UIDE y UNIANDES no se tiene valor de H-index, debido a que dentro de Scopus, en los años considerados, no poseen artículos citados.

En la Figura 5.19 se muestra los niveles de impacto del Sistema de Educación Superior según el H-index final obtenido, considerando los años 2010 al 2017*, bajo la asignación del percentil 25 (5,25), percentil 50 (9), y el percentil 75 (15,75) de manera que:

- Impacto Bajo: Si $H\text{-index} \leq 5,25$.
- Impacto Moderado: Si $5,25 < H\text{-index} \leq 9$.
- Impacto Medio: Si $9 < H\text{-index} \leq 15,75$.
- Impacto Alto: Si $H\text{-index} > 15,75$.

	IES	H	Nivel Impacto	IES	H	Nivel Impacto	
1	USFQ	43	Alto	16	UTA	9	Moderado
2	PUCE	31	Alto	17	UTMACH	9	Moderado
3	UCUENCA	22	Alto	18	UPS	8	Moderado
4	EPN	20	Alto	19	ULEAM	7	Moderado
5	UCSG	19	Alto	20	UNL	7	Moderado
6	UTPL	19	Alto	21	UTE	7	Moderado
7	UCE	17	Alto	22	UDA	6	Moderado
8	ESPE	16	Alto	23	UTN	5	Bajo
9	ESPOL	15	Medio	24	UPSE	4	Bajo
10	UDLA	12	Medio	25	SEK	3	Bajo
11	UEES	12	Medio	26	UAE	3	Bajo
12	UNACH	10	Medio	27	UTM	3	Bajo
13	YACHAY	10	Medio	28	UASB	1	Bajo
14	ESPOCH	9	Moderado	29	FLACSO	1	Bajo
15	UG	9	Moderado	30	UTI	1	Bajo

Nivel Impacto: Bajo ($H \leq$ percentil 25), Moderado (percentil 25 $< H \leq$ percentil 50), Medio (percentil 50 $< H \leq$ percentil 75), Alto ($H >$ percentil 75).

Figura 5.19: H-index: Impacto Sistema de Educación Superior.

A continuación se muestra los valores H-index de las IES dentro de sus categorías, considerando los siguientes subperíodos acumulativos:

1	2010-2011.	3	2010-2013.	5	2010-2015.
2	2010-2012.	4	2010-2014.	6	2010-2016.

- Cabe mencionar que para las Categoría B y Categoría C se muestra el H-index desde el subperíodo 3 en adelante, ya que en los subperíodos 1 y 2, existían IES sin producción. A partir del subperíodo 3, todas las IES en cada categoría presentaban al menos un documento.

5.1.4.1 *Categoría A.

	IES1	H1	IES2	H2	IES3	H3	IES4	H4	IES5	H5	IES6	H6
1	USFQ	35	USFQ	39	USFQ	42	USFQ	42	USFQ	42	USFQ	43
2	EPN	14	EPN	16	UCUENCA	18	UCUENCA	20	UCUENCA	21	UCUENCA	21
3	UCUENCA	11	UCUENCA	15	EPN	17	EPN	18	EPN	20	EPN	20
4	ESPOL	9	ESPOL	11	ESPOL	13	ESPOL	15	ESPE	16	ESPE	16
5	ESPE	8	ESPE	11	ESPE	12	ESPE	15	ESPOL	15	ESPOL	15
6	FLACSO	1	UEES	2	UEES	8	UEES	12	UEES	12	UEES	12
7	UASB	1	FLACSO	1	FLACSO	1	FLACSO	1	FLACSO	1	FLACSO	1
8	UEES	0	UASB	1	UASB	1	UASB	1	UASB	1	UASB	1

IES_i: IES en orden decreciente según el H_i para el subperíodo i; H_i: H-index para el subperíodo i, i: 1,...,6.

Figura 5.20: H-index: Categoría A (subperíodos).

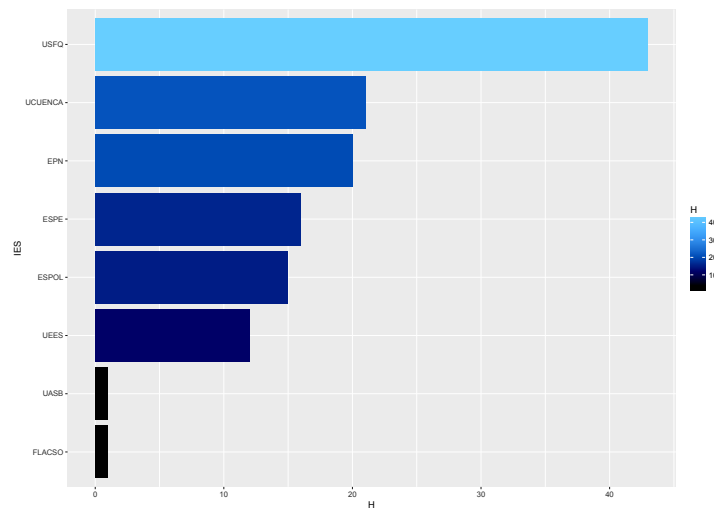


Figura 5.21: H-index: Categoría A (período de análisis).

Con base en la Figura 5.20, se evidencia que dentro de los subperíodos, la USFQ ha mantenido la primera posición dentro de su categoría según su H-index. La segunda y tercera posición se han mantenido entre la EPN y UCUENCA.

Finalmente, en la Figura 5.22 se muestra los niveles de impacto dentro de esta categoría según el H-index final obtenido considerando el período de análisis, bajo la asignación del percentil 25 (9,25), percentil 50 (15,5), y el percentil 75 (20,25) de manera que:

- Impacto Bajo: Si $H\text{-index} \leq 9,25$.
- Impacto Moderado: Si $9,25 < H\text{-index} \leq 15,5$.
- Impacto Medio: Si $15,5 < H\text{-index} \leq 20,25$.
- Impacto Alto: Si $H\text{-index} > 20,25$.

IES	H	Nivel.Impacto
1 USFQ	43	Alto
2 UCUENCA	21	Alto
3 EPN	20	Medio
4 ESPE	16	Medio
5 ESPOL	15	Moderado
6 UEES	12	Moderado
7 FLACSO	1	Bajo
8 UASB	1	Bajo

Nivel.Impacto: Bajo ($H \leq \text{percentil } 25$), Moderado ($\text{percentil } 25 < H \leq \text{percentil } 50$), Medio ($\text{percentil } 50 < H \leq \text{percentil } 75$), Alto ($H > \text{percentil } 75$).

Figura 5.22: H-index: Impacto Categoría A.

5.1.4.2 *Categoría B.

	IES3	H3	IES4	H4	IES5	H5	IES6	H6
1	PUCE	28	PUCE	30	PUCE	31	PUCE	31
2	UCSG	18	UCSG	19	UCSG	19	UCSG	19
3	UCE	15	UTPL	18	UTPL	19	UTPL	19
4	UTPL	14	UCE	16	UCE	17	UCE	17
5	UDLA	9	UDLA	11	UDLA	12	UDLA	12
6	UG	9	ESPOCH	9	ESPOCH	9	ESPOCH	9
7	ESPOCH	8	UG	9	UTA	9	UTA	9
8	UTE	6	UTMACH	8	UG	9	UG	9
9	UTMACH	6	UTA	7	UTMACH	8	UTMACH	9
10	UNL	4	UTE	7	UNL	7	UPS	8
11	UPS	4	UNL	6	UTE	7	UNL	7
12	UDA	3	UPS	5	UPS	6	UTE	7
13	UTA	3	UDA	4	UDA	5	UDA	6
14	SEK	2	UTN	4	UTN	5	UTN	5
15	UTN	2	SEK	2	SEK	3	SEK	3
16	UTM	2	UTM	2	UTM	3	UTM	3
17	UTI	1	UTI	1	UTI	1	UTI	1

IES_i: IES en orden decreciente según el H_i para el subperíodo i; H_i: H-index para el subperíodo i, i: 3,...,6.

Figura 5.23: H-index: Categoría B (subperíodos).

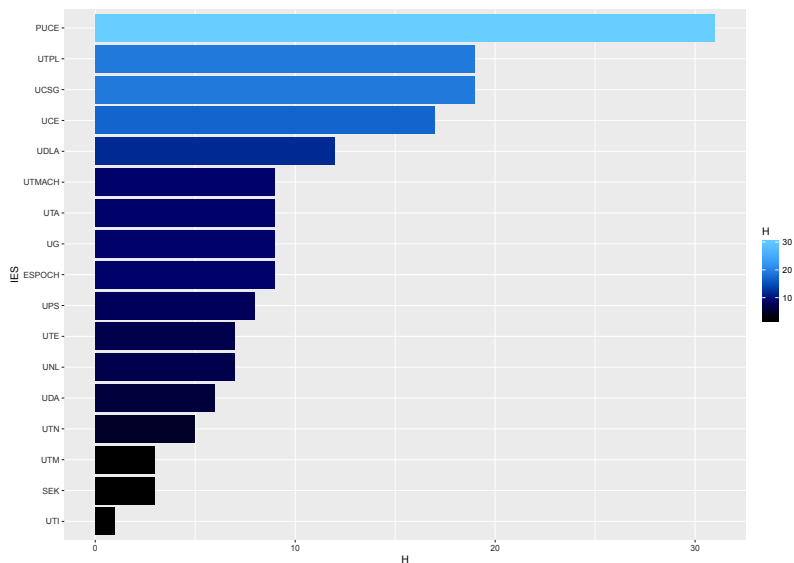


Figura 5.24: H-index: Categoría B (período de análisis).

Las IES: IAEN y UIDE no poseen H-index debido a no tener producción citada.

Con base en la Figura 5.23, se evidencia que dentro de los subperíodos, la PUCE ha mantenido la primera posición dentro de su categoría según su H-index. La segunda y tercera

posición se han mantenido entre la UCSG y UTPL respectivamente.

Finalmente, en la Figura 5.25 se muestra los niveles de impacto dentro de esta categoría según el H-index final obtenido considerando el período de análisis, bajo la asignación del percentil 25 (6), percentil 50 (9), y el percentil 75 (12) de manera que:

- Impacto Bajo: Si $H\text{-index} \leq 6$.
- Impacto Moderado: Si $6 < H\text{-index} \leq 9$.
- Impacto Medio: Si $9 < H\text{-index} \leq 12$.
- Impacto Alto: Si $H\text{-index} > 12$.

IES	H	Nivel.Impacto	
1	PUCE	31	Alto
2	UCSG	19	Alto
3	UTPL	19	Alto
4	UCE	17	Alto
5	UDLA	12	Medio
6	ESPOCH	9	Moderado
7	UTA	9	Moderado
8	UG	9	Moderado
9	UTMACH	9	Moderado
10	UPS	8	Moderado
11	UNL	7	Moderado
12	UTE	7	Moderado
13	UDA	6	Bajo
14	UTN	5	Bajo
15	SEK	3	Bajo
16	UTM	3	Bajo
17	UTI	1	Bajo

Nivel.Impacto: Bajo ($H \leq$ percentil 25), Moderado (percentil 25 < $H \leq$ percentil 50), Medio (percentil 50 < $H \leq$ percentil 75), Alto ($H >$ percentil 75).

Figura 5.25: H-index: Impacto Categoría B.

5.1.4.3 *Categoría C.

IES3	H3	IES4	H4	IES5	H5	IES6	H6	
1	UNACH	2	UNACH	5	UNACH	10	UNACH	10
2	ULEAM	2	ULEAM	4	ULEAM	6	ULEAM	7
3	UPSE	1	UAE	3	UPSE	4	UPSE	4
4	UAE	1	UPSE	2	UAE	3	UAE	3

IESi: IES en orden decreciente según el H_i para el subperíodo i ; H_i : H-index para el subperíodo i , $i: 3, \dots, 6$.

Figura 5.26: H-index: Categoría C (subperíodos).

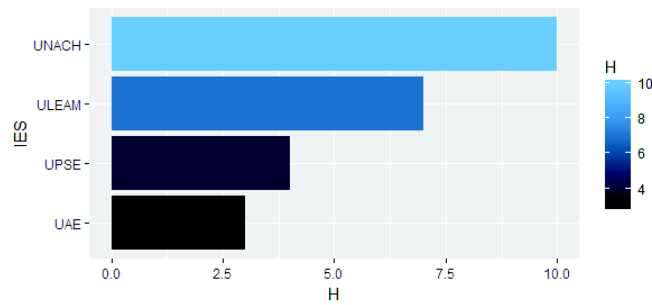


Figura 5.27: H-index: Categoría C (período de análisis).

La IES: UNIANDES no posee H-index debido a no tener producción citada.

Con base en la Figura 5.26, se evidencia que dentro de los subperíodos, la UNACH ha mantenido la primera posición dentro de su categoría según su H-index. La segunda y tercera posición se han mantenido entre la ULEAM y UPSE respectivamente.

Finalmente, en la Figura 5.28 se muestra los niveles de impacto dentro de esta categoría según el H-index final obtenido considerando el período de análisis, bajo la asignación del percentil 25 (3,75), percentil 50 (5,5), y el percentil 75 (7,75) de manera que:

- Impacto Bajo: Si $H\text{-index} \leq 3,75$.
- Impacto Moderado: Si $3,75 < H\text{-index} \leq 5,5$.
- Impacto Medio: Si $5,5 < H\text{-index} \leq 7,75$.
- Impacto Alto: Si $H\text{-index} > 7,75$.

	IES	H	Nivel.Impacto
1	UNACH	10	Alto
2	ULEAM	7	Medio
3	UPSE	4	Moderado
4	UAE	3	Bajo

Nivel.Impacto: Bajo ($H \leq \text{percentil } 25$), Moderado ($\text{percentil } 25 < H \leq \text{percentil } 50$), Medio ($\text{percentil } 50 < H \leq \text{percentil } 75$), Alto ($H > \text{percentil } 75$).

Figura 5.28: H-index: Impacto Categoría C.

5.2 Áreas de Mayor Producción Científica.

Uno de los intereses del presente trabajo es conocer las áreas o campos de investigación que tienen más difusión o impacto dentro de la producción científica. Para poder establecer dichas áreas, se utiliza la codificación ASJC (All Science Journals Classification), propia de

Scopus y que permite clasificar a las revistas indexadas de acuerdo a sus áreas o campos de investigación, es decir, estos códigos de clasificación permiten asociar a la revista donde el documento se encuentra publicado con una o varias líneas de investigación, siendo así que una revista multidisciplinaria puede tener varios códigos ASJC.

El criterio a usarse en este apartado, es realizar un empate entre las revistas asociadas a los documentos de la base y los códigos ASJC proporcionados por el paquete *CITAN*.

En la Figura 5.29 se presenta el número de documentos por área de producción dentro del período de análisis, tomando como condición que el área tenga más del 0,95 % de documentos de la base:

	AREA	TOT	POR	POR.A
1	Computer Science	494	5.25	5.25
2	Engineering	460	4.89	10.14
3	Ecology, Evolution, Behavior and Systematics	417	4.43	14.57
4	Medicine	411	4.37	18.94
5	Animal Science and Zoology	230	2.44	21.38
6	Ecology	189	2.01	23.39
7	Plant Science	175	1.86	25.25
8	Infectious Diseases	172	1.83	27.08
9	Physics and Astronomy	168	1.79	28.87
10	Electrical and Electronic Engineering	158	1.68	30.55
11	Agricultural and Biological Sciences	136	1.45	32.00
12	Condensed Matter Physics	135	1.41	33.41
13	Public Health, Environmental and Occupational Health	126	1.34	34.75
14	Biochemistry, Genetics and Molecular Biology	123	1.31	36.06
15	Computer Science Applications	114	1.21	37.27
16	Chemistry	111	1.18	38.45
17	Genetics	108	1.15	39.60
18	Aquatic Science	104	1.11	40.71
19	Agronomy and Crop Science	96	1.02	41.73
20	Parasitology	95	1.01	42.74
21	Food Science	94	1.00	43.74
22	Applied Mathematics	91	0.97	44.71
23	Materials Science	90	0.96	45.67

AREA: Área de Producción; TOT: Número de documentos por área; POR: TOT en porcentaje; POR.A: POR acumulado.

Figura 5.29: Documentos por áreas de producción 2010-2016.

Se observa que existen áreas con más de 400 documentos dentro de la base, convirtiendo a las Ciencias de la Computación, Ingeniería, Ecología y Evolución y Medicina en las áreas de producción e interés predominantes.

En la Figura 5.30 se presenta las áreas de mayor producción dentro del período de análisis, tomando como condición que existan más de 130 documentos generados en el área:

Áreas de producción: 2010-2016.

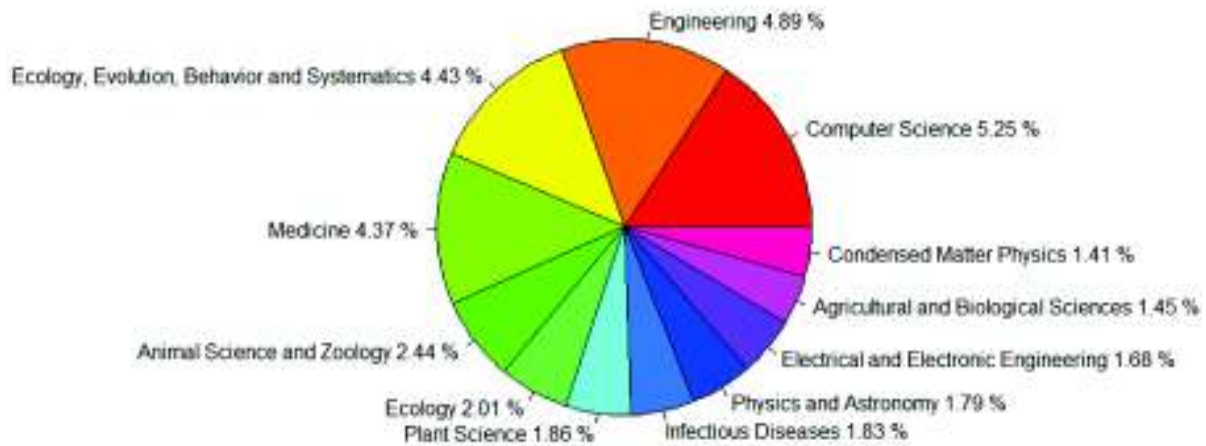


Figura 5.30: Áreas de alta producción 2010-2016.

Se observa que en general, existe un marcado interés por las Ciencias Naturales (Ciencias Animales, Ecología, Ciencias de las Plantas, Agricultura y Ciencias Biológicas), seguido por el interés en Ciencias de la Física (Física de la Materia Condensada, Física y Astronomía).

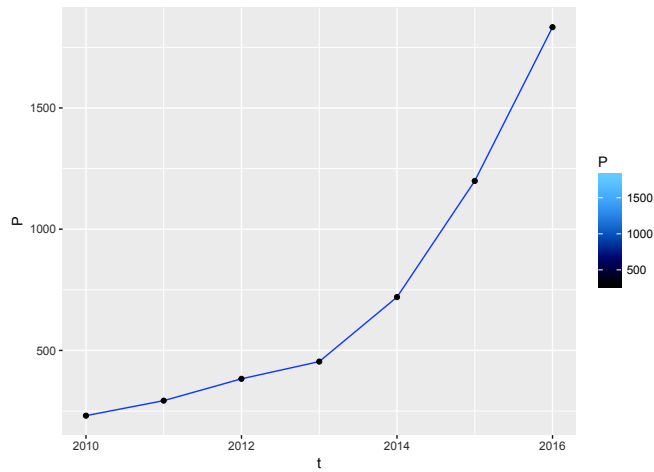
Agricultura y Medicina (intereses de producción clásicos) aún se encuentran dentro de las mayores áreas de producción.

5.3 Análisis y prueba de las Leyes Bibliométricas.

5.3.1 Crecimiento de la producción: Price.

El modelo de crecimiento de Price permite obtener un modelo predictivo de tipo exponencial para la producción científica.

De manera gráfica, basándonos en la Figura 5.31, se esperaría que los datos reales se ajusten al modelo de Price:



P: Producción Científica; t: 0(año 2010), 2(año 2012), 4(año 2014), 6(año 2016).

Figura 5.31: Datos reales PC.

Considerando el período de análisis, se procederá con la verificación analítica de la idea intuitiva. En la Figura 5.32 se muestra los resultados del ajuste realizado:

```
> regresion1
Nonlinear regression model
model: P ~ Po * exp(a * t)
data: Datos
a
0.3346
residual sum-of-squares: 76265

Number of iterations to convergence: 9
Achieved convergence tolerance: 1.913e-06
> summary(regresion1)

Formula: P ~ Po * exp(a * t)

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a 0.334604  0.008878  37.69 2.33e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 112.7 on 6 degrees of freedom

Number of iterations to convergence: 9
Achieved convergence tolerance: 1.913e-06

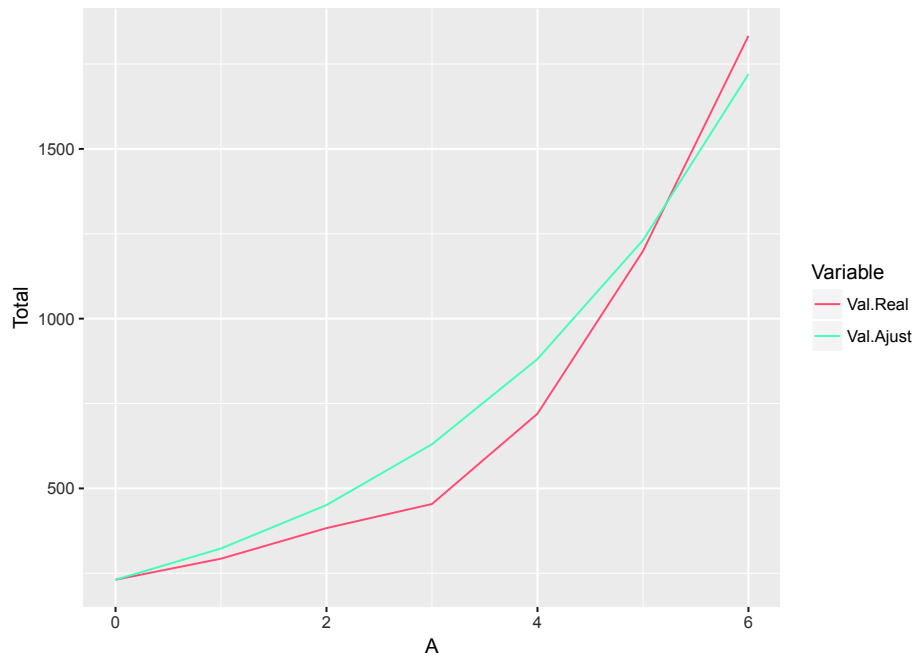
> KS

Two-sample Kolmogorov-Smirnov test

data: Datos$P and #R$fitted.reg
D = 0.14286, p-value = 1
alternative hypothesis: two-sided

> R.cuad
[1] 0.9632145
```

Figura 5.32: Ajuste de la producción: Modelo de Price.



Total: Producción Científica; t: 0(año 2010), 2(año 2012), 4(año 2014), 6(año 2016).

Figura 5.33: PC Real vs PC Ajustado.

Debido a que se cumple la prueba Kolmogórov-Smirnov a un nivel de confianza del 95 % y el coeficiente de determinación obtenido es: $R^2 = 0,96$; se puede asegurar que la proporción de crecimiento de la producción científica es de 0,3346; es decir, en el período de análisis, se tiene un nivel de crecimiento anual moderado y que el tiempo de duplicación de la producción científica $\left(D = \frac{\ln(2)}{0,3346} = 2,071\right)$ es de 2 años.

Utilizando la proporción de crecimiento evidenciada, se establece las siguientes proyecciones de producción científica:

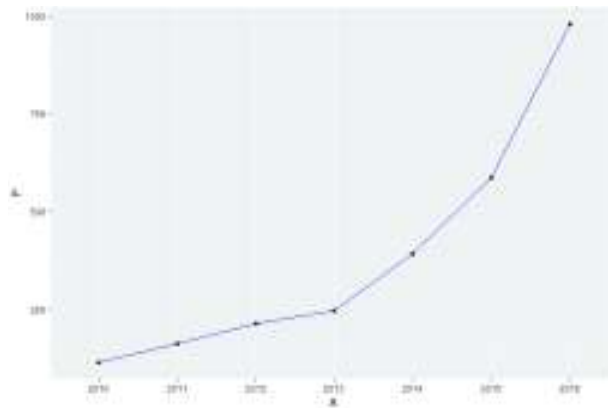
	A	PCP
1	2017	2403
2	2018	3358
3	2019	4693
4	2020	6558

A: Año; PCP: Proyecciones resultantes del modelo ajustado.

Figura 5.34: Proyecciones: Modelo de Price.

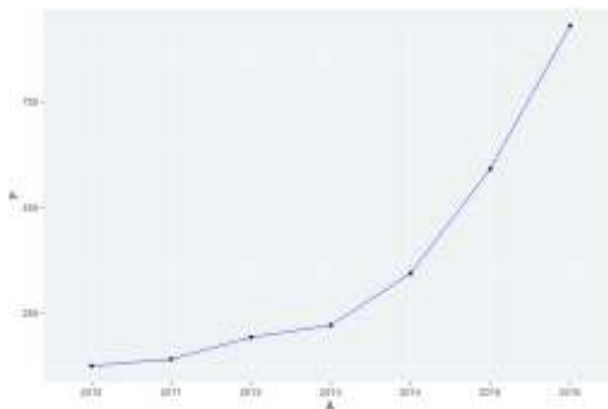
5.3.2 Modelo de Price en las Categorías del CEAACES.

A continuación, se analiza el comportamiento de la producción científica de cada una de las 3 categorías definidas por el CEAACES:



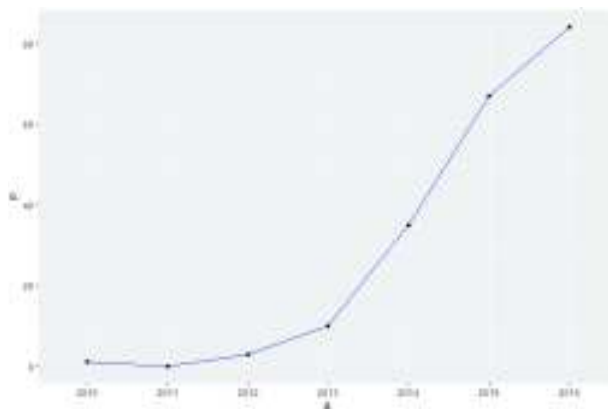
A: Año; P: Producción Científica.

Figura 5.35: Comportamiento Producción Científica Categoría A 2010-2016.



A: Año; P: Producción Científica.

Figura 5.36: Comportamiento Producción Científica Categoría B 2010-2016.



A: Año; P: Producción Científica.

Figura 5.37: Comportamiento Producción Científica Categoría C 2010-2016.

De manera gráfica se observa que el comportamiento de las categorías A y B a nivel interno puede ajustarse al Modelo exponencial de Price, sin embargo, se observa que el

comportamiento de la categoría C puede no ajustarse satisfactoriamente a un modelo exponencial.

A continuación se procederá con la verificación analítica de la idea intuitiva.

En la tabla 5.1 se muestran los resultados obtenidos al realizar el ajuste de tipo exponencial:

IES	a	Ks (p-valor)	R^2
Categoría A	0,34594	1	0,97
Categoría B	0,32091	0,9375	0,94
Categoría C	0,75874	1	0,88

Cuadro 5.1: Resultados ajuste exponencial por categorías.

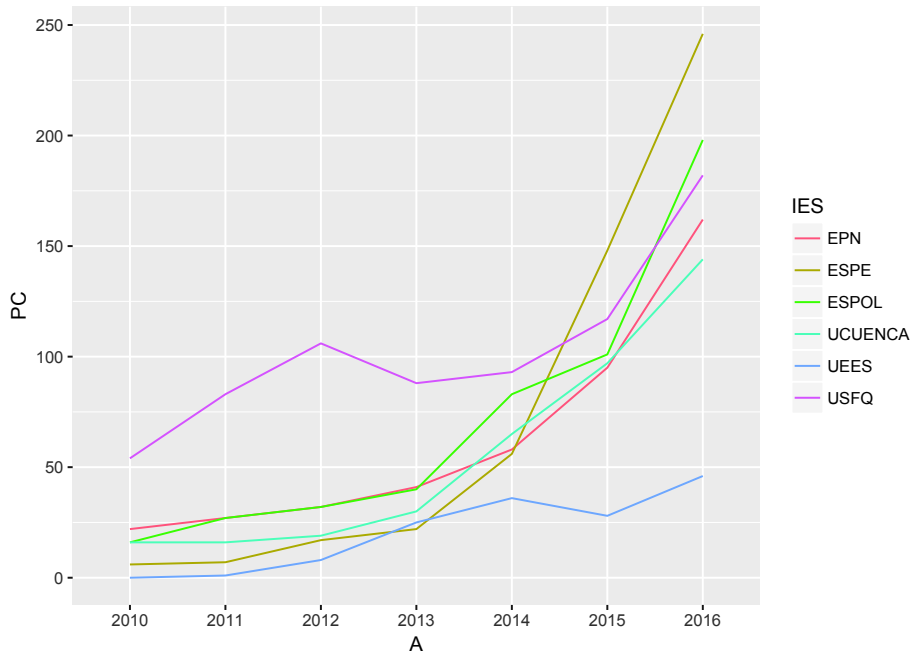
Se puede observar que para las categorías A y B en el sentido de grupos individuales, el ajuste exponencial de Price es el adecuado para la modelización del comportamiento de la producción científica, mientras que para la Categoría C, el ajuste exponencial no es el adecuado.

5.3.2.1 Modelo de Price en la Categoría A.

Nota: En las secciones 5.3.2.1 y 5.3.2.2 se desarrolla el análisis considerando únicamente las IES pertenecientes a la Categoría A. Es importante mencionar que el desarrollo de estas secciones puede realizarse para las Categorías B y C (el código de programación en R permite realizar esta particularización).

A continuación, se analiza el comportamiento de producción científica de cada IES perteneciente a la categoría A con más de 2 artículos dentro del período de análisis (FLACSO y UASB no se consideran).

En la Figura 5.38, se muestra la producción científica dentro del período de análisis asociado a las IES consideradas:



A: Año; PC: Producción Científica.

Figura 5.38: Comportamiento Producción Científica 2010-2016.

De manera gráfica se observa que el comportamiento de las IES: USFQ y UEES difieren del comportamiento de las otras IES consideradas (el comportamiento de la USFQ y UEES puede no ajustarse con un modelo exponencial).

A continuación se procederá con la verificación analítica de la idea intuitiva.

En la tabla 5.2 se muestran los resultados obtenidos al realizar el ajuste de tipo exponencial:

IES	a	Ks (p-valor)	R ²
EPN	0,3138953	0,9375	0,93
ESPOL	0,4072004	1	0,96
USFQ	0,1846834	0,9375	0,77
UCUENCA	0,360302	0,9375	0,96
ESPE	0,6203313	1	0,98
UEES	0,7887075	0,4413	0,21

Cuadro 5.2: Resultados ajuste exponencial.

Se puede observar que para las IES: USFQ y UEES, el ajuste exponencial de Price no es el adecuado para la modelización del comportamiento de la producción científica.

Las IES restantes, cumplen con el comportamiento exponencial de Price; se puede observar que la ESPE tiene una proporción de crecimiento mayor, seguida por la ESPOL.

Para las IES: EPN, ESPOL, UCUENCA y ESPE se puede utilizar el modelo de Price para

establecer proyecciones al año 2020:

En la Figura 5.39 se muestran los resultados de las proyecciones:

	A	EPN	ESPOL	UCUENCA	ESPE
1	2017	198	277	199	461
2	2018	271	416	286	858
3	2019	371	625	410	1595
4	2020	508	939	587	2966

En las filas, el año de proyección; en las columnas, las proyecciones del modelo ajustado por IES.

Figura 5.39: Proyecciones: Modelo de Price EPN, ESPOL, UCUENCA, ESPE.

5.3.2.2 Alternativas para las IES de la Categoría A que no cumplen con Price.

Una de las posibilidades para establecer proyecciones de las IES: USFQ y UEES es ajustar un modelo paramétrico ya sea diferente en su estructura o en su período de análisis.

Para la UEES, el análisis se realizará modificando su estructura, es decir, se utilizará un modelo lineal, tomando en consideración la idea intuitiva del comportamiento de su producción proporcionada por la Figura 5.38.

En la Figura 5.40 se muestra los resultados del ajuste realizado:

```
> regresion2
Call:
lm(formula = P ~ t - 1, data = Datos)

Coefficients:
t
7.165

> summary(regresion2)

Call:
lm(formula = P ~ t - 1, data = Datos)

Residuals:
    Min       1Q   median       3Q      Max
-7.824 -6.247  0.000  3.258  7.341

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
t    7.1648    0.6268    11.43 2.69e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.98 on 6 degrees of freedom
Multiple R-squared:  0.9561,    Adjusted R-squared:  0.9488
F-statistic: 130.7 on 1 and 6 DF,  p-value: 2.69e-05

> R.cuad
[1] 0.9560934
```

Figura 5.40: Ajuste de la producción: Modelo lineal en el período de análisis.

Se puede observar que el modelo lineal sin intercepto es adecuado para el ajuste de la producción de la UEES debido a cumple el Test F con un $F > 1$ a un nivel de confianza del

95 % y con un coeficiente de determinación cercano a 1 ($R^2 = 0,96$).

En la Figura 5.41, usando el parámetro estimado asociado a la variable t, se tienen las proyecciones para el 2017-2020:

	A	PCP
1	2017	50
2	2018	57
3	2019	64
4	2020	72

Figura 5.41: Proyecciones UEES

Para la USFQ, el análisis se realizará modificando el período, es decir, se utilizará un modelo exponencial (propuesto por Price), considerando un período menor al período de análisis, específicamente la mitad del período (del 2013 al 2016), ya que un modelo ajustado en el subperíodo permitirá proyectar a los años posteriores consecutivos del mismo.

En la Figura 5.42 se muestra los resultados del ajuste realizado:

```
> regresion1
Nonlinear regression model
model: P ~ Po * exp(a * t)
data: Datos
a
0.2165
residual sum-of-squares: 796.6

Number of iterations to convergence: 7
Achieved convergence tolerance: 7.619e-06
> summary(regresion1)

Formula: P ~ Po * exp(a * t)

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a 0.21645    0.02791    7.756 0.00446 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.3 on 3 degrees of freedom

Number of iterations to convergence: 7
Achieved convergence tolerance: 7.619e-06

> KS

Two-sample Kolmogorov-Smirnov test

data: Datos$P and PR$fitted.reg
D = 0.25, p-value = 0.9996
alternative hypothesis: two-sided

> R.cuad
[1] 0.8579025
```

Figura 5.42: Ajuste de la producción: Modelo exponencial en el subperíodo.

Debido a que se cumple la prueba Kolmogórov-Smirnov, a un nivel de confianza del 95 % y que el coeficiente de determinación obtenido es: $R^2 = 0,86$ se puede decir que el ajuste exponencial en el subperíodo es adecuado para modelar el comportamiento de la producción científica y establecer proyecciones.

En la Figura 5.43, con el parámetro α estimado, se tienen las proyecciones para el 2017-2020:

	A	PCP
1	2017	400
2	2018	497
3	2019	617
4	2020	767

Figura 5.43: Proyecciones USFQ

5.3.3 Ley de Lotka. Índice de Productividad Personal.

El índice de productividad personal permite determinar una clasificación de autores, sin embargo, para que ese índice pueda ser utilizado, debe probarse que el comportamiento de la productividad de los autores cumple la Ley de Lotka clásica.

En las Figuras 5.44, 5.45 y 5.46 se presenta las probabilidades de que se cumpla la Ley de Lotka clásica dentro de cada categoría considerando la productividad de los autores, en los años 2010 al 2017*:

5.3.3.1 *Categoría A.

	A	LA
1	UASB	1
2	FLACSO	1
3	EPN	0.454
4	ESPE	0.387
5	ESPOL	0.24
6	UCUENCA	0.082
7	UEES	0.009
8	USFQ	0

A: IES de la categoría A; LA: p-valor.

Figura 5.44: Lotka. Categoría A.

Con base en la Figura 5.44, dentro de la Categoría A, las IES: USFQ y UEES no cumplen con la Ley de Lotka clásica. Una de las posibles causas es que la productividad de sus autores pueda ajustarse a la Ley de Lotka Generalizada y no a la Clásica.

5.3.3.2 *Categoría B.

	B	LB
1	UTB	1
2	UTA	0.978
3	UNL	0.938
4	UDLA	0.847
5	UTPL	0.591
6	UDA	0.541
7	UMACH	0.441
8	UPE	0.4
9	ESPOCH	0.336
10	UG	0.330
11	SEK	0.329
12	UTM	0.329
13	UTN	0.27
14	PUCE	0.213
15	UCSO	0.181
16	UCE	0.113
17	UPS	0.108

B: IES de la categoría B; LB: p-valor.

Figura 5.45: Lotka. Categoría B.

Con base en la Figura 5.45, dentro de la Categoría B, todas las IES cumplen con la Ley de Lotka clásica.

5.3.3.3 *Categoría C.

	C	LC
1	UNIANDES	1
2	UAE	0.819
3	UNACH	0.699
4	ULEAM	0.627
5	UPSE	0.4

C: IES de la categoría C; LC: p-valor.

Figura 5.46: Lotka. Categoría C.

Con base en la Figura 5.46, dentro de la Categoría C, todas las IES cumplen con la Ley de Lotka clásica.

Como podemos observar, la mayoría de las IES ecuatorianas cumplen con la ley de Lotka clásica a un nivel de confianza del 95 % ($p\text{-valor} > 0,05$), por lo cual, para las IES que cumplen esta ley, se puede proceder al cálculo de la productividad personal.

Nota: En la sección 5.3.3.4 se desarrolla el análisis considerando únicamente la EPN. Es importante mencionar que el desarrollo de esta sección puede realizarse para cualquier IES según el propósito del lector (el código de programación en R permite realizar esta particularización).

5.3.3.4 Productividad Personal para la EPN.

En la Figura 5.47, se observa que en los años 2010-2017*, la población de autores fueron en su mayoría productores transitorios (75,48%), seguidos de productores intermedios (23,63%) y finalmente por grandes productores (0,89%):

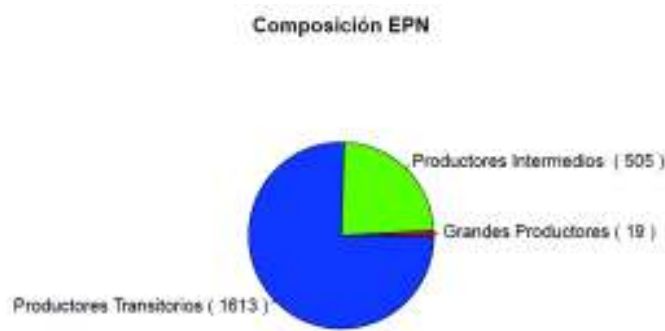


Figura 5.47: Composición Productores EPN.

Cabe mencionar que parte de los productores transitorios son colaboradores extranjeros ocasionales.

En la Figura 5.48 se presenta los 19 grandes productores localizados:

AU	Nro.Documentos	IP	TIPO	
1	MOYHES,P	30	1.4771213	Grandes Productores
2	VEPES,H	20	1.5010300	Grandes Productores
3	RUIZ,M	17	1.2800089	Grandes Productores
4	LUIÁN-MORA,S	16	1.2041200	Grandes Productores
5	RAZ,I	15	1.1760015	Grandes Productores
6	AGUIRACOLA	12	1.0791812	Grandes Productores
7	BAQUERO-ESCUDERO,M	12	1.0791812	Grandes Productores
8	BAGLE,L	12	1.0791812	Grandes Productores
9	TORRES,L	12	1.0791812	Grandes Productores
10	DELOSREYES,J	11	1.0413927	Grandes Productores
11	HIDALGO,S	11	1.0413927	Grandes Productores
12	IDROBOL,I	11	1.0413927	Grandes Productores
13	VALENCA-CHAMBEROS	11	1.0413927	Grandes Productores
14	BEATE,B	10	1.0000000	Grandes Productores
15	DOMÍNGUEZ-CASTRO,P	10	1.0000000	Grandes Productores
16	HALL,M	10	1.0000000	Grandes Productores
17	LEPENEC,I	10	1.0000000	Grandes Productores
18	HOCQUET,I	10	1.0000000	Grandes Productores
19	RAMÓN,P	10	1.0000000	Grandes Productores

AU: Investigador; Nro.artículos: Número de documentos generados por el autor en 2010 – 2017; IP: Índice de Productividad Personal; TIPO: segmentación según el IP.*

Figura 5.48: Grandes Productores EPN.

Es importante mencionar que los autores catalogados como grandes productores tienen como mínimo 10 documentos generados dentro de los años 2010-2017*.

Capítulo 6

Conclusiones y Recomendaciones.

6.1 Conclusiones.

Nota: Se considera como documentos científicos a todos los posibles tipos de presentación de la investigación científica: artículos, libros, conferencias, etc.

Con base en los resultados obtenidos en el capítulo 4, y recordando que los indicadores bibliométricos permiten obtener una visión del estado actual de la producción científica mientras que los modelos de Lotka y Price permiten obtener una visión del estado futuro de la producción científica, se tienen las siguientes conclusiones del presente estudio:

- Al considerar la producción científica en los últimos años se evidencia que las IES (Instituciones de Educación Superior) se están convirtiendo de manera paulatina en los generadores principales de documentos científicos en revistas de importancia internacional. Esto se puede evidenciar al comparar los porcentajes de composición de los documentos generados por IES y las instituciones que no son IES del 2010 y 2016 (Ver 5.2), ya que en el 2010 la diferencia entre los dos tipos de productores fue del 1,54 % y al 2016, la diferencia fue de 57,14 %.
- La representación más explotada de la actividad investigativa ecuatoriana sigue siendo el artículo científico (representación clásica) mientras que la representación menos explotada es la publicación de libros o capítulos de libros.
- Se puede observar que las IES con un mayor número de producción considerando los años 2010-2017* son Universidad San Francisco de Quito (USFQ: 892 documentos), Universidad Técnica Particular de Loja (UTPL: 683 documentos) y Universidad Nacional del Chimborazo (UNACH: 112 documentos) dentro de las categorías A, B y C

respectivamente. Se tiene que la diferencia entre la USFQ y UTPL es de 209 documentos y entre la UTPL y UNACH es de 571 documentos. La universidad pública con mayor número de publicaciones es la Universidad de las Fuerzas Armadas (ESPE: 631 documentos), seguida por la Escuela Superior Politécnica del Litoral (ESPOL: 618 documentos) y la Escuela Politécnica Nacional (EPN: 546 documentos). La diferencia entre la ESPE y la ESPOL son 13 documentos mientras que entre la ESPOL y la EPN se tiene una diferencia de 72 documentos (Ver [A.1](#)).

- Dentro del período de análisis, se observa que el número de documentos generados por un solo autor (afiliado a una sola IES) constituyen el 4,9% de la producción total; también se puede observar una tendencia a generar documentos con 3 autores (sean estos afiliados a una misma IES o diferentes IES) constituyendo el 17,9% de la producción total. Esta composición muestra el comportamiento del fenómeno colaborativo, lo que evidencia que la producción de las IES va a la par con el proceso evolutivo de la Ciencia [[41](#)], ya que se tiene una característica arraigada de generación cooperativa y un declive en la tendencia de generación unipersonal.
- Debido a que en el período de análisis se tiene que solo el 9,5% de la producción tiene más de 11 autores, mostrando casos extremos de documentos con más de 2000 autores, se puede observar que la tendencia cooperativa no se encuentra altamente influenciada por posibles aspectos controversiales al momento de producir en colaboración, como por ejemplo, la autoría honoraria ¹, cuya visualización se da en artículos con un alto número de autores, siendo que los documentos no son un compendio de estudios particulares.
- El fenómeno evidenciado entre la citación y la producción científica (tendencia inversa de comportamiento en los últimos años) puede comprenderse como natural dentro del progreso de la Ciencia debido a que las publicaciones más recientes no han tenido el mismo tiempo u oportunidad de ser citado en comparación con publicaciones de fechas previas [[42](#)].
- La citación dentro del período de análisis puede considerarse a simple vista (tomando en cuenta cantidades netas) como un fenómeno alentador, debido a que en cada año, las citas superan a los valores netos de producción (Ver [5.9](#)). Sin embargo, no se recomienda medir el impacto de esta manera ya que distintos tipos de áreas y enfoques

¹**Autoría Honoraria:** Práctica de nombrar como autores de un documento científico a aquellas personas cuya participación no ha sido de tipo investigativa generadora de conocimiento, por ejemplo, proveedores de fondos, materiales etc., inflando de esta manera el grado de autoría.

de investigación tienen distintas prácticas de citación. Debido a que en el presente estudio se consideró a todas las IES ecuatorianas, y estas tienen diferentes enfoques de vinculación con la sociedad, no es posible concluir que el impacto de la producción sea equilibrado (en el sentido de que todos los artículos causan impacto en la comunidad científica); de hecho, dentro del estudio se evidencian artículos con más de 500 citas, los cuales podrían considerarse como de mayor impacto, mientras que los 1.938 documentos con 0 citas como de impacto nulo.

- Se puede observar que las IES con un mayor número de citación en los años 2010-2017* son Universidad San Francisco de Quito (USFQ: 9.649 citas), Pontificia Universidad Católica del Ecuador (PUCE: 4.854 citas) y Universidad Nacional del Chimborazo (UNACH: 287 citas) dentro de las categorías A, B y C respectivamente. La universidad pública con mayor número de citación es la Universidad de Cuenca (UCUENCA: 2.833 citas) seguida por la Escuela Politécnica Nacional (EPN: 1.988 citas) con una diferencia entre las mismas de 845 citas (Ver [A.3](#)).
- Se observa que la producción científica dentro del período de análisis tiene un alto componente internacional (el 76,43 % de la producción total tiene al menos una afiliación extranjera), lo que evidencia que publicar en colaboración con investigadores extranjeros sigue siendo la preferencia de las IES; sin embargo, las IES también se encuentran fomentando la producción netamente ecuatoriana ya sea en documentos generados de manera individual o de manera conjunta con otras IES (Ver [5.15](#)). Este escenario puede considerarse adecuado para fines futuros de producción, ya que se evidencia el interés de promover la producción únicamente ecuatoriana conservando a la vez el interés por establecer relaciones internacionales y compartir conocimientos.
- Considerando los años 2010-2017*, las IES mejor posicionadas considerando el H-index son: Universidad San Francisco de Quito (USFQ), Pontificia Universidad Católica del Ecuador (PUCE), Universidad de Cuenca (UCUENCA) y Escuela Politécnica Nacional (EPN), siendo USFQ, UCUENCA y EPN instituciones pertenecientes a la Categoría A y la PUCE institución perteneciente a la Categoría B (Ver [5.18](#)).
- Al observar el comportamiento dentro del período de análisis de las IES pertenecientes a la Categoría A, se observa que la Universidad San Francisco de Quito (USFQ) ha mantenido la primera posición dentro de la categoría con un H-index mayor y una notoria diferencia al compararse con las demás IES, mientras que la Universidad de

Cuenca (UCUENCA) y la Escuela Politécnica Nacional (EPN) han compartido de manera alternante la segunda posición debido a la cercanía de sus valores en H-index (Ver 5.20). Al utilizar el H-index como medida de Impacto, se observa que estas 3 IES constituyen las instituciones de mayor impacto.

- Para la Categoría B, la Pontificia Universidad Católica del Ecuador (PUCE) ha mantenido la primera posición dentro de la categoría con un H-index mayor y una notoria diferencia al compararse con las demás IES (Ver 5.23).
- Al ser el H-index una medida de volumen y calidad, se puede decir que la Pontificia Universidad Católica del Ecuador (PUCE) sobresale en su categoría ya que considerando los años 2016-2017* tiene un H-index mayor al de 2 instituciones de Categoría A: Universidad de Cuenca (UCUENCA) y Escuela Politécnica Nacional (EPN) (Ver 5.18).
- Debido a que se verifica que la producción científica dentro del período de análisis cumple con la Ley de Price, se evidencia que el sistema de educación superior ecuatoriano se halla en un momento favorable de inserción en la comunidad investigativa internacional que debe ser aprovechado. Este crecimiento podría estar altamente correlacionado con las medidas tomadas desde la creación del CEAACES y el proceso de categorización. Utilizar el modelo de Price para establecer proyecciones, permite tener una idea general del futuro comportamiento de las IES, sin embargo, cabe recordar que este modelo es aplicable si los recursos operativos y financieros invertidos se mantienen o se incrementan. Debido a que la Ciencia no puede crecer infinitamente, llegará el momento en que el crecimiento llegue a un punto de saturación, generalmente generado por la falta o disminución de recursos.
- Al observar que el modelo de Price ajusta de manera satisfactoria al crecimiento de la producción científica de las IES (debido a que satisface la prueba de prueba de bondad de ajuste de Kolmogórov-Smirnov y el R^2 obtenido es de 0,96), se puede asegurar que el tiempo de duplicación de la producción es de prácticamente 2 años; es decir que, para el año 2020 la producción científica tendrá un valor de aproximadamente el doble de la existente en el año 2018. De hecho, usando la proporción de crecimiento obtenida en el ajuste del modelo exponencial de Price: $a=0,3346$, se observa que en el año 2020, se tendrá un Número Total de Publicaciones (PC) de 6.558 documentos siendo que el año 2018 el PC se estima será de 3.358 documentos (Ver 5.34).
- Al observar los ajustes realizados dentro de la Categoría A, se tiene que la Univer-

sidad San Francisco de Quito (USFQ) y la Universidad de Especialidades Espíritu Santo (UEES) no satisfacen el modelo de Price considerando el período 2010-2016. En casos como estos, con base en la teoría Bibliométrica, se puede probar ajustes paramétricos de tipo lineal. El uso de este modelo alternativo tiene una relación directa con el momento en el que se encuentre el entorno de producción científica de la IES. En el presente trabajo, se tomó una alternativa lineal satisfactoria para la UEES (debido a que satisface la prueba de prueba de bondad de ajuste de Kolmogórov-Smirnov y el R^2 obtenido es de 0,96). Por otro lado, para la USFQ, un modelo de tipo lineal no fue adecuado para el ajuste, frente a lo cual, se consideró un modelo exponencial con la modificación del período de análisis lo cual es válido ya que en los recientes años: 2013-2016, el comportamiento de la USFQ no ha tenido irregularidades (Ver 5.38), dando como resultado un ajuste satisfactorio en la prueba de Kolmogórov-Smirnov y un R^2 de 0,86 .

- Debido a que la Ley de Lotka se verifica para todas las IES del estudio con excepción de la Universidad San Francisco de Quito (USFQ) y la Universidad de Especialidades Espíritu Santo (UEES), (se verifica la prueba de Kolmogórov-Smirnov a un nivel de confianza del 95 %), es posible calcular el índice de productividad personal para los investigadores a nivel interno de cada IES que cumple la Ley de Lotka Clásica. El índice de productividad personal muestra como pocos investigadores afiliados a una institución en particular, son los generadores de la mayor parte de documentos de dicha IES, recordando que dichos productores pueden o no estar asociados a la misma área investigativa. Ante esto, es importante que las medidas establecidas por las IES para fomentar un incremento en la productividad personal, se diferencien de acuerdo a la área de investigación.
- Debido a que la publicación científica se ha convertido en una prioridad para la comunidad investigativa, existen aspectos que pueden afectar los niveles de publicación para las IES, en especial, si el documento que se desea publicar en una revista indexada pertenece a disciplinas con un número bajo de científicos incursionando en las mismas.

Algunos de los aspectos que determinan el nivel de publicación por área se encuentran: el tiempo de demora desde el envío hasta la publicación debido al incremento en la oferta de documentos científicos, el tipo de idioma de la publicación, y la asignación de los pares revisores. Existen áreas donde la producción de documentos es alta, como por ejemplo Medicina (Ver 5.30), provocando que a pesar de que hay documentos

que se encuentran en revisión para publicación, hay los suficientes documentos ya publicados como para mencionar que se conserva el interés en dicha área. El tipo de idioma también es un factor importante al momento de publicar ya que de acuerdo a la revista objetivo, el idioma puede variar; en el caso del presente estudio, existen revistas donde se puede estar publicando y que por algún filtro, no se encuentre en la base indexada de Scopus. Finalmente, la asignación en la revisión por pares, que constituye la validación del documento presentado, puede verse afectada por el grado de especialización de los científicos en una determinada área como ocurre con disciplinas asociadas a la Matemática (Álgebra, Optimización, etc.).

6.2 Recomendaciones.

Las recomendaciones del presente estudio se encuentran enfocadas en dos aspectos importantes: **fFuente de datos** y **futuros estudios**.

6.2.1 Fuente de Datos: Situaciones Observadas a la fecha de corte.

A lo largo del estudio, se evidenciaron situaciones que afectan el cálculo de los indicadores. Se menciona estas observaciones con el objetivo de que antes de desarrollar futuros estudios bibliométricos, se establezca una medida correctiva sobre las mismas.

- **Existencia de artículos que Scopus no considera como producción ecuatoriana al momento de realizar la consulta por afiliación Ecuador, pero si se evidencian en la consulta por IES.**

Como se pudo observar en el capítulo 3, para el cálculo de uno de los indicadores bibliométricos del estudio, se creó una base auxiliar que contiene los documentos asociados al criterio AFFILCOUNTRY(ECUADOR) en los años 2010-2016; este criterio filtra dentro de las bases de Scopus, todos los documentos que posean al menos una afiliación ecuatoriana:



Figura 6.1: Detalle criterio Scopus.

Es importante mencionar que esta base auxiliar es de una dimensión superior a la base generada por los documentos de las IES debido a que Scopus reconoce como afiliación a toda institución que genera producción científica, es decir, considera a las IES, y también a hospitales, centros de investigación, entidades privadas, etc.

Al realizar la comparación entre la base E_TOT y la base B_IES_C2* (considerando B_IES_C2 en los años 2010-2016), se evidenció que algunos documentos de B_IES_C2*

no se encontraban en la base E_TOT.

En las Figuras 6.2 y 6.3 se muestra un ejemplo:

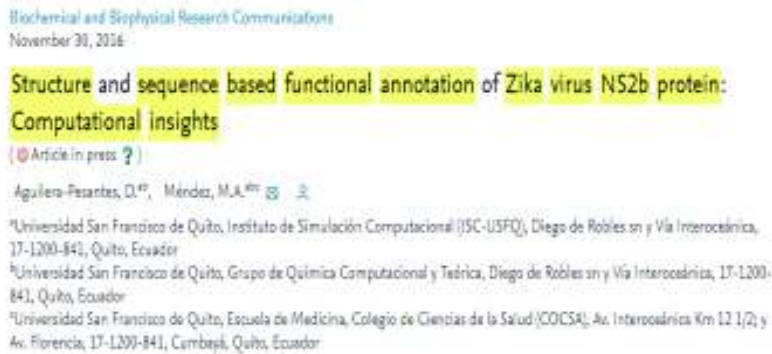


Figura 6.2: Documento en la base B_IES_C2*.



Figura 6.3: Consulta en la base E_TOT.

El ejemplo mencionado, nos permite establecer que al momento de realizar un estudio a nivel de Sistema de Educación Superior, el esquema seguido en el capítulo 3, da como resultado una base fiable para los análisis.

- **Existen artículos que Scopus no considera como producción ecuatoriana, y que tampoco aparecen en la consulta de las IES a pesar de que su autor pertenece a una IES.**

Como se pudo observar en el capítulo 3, Scopus establece afiliaciones potenciales, hecho que permite obtener una base más acorde a la producción científica real de las IES; sin embargo, existen ocasiones en las cuales, a pesar de realizar este refinamiento, los documentos no se asocian a la IES, mermando así el número de artículos considerados para el estudio.

En las Figuras 6.4, 6.5 y 6.6 se muestra un ejemplo:



Figura 6.4: Documento asociado a un autor de la EPN.



Figura 6.5: Consulta del documento en la IES.



Figura 6.6: Consulta en la base E_TOT.

Esta situación ocurre por que Scopus no logra establecer una conexión entre la forma de escritura de la afiliación y la IES, y por ende no lo considera como un artículo de la EPN. También se evidencia que no existe concordancia entre la escritura de la afiliación y el país provocando que tampoco se reconozca al documento como artículo ecuatoriano.

Frente a estas dos situaciones se recomienda las siguientes **medidas de corrección**:

- Comunicarse con Scopus para redefinir la afiliación y el país de afiliación de los documentos; para lo cual, los autores podrían verificar en Scopus la forma de registro de sus documentos y en caso de existir alguna incongruencia pedir la corrección del mismo.
- Establecer un patrón, o formato de escritura de afiliación normalizado e incluso automatizado dentro de las IES para futuros documentos.

6.2.2 Futuros estudios.

- Para futuros estudios y de acuerdo a las condiciones propias del sistema de educación superior (cambio de normativas, distribución de fondos, etc), se recomienda probar un ajuste logístico para la producción científica nacional, debido a que este modelo representa comportamientos con un nivel de saturación.
- En futuros estudios, se recomienda probar modelos de tipo no paramétricos para aquellos casos en los que no se cumpla la Ley de Price. Estos modelos constituirían una alternativa a la opción de ajuste lineal y podrían tener un mejor desempeño al momento de representar el comportamiento de la producción científica.
- Se recomienda incursionar en mecanismos de automatización de estudios bibliométricos con el objetivo de tener una visualización al día del comportamiento de la investigación científica.
- Se recomienda crear un sistema de información basado en Bibliometría que permita obtener datos, procesarlos y generar conclusiones sobre el comportamiento de la producción, citación, colaboración, etc.
- El presente estudio utiliza datos de Scopus, en un futuro se podría tener una extensión al considerarse otras bases de datos internacionales tales como ISI Web of Knowledge; para esto, es importante establecer una normalización entre los datos ya que su forma de acceso, extracción y escritura es diferente.

Apéndice A

Anexos

A.1 Producción Científica IES por año .

	IES	2010	2011	2012	2013	2014	2015	2016	2017	2010-2016	2010-2017*
1	USFQ	54	83	106	88	93	117	182	169	723	892
2	UTPL	32	24	37	44	87	135	232	92	591	683
3	ESPE	6	7	17	22	56	148	246	129	502	631
4	ESPOL	10	27	32	40	83	101	198	121	497	618
5	EPN	22	27	32	41	58	95	162	109	437	546
6	PUCE	50	41	54	70	59	79	96	76	453	529
7	UCUENCA	16	16	19	30	65	97	144	68	387	455
8	UPS	1	2	13	13	31	103	142	55	305	360
9	UCE	13	22	23	24	44	57	91	47	276	323
10	UDLA	7	3	6	9	27	40	52	50	144	194
11	UEES	0	1	8	25	36	28	46	40	144	184
12	ESPOCH	1	2	5	10	25	44	41	28	128	156
13	YACHAY	0	0	0	0	10	47	45	54	102	156
14	UCSG	9	23	21	11	15	17	33	20	129	149
15	UG	6	5	10	8	9	28	38	29	104	133
16	UTA	0	1	3	2	14	22	61	20	103	123
17	UNACH	0	0	0	2	16	29	34	31	81	112
18	UNL	1	7	2	6	5	14	40	17	75	92
19	UTE	1	3	8	6	4	9	18	33	49	82
20	ULEAM	0	0	0	3	12	27	26	13	68	81
21	UTMACH	1	3	0	2	7	17	26	16	56	72
22	UDA	0	3	2	4	6	13	14	22	42	64
23	UTN	0	0	1	2	7	4	24	23	38	61
24	UPSE	0	0	2	4	5	10	11	10	32	42
25	UTM	0	0	0	2	0	7	14	17	23	40
26	SEK	0	0	4	2	3	2	6	11	17	28
27	UAE	0	0	1	1	2	1	13	5	18	23
28	UASB	0	1	1	0	0	0	0	0	2	2
29	FLACSO	0	1	0	0	0	0	0	0	1	1
30	IAEN	0	0	0	1	0	0	0	0	1	1
31	UIDE	0	0	1	0	0	0	0	0	1	1
32	UNIANDÉS	1	0	0	0	0	0	0	0	1	1
33	UTI	0	0	1	0	0	0	0	0	1	1

Figura A.1: Producción Científica IES por Año: 2010-2017*.

A.2 Producción según el tipo de documento por año.

A	ARTÍCULO	LIBRO	CONFERENCIA	Total
2010	194	4	33	231
2011	264	8	21	293
2012	328	3	52	383
2013	393	9	52	454
2014	589	16	115	720
2015	885	27	287	1199
2016	1371	11	451	1833
2017*	1029	4	147	1180

Figura A.2: Producción según el tipo de documento por año: 2010-2017*.

A.3 Citas por IES.

	IES	2010	2011	2012	2013	2014	2015	2016	2017*	2010-2016	2010-2017*
1	USFQ	1888	2125	2347	1315	869	501	476	128	9521	9649
2	PUCE	1550	586	777	872	522	313	218	16	4838	4854
3	UCUENCA	332	170	190	895	405	321	483	37	2796	2833
4	UTPL	170	152	711	303	491	361	216	9	2404	2413
5	EPN	339	291	243	289	357	296	162	11	1977	1988
6	ESPOL	49	235	333	220	328	242	168	20	1575	1595
8	ESPE	58	123	377	79	256	329	219	25	1441	1466
7	UCE	104	100	171	425	305	237	115	7	1457	1464
9	UCSG	164	252	400	90	91	44	23	1	1064	1065
10	UEES		0	12	197	274	84	44	5	611	616
11	UDLA	101	28	11	64	151	85	73	15	513	528
12	UDA		11	2	12	8	236	211	1	480	481
13	YACHAY					38	272	120	13	430	443
14	UG	150	38	93	45	21	35	15	3	397	400
15	UPS	0	7	15	22	65	156	106	2	371	373
16	UTN			0	228	91	16	10	1	345	346
17	ESPOCH	3	11	81	64	44	49	35	18	287	305
18	UNACH				17	61	164	36	9	278	287
19	UTE	4	18	80	122	20	21	8	3	273	276
20	UTMACH	27	102		26	75	15	28	1	273	274
21	UTA		0	24	15	116	57	26	3	238	241
22	UNL	19	47	6	22	20	71	37	0	222	222
23	ULEAM				7	39	110	29	2	185	187
24	UPSE			1	8	10	26	5	1	50	51
26	SEK			13	7	0	3	9	3	32	35
25	UTM				20		12	2	0	34	34
27	FLACSO		16							16	16
28	UAE			0	9	6	0	1	0	16	16
29	UTI			3						3	3
30	UASB		1	0						1	1
31	IAEN				0					0	0
32	UIDE			0						0	0
33	UNIANDES	0								0	0

Figura A.3: Citación por IES y año: 2010-2017*.

A.4 Citas y Frecuencia de documentos.

	x	y	x	y	x	y	x	y	x	y	x	y	x	y
1	0	1938	21	20	18	41	40	1	61	63	3	81	96	1
2	1	775	22	21	21	42	41	5	62	64	2	82	98	1
3	2	460	23	22	17	43	42	1	63	66	1	83	109	1
4	3	315	24	23	13	44	43	4	64	68	1	84	111	1
5	4	222	25	24	19	45	44	3	65	70	1	85	112	1
6	5	165	26	25	12	46	45	3	66	71	3	86	113	1
7	6	147	27	26	18	47	46	3	67	72	1	87	115	1
8	7	141	28	27	10	48	47	2	68	75	1	88	120	1
9	8	86	29	28	9	49	48	2	69	77	2	89	124	2
10	9	91	30	29	5	50	49	3	70	79	1	90	125	1
11	10	63	31	30	9	51	50	1	71	81	1	91	132	1
12	11	81	32	31	7	52	51	1	72	82	2	92	134	1
13	12	48	33	32	9	53	52	2	73	83	2	93	140	1
14	13	50	34	33	15	54	53	3	74	85	1	94	144	1
15	14	41	35	34	9	55	54	3	75	87	1	95	146	1
16	15	30	36	35	4	56	55	4	76	88	1	96	154	1
17	16	45	37	36	9	57	57	1	77	89	1	97	155	1
18	17	26	38	37	8	58	58	2	78	91	1	98	162	1
19	18	31	39	38	6	59	59	1	79	92	1	99	169	1
20	19	30	40	39	3	60	62	1	80	94	1	100	171	1

Figura A.4: Citas y Frecuencia de documentos.

A.5 Programación en R.

```

library(bibliometrix)
library(dplyr)
library(sqldf)
library(stats)
library(ggplot2)
library(reshape2)
library(CITAN)
##### CONSTRUCCIÓN DE BASES #####

#### CARGA DE DATOS: BASE INICIAL ####
#Construcción Base: B_IES_C1 (MCONDU)
uni<-c("EPN", "ESPE", "ESPOCH", "ESPOL", "PUCE",
       "SEK", "UAE", "UASB", "UCE", "UCSG", "UCUENCA", "UDA",
       "UDLA", "UEES", "UG", "ULEAM", "UNL",
       "UPS", "UPSE", "USFQ", "UTA", "UTE", "UTM", "UTMACH",
       "UTN", "UTPL", "UNACH", "YACHAY", "FLACSO", "IAEN", "UIDE", "UNIANDES", "UTI
       ")

i <- 1
nombre <- paste0(uni[i], ".bib")

```

```

nombre
D <- readFiles(nombre)
M <- scopus2df(D)
x<-dim(M)
print(x)
M$IES<-as.character(uni[i])
for(i in 2:length(uni)){
  nombre <- paste0(uni[i], ".bib")
  D <- readFiles(nombre)
  N <- scopus2df(D)
  x<-dim(N)
  print(x)
  N$IES<-as.character(uni[i])
  M <- mergeDbSources(M,N, remove.duplicated=FALSE)
}
dim(M)
M<-distinct(M)##Retiro los auxiliares (IES con 1 doc)
names(M)
dim(M)
View(M)
save(M, file = "MCONDU.RData") ##load("MCONDU.RData")

#### PROCEDIMIENTO DE DEPURACIÓN ####
## TEMPORALIDAD 2010-2017 ##
#Construcción Base: B_IES_C2 (M1VT)
M1<- M %>% select(AU, TI, SO, DT, C1, TC, PY, UT, IES) %>% filter(PY>=2010 & PY
  <=2017)
dim(M1)
names(M1)
View(M1)
save(M1, file = "M1VT.RData")##load("M1VT.RData")
## CORRECCIÓN DATOS ##
#Construcción Base: B_IES (M1F)
# 1) Verificación de la presencia de "Ecuador"
pru1<-select(M1, TI, C1, IES)
pru1$UBIC<-strapply(pru1$C1, "ECUADOR")
View(pru1)
pru2<-filter(pru1, UBIC %in% c("NULL"))
registros_sin_ECUADOR<-dim(pru2)[1]

```



```
registros_sin_ECUADOR
# 2) Realizando la normalización Ciudad-País
patron <-"$"
M1<-apply(M1, 2, function(M1) gsub(patron,";",M1,ignore.case=T))
patron <-"Q(U|Ü|Ú) (I|Í|Ĭ)T(O|Ó|Ö),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"QUITO, ECUADOR;",M1,ignore.case=
T))
patron <-", (U|Ü) (I|Í)T(O|Ó),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron," , QUITO, ECUADOR;",M1,ignore.
case=T))
patron <-"ESCUELA POLITÉCNICA NACIONAL,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"ESCUELA POLITÉCNICA NACIONAL,
QUITO, ECUADOR;",M1,ignore.case=T))
patron <-"ESPE,\\s\\w+(;|\\s+)"
M1<-apply(M1, 2, function(M1) gsub(patron,"ESPE, SANGOLQUÍ, ECUADOR;",M1,
ignore.case=T))
patron <-"PICHINCHA,\\s\\w+\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"SANGOLQUÍ, ECUADOR;",M1,ignore.
case=T))
patron <-"ARMADAS,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"ARMADAS, SANGOLQUÍ, ECUADOR;",M1
,ignore.case=T))
patron <-"ESPE,\\w+,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"ESPE, SANGOLQUÍ, ECUADOR;",M1,
ignore.case=T))
patron <-"R(I|Í|Ĭ) (O|Ó|Ö)B(A|Ä|Á)MB(A|Ä|Á),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"RIOBAMBA, ECUADOR;",M1,ignore.
case=T))
patron <-"LITORAL,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"LITORAL, GUAYAQUIL, ECUADOR;",M1
,ignore.case=T))
patron <-"LITORAL,\\s\\w+\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"LITORAL, GUAYAQUIL, ECUADOR;",M1
,ignore.case=T))
patron <-"G(U|Ü|Ú) (A|Ä|Á) (Y|LL) (A|Ä|Á) (Q|K|C) (U|Ü|Ú) (I|Í|Ĭ) (L|LL),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"GUAYAQUIL, ECUADOR;",M1,ignore.
case=T))
patron <-"SEK,\\s\\w+;"
```

```

M1<-apply(M1, 2, function(M1) gsub(patron,"SEK, QUITO, ECUADOR;",M1,ignore.
  case=T))
patron <-"C(U|Ü|Ú)(E|É|Ë)NC(A|Ä|Á),\\s\\w+,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"CUENCA, ECUADOR;",M1,ignore.case
  =T))
patron<-"M(A|Ä|Á)N(A|Ä|Á)B(I|Í|Ï),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"MANABÍ, ECUADOR;",M1,ignore.case
  =T))
patron <-"GUANO\\sRE,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"RIOBAMBA, ECUADOR;",M1,ignore.
  case=T))
patron <-"SALESIANA,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"SALESIANA, QUITO, ECUADOR;",M1,
  ignore.case=T))
patron <-"(A|Ä|Á)MB(A|Ä|Á)T(O|Ó|Ö),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"AMBATO, ECUADOR;",M1,ignore.case
  =T))
patron <-"M(A|Ä|Á)CH(A|Ä|Á)L(A|Ä|Á),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"MACHALA, ECUADOR;",M1,ignore.
  case=T))
patron <-"(I|Í|Ï)B(A|Ä|Á)RR(A|Ä|Á),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"IBARRA, ECUADOR;",M1,ignore.case
  =T))
patron <-"([UTPL]),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,", LOJA, ECUADOR;",M1,ignore.case
  =T))
patron <-", UTPL,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,", LOJA, ECUADOR;",M1,ignore.case
  =T))
patron <-",\\sL(O|Ó|Ö)J(A|Ä|Á),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,", LOJA, ECUADOR;",M1,ignore.case
  =T))
patron <-"DE\\sLOJA,\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"DE LOJA, ECUADOR;",M1,ignore.
  case=T))
patron <-"(U|Ü|Ú)RC(U|Ü|Ú)Q(U|Ü|Ú)(I|Í|Ï);"
M1<-apply(M1, 2, function(M1) gsub(patron,"URCUQUÍ, ECUADOR;",M1,ignore.
  case=T))
patron<-"M(A|Ä|Á)NT(A|Ä|Á),\\s\\w+;"

```

```

M1<-apply(M1, 2, function(M1) gsub(patron,"MANTA, ECUADOR;",M1,ignore.case=
  T))
patron <-"S(A|Ä|Á)NT(A|Ä|Á)\\s(E|É|Ë)L(E|É|Ë)N(A|Ä|Á),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"SANTA ELENA, ECUADOR;",M1,ignore
  .case=T))
patron <-"P(O|Ó|Ö)RT(O|Ó|Ö)V(I|Í|Ï)(E|É|Ë)J(O|Ó|Ö),\\s\\w+;"
M1<-apply(M1, 2, function(M1) gsub(patron,"PORTOVIEJO, ECUADOR;",M1,ignore.
  case=T))
patron<-";;"
M1<-apply(M1, 2, function(M1) gsub(patron,";",M1,ignore.case=T))
patron<-";;,"
M1<-apply(M1, 2, function(M1) gsub(patron,";",M1,ignore.case=T))
patron <-";$"
M1<-apply(M1, 2, function(M1) gsub(patron,"",M1,ignore.case=T))
M1<-as.data.frame(M1)
M1$C1<-as.character(M1$C1)
View(M1)
#3) Verificación corrección
pru3<-select(M1, TI, C1, IES)
pru3$UBIC<-strapply(pru3$C1, "ECUADOR")
pru4<-filter(pru3, UBIC %in% c("NULL"))
registros_sin_ECUADOR_F<-dim(pru4)[1]
registros_sin_ECUADOR_F
#####
M1F<-M1
View(M1F)
dim(M1F)
save(M1F, file = "M1F.RData")##load("M1F.RData")
## ELIMINACIÓN DATOS DUPLICADOS ##
#Construcción Base: B_T_C1 (M2VT)
M21<- M1F %>% select(AU, TI, SO, DT, C1, TC, PY, UT)
dim(M21)
M2<-distinct(M21)## retira registros duplicados exactos entre las IES.
dim(M2)
View(M2)
save(M2, file = "M2VT.RData") ###load("M2VT.RData")
## CONSOLIDACIÓN DE CITACIONES ##
#Construcción Base: B_T_C2 (M2F)

```

```

codigo <- paste(M2$AU, "-", M2$TI) ## consolida citas para un mismo
  artículo
M2$codigo <- codigo
View(M2)
rd<-sqldf("select codigo, count(codigo) n, max(TC) maxi, min(TC) mini, sum(
  TC) suma from M2 group by codigo")
View(rd)
dim(rd)
save(rd, file = "rd.RData") ###load("rd.RData")
M2F<-merge(x = M2 , y = rd, all.x = TRUE)
M2F$PY<-as.character(M2F$PY)
dim(M2F)
View(M2F)
save(M2F, file = "M2F.RData") #####load("M2F.RData")
## BASE DEPURADA ##
#Construcción Base: B_T (MC)
MC<- M2F %>% select (codigo,AU, TI, SO,DT,C1,PY,UT,n,maxi,mini,suma)
dim(MC)
View(MC)
save(MC, file = "MC.RData") #####load("MC.RData")

### Se crea una base auxiliar asociada tomando como consulta AFILIACIÓN:
  ECUADOR ###
#Construcción Base: E_TOT
per<-c("E_2010", "E_2011", "E_2012", "E_2013", "E_2014", "E_2015", "E_2016_1", "E_
  2016_2")
i <- 1
nb <- paste0(per[i], ".bib")
nb
D <- readFiles(nb)
E <- scopus2df(D)
x<-dim(E)
print(x)
E$IES<-as.character(per[i])
for(i in 2:length(per)){
  nb<- paste0(per[i], ".bib")
  D <- readFiles(nb)
  N <- scopus2df(D)
  N$A<-as.character(per[i])
}

```

```

E <- mergeDbSources(E,N, remove.duplicated=FALSE)
}
dim(E)
View(E)
save(E, file = "E_1.RData") ##load("E_1.RData")
EV2<- E %>% select (AU, TI, SO, DT, C1, TC, PY, UT)
dim(EV2)
E2<-distinct (EV2) ## retira registros duplicados
dim(E2)
View(E2)
save(E2, file = "E_TOT.RData") ###load("E_TOT.RData")

#####INDICADORES BIBLIOMÉTRICOS #####

#####Indicadores de Producción
###PRODUCCIÓN POR AÑOS
#load("MC.RData")
#View(MC)
PC<-as.data.frame(table(MC$PY))
names(PC)<-c("A", "Nro.artículos")
View(PC)
PCT<-sum(PC$Nro.artículos)
PCT## 2010-2017
PCP<-sum(PC$Nro.artículos[1:7])
PCP## 2010-2016
TCP<-((PC$Nro.artículos[7]/PC$Nro.artículos[1])^(1/(7-1)))-1## 2010-2016
TCP<-round(TCP*100,2)
TCP
#
names(PC)<-c("A", "PC")
View(PC)
###CONCENTRACIÓN POR AÑO
FT1<-PC[1:7,]
FT1$"%Conc."<-round(FT1$PC*100/sum(FT1$PC),2)
View(FT1)
#
gr<-ggplot(data=FT1,aes(x=A,y=PC, fill=PC)) +
  geom_bar(stat="identity", position=position_dodge())
gr + scale_y_continuous(limit = c(0,2000))

```

```

gr
###Ecuador, IES, NoIES
#load("E_TOT.RData")
TY<-as.data.frame(table(E2$PY))
TY
C1<-as.data.frame(cbind(FT1,TY))
C1<-select(C1,A,Nro.articulos_E=Freq,Nro.articulos_IES=PC)
C1$Nro.articulos_NoIES<- C1$Nro.articulos_E-C1$Nro.articulos_IES
C1$"%Resp_E_IES"<-round((C1$Nro.articulos_IES*100)/C1$Nro.articulos_E,2)
C1$"%Resp_E_NoIES"<-round((C1$Nro.articulos_NoIES*100)/C1$Nro.articulos_E
,2)
View(C1)
#
names(C1)<-c("A", "PC.E", "PC.I", "PC.NI", "%.E.I", "%.E.NI")
#
test_data_long <- melt(C1[,1:4], id="A")
names(test_data_long)<-c("A","Variable","Total")
gr1<-ggplot(data=test_data_long,
            aes(x=A, y=Total, colour=Variable,group=Variable)) +
  geom_line()+geom_point()
gr1
### Documentos y Autores
##load("MC.RData")
##View(MC)
AA<-MC %>% select(AU, TI, PY, C1, n, suma)
AA<-filter(AA, PY!="2017")
aut<- as.character(AA$AU)
resaf<- strsplit(aut, ";")
AA$N_AU<-as.numeric(summary(resaf)[,1])
View(AA)
Q<-as.data.frame(AA %>% group_by (N_AU) %>%
                summarise (tot=n()))
Q<-arrange(Q, desc(tot))
names(Q)<-c("Nro.autores", "Total")
Q$Doc.acu<-cumsum(Q$Total)
Q$Porc.doc<-round((Q$Total*100)/max(Q$Doc.acu),1)
Q$Porc.acum.doc<-round((Q$Doc.acu*100)/max(Q$Doc.acu),1)
View(Q)
###Composición documentos

```



```

HA<-sqldf("select *from vA A left join vl B on A.IES=B.IES")
HA<-as.data.frame(HA)
names(HA)<-c("A","IES","N.Doc","IES","TOT")
gr3 <- ggplot(HA, aes(x=reorder(IES,-TOT), y=N.Doc,fill= N.Doc)) +
  geom_bar(stat="identity",position="dodge") + facet_grid(A ~.)+
  labs(x = "IES",y = "Nro.Documentos")
gr3
vl1<- as.data.frame(vB %>% group_by (IES) %>%
  summarise (TOT=sum(Nro.artículos)))
HA1<-sqldf("select *from vB A left join vl1 B on A.IES=B.IES")
HA1<-as.data.frame(HA1)
names(HA1)<-c("A","IES","N.AR","IES","TOT")
gr4 <- ggplot(HA1, aes(x=reorder(IES,-TOT), y=N.AR,fill= N.AR)) +
  geom_bar(stat="identity",position="dodge") + facet_grid(A ~.) +
  labs(x = "IES",y = "Nro.Artículos")
gr4
##Categoría B
uniB<-c("ESPOCH","PUCE","UCSG","UCE","UDA","UNL","SEK","UPS",
  "UTA","UTN","UTPL","UTE","UTM","UDLA",
  "UG","UTMACH")
uniB1<-c("UIDE","UTI","IAEN") ## IES con un solo documento
vA<-filter(PC_I, IES %in% uniB)
vB<-filter(PC_I, IES %in% uniB1)
vl<- as.data.frame(vA %>% group_by (IES) %>%
  summarise (TOT=sum(Nro.artículos)))
HA<-sqldf("select *from vA A left join vl B on A.IES=B.IES")
HA<-as.data.frame(HA)
names(HA)<-c("A","IES","N.Doc","IES","TOT")
gr5 <- ggplot(HA, aes(x=reorder(IES,-TOT), y=N.Doc,fill= N.Doc)) +
  geom_bar(stat="identity",position="dodge") + facet_grid(A ~.)+
  labs(x = "IES",y = "Nro.Documentos")
gr5
vl1<- as.data.frame(vB %>% group_by (IES) %>%
  summarise (TOT=sum(Nro.artículos)))
HA1<-sqldf("select *from vB A left join vl1 B on A.IES=B.IES")
HA1<-as.data.frame(HA1)
names(HA1)<-c("A","IES","N.AR","IES","TOT")
gr6 <- ggplot(HA1, aes(x=reorder(IES,-TOT), y=N.AR,fill= N.AR)) +
  geom_bar(stat="identity",position="dodge") + facet_grid(A ~.)+

```



```

labs(x = "IES",y = "Nro.Artículos")
gr6
##Categoría C
uniC<-c("UNACH", "UPSE", "ULEAM", "UAE", "UNIANDES")
# UNIANDES con un solo documento
vA<-filter(PC_I, IES %in% uniC)
v1<- as.data.frame(vA %>% group_by (IES) %>%
                  summarise (TOT=sum(Nro.artículos)))
HA<-sqldf("select *from vA A left join v1 B on A.IES=B.IES")
HA<-as.data.frame(HA)
names(HA)<-c("A", "IES", "N.Doc", "IES", "TOT")
gr7 <- ggplot(HA, aes(x=reorder(IES, -TOT), y=N.Doc, fill= N.Doc)) +
  geom_bar(stat="identity", position="dodge") + facet_grid(A ~.)+
  labs(x = "IES",y = "Nro.Documentos")
gr7

#####Indicadores de Visibilidad e Impacto
###Número Total de Citas.
##Por año
#load("M2F.RData")
CPA<- M2F %>% select (A=PY, suma, codigo)
View(CPA)
TCI<- as.data.frame(CPA %>% group_by (A) %>%
                  summarise (Nro.Citas=sum(as.numeric(suma)),
                              Nro.artículos=n()))
TC<-sum(TCI$Nro.Citas[1:7])
TC
TCI$Cit.por.art<-round(TCI$Nro.Citas/TCI$Nro.artículos,1)
View(TCI)
#
names(TCI)<-c("A", "CT", "PC", "C.D")
View(TCI)
#
test_data_long <- melt(TCI[1:7,1:3], id="A")
names(test_data_long)<-c("A", "Variable", "Total")
gr8<-ggplot(data=test_data_long,
            aes(x=A, y=Total, colour=Variable, group=Variable)) +
  geom_line()+geom_point()

```

```

gr8
###Citaciones por IES.
View(MC)
View(M1F)
A$codig<-paste(A$AU, "-", A$TI, "-", A$C1, "-", A$PY, "-", A$SO, "-", A$DT, "-", A$UT)
B$codig2<-paste(B$codigo, "-", B$C1, "-", B$PY, "-", B$SO, "-", B$DT, "-", B$UT)
C<-select(A, codig, TI, PY, IES)
D<-select(B, codig2, suma)
E<-as.data.frame(sqldf("select * from C A left join D B on A.codig=B.codig2
  "))
View(E)
G<- as.data.frame(E %>% group_by (IES, PY) %>%
  summarise (Nro.Citas=sum(as.numeric(suma))))
View(G)
##Documentos, Citaciones y Autores
#load("MC.RData")
#View(MC)
CAA<- MC %>% select(AU, TI, PY, C1, n, suma)
CAA<-filter(CAA, PY!="2017")
aut<- as.character(CAA$AU)
resaf<- strsplit(aut, ";")
sum(CAA$suma)
CAA$N_AU<-as.numeric(summary(resaf)[,1])
View(CAA)
R<-as.data.frame(CAA %>% group_by (suma, N_AU) %>%
  summarise (tot=n()))
R<-arrange(R, desc(suma))
names(R)<-c("Nro.citas", "Nro.autores", "Nro.artículos")
sum(R$Nro.artículos)
R$Cit.acu<-cumsum(R$Nro.citas*R$Nro.artículos)
R$Doc.acu<-cumsum(R$Nro.artículos)
R$Porc.acum.cit<-round((R$Cit.acu*100)/max(R$Cit.acu), 1)
View(R)
#Citaciones y Documentos
R1<-select(R, "Nro.citas", "Nro.artículos")
View(R1)
R2<-as.data.frame(R1 %>% group_by (Nro.citas) %>%
  summarise (Nro.artículos=sum(Nro.artículos)))
View(R2)

```

```

#
names (R2) <-c ("x", "y")
gr9<-ggplot (data=R2, aes (x=x, y=y)) +
  geom_point (color="blue") #+ geom_line (color="blue")
gr9
#Sin los superiores a 300
R3<-as.data.frame (sqldf ("select * from R2 where x<=300"))
dim (R3)
View (R3)
gr10<-ggplot (data=R3, aes (x=x, y=y)) +
  geom_point (color="blue") #+ geom_line (color="blue")
gr10
#Citaciones, autores
R$Cit.por.art<-R$Nro.citas*R$Nro.artículos
View (R)
R4<-select (R, "Cit.por.art", "Nro.autores")
View (R4)
R5<-as.data.frame (R4 %>% group_by (Nro.autores) %>%
  summarise (Citas.tot=sum (Cit.por.art)))
R5<-arrange (R5, desc (Citas.tot))
R5$Porc.Cit<-round (R5$Citas.tot*100/sum (R5$Citas.tot), 2)
R5$Porc.Cit.acu<-cumsum (R5$Porc.Cit)
View (R5)
sum (R5$Citas.tot [1:9])
#
names (R5) <-c ("N.AU", "N.C", "%C", "%C.A")

#####Indicadores de Colaboración
#load ("MC.RData")
## Autoría ecuatoriana-nacional y extranjera por documento
AEE<-MC
AEE$suma<-as.numeric (AEE$suma)
#Seleccionar procedencia
afil <- as.character (AEE$C1)
ti <- as.character (AEE$TI)
a <- as.character (AEE$PY)
cod<-as.character (AEE$codigo)
so<-as.character (AEE$SO)
dt<-as.character (AEE$DT)

```

```

ut<-as.character(AEE$UT)
suma<-as.numeric(AEE$suma)
resaf<- strsplit(afil, ";")
AEE$N_af<-as.numeric(summary(resaf)[,1])
View(AEE)
n <- as.numeric(summary(resaf)[,1])
TI<- character()
PY <- character()
SO <- character()
codigo<-character()
DT <- character()
UT <- character()
C1 <- character()
SU<-numeric()
for(i in 1:length(n)){
  TI <- c(TI,rep(ti[i],each = n[i]))
  PY <- c(PY,rep(a[i],each = n[i]))
  codigo<-c(codigo,rep(cod[i],each = n[i]))
  SO<-c(SO,rep(so[i],each = n[i]))
  DT<-c(DT,rep(dt[i],each = n[i]))
  UT<-c(UT,rep(ut[i],each = n[i]))
  C1<-c(C1,rep(afil[i],each = n[i]))
  SU<-c(SU,rep(suma[i],each = n[i]))
}
Af<-data.frame(AF=unlist(resaf),TI,PY,SO,DT,UT,C1,codigo,SU)
View(Af) ##sum(AEE$N_af)
x<-dim(Af)
patron <-"ECUADOR"
CN <- vector(x[1],mode = "integer")
CE <- vector(x[1],mode = "integer")
for(i in 1:x[1]){
  CN[i] <- ifelse(length(grep(patron,Af$AF[i])) == 0,0,1)
  CE[i] <- 1-CN[i]
}
Af$CN <- CN
Af$CE <- CE
View(Af)
AUCO<-Af %>% select(PY, TI, SO, codigo, CN, CE, DT, UT, C1, SU)
dim(AUCO)

```

```

View(AUCO)
COLAB<- as.data.frame(AUCO %>%
                    group_by (codigo,C1,PY,SO,DT,UT,SU) %>%
                    summarise ( TOT_EC_XA =sum(CN), TOT_EX_XA = sum(CE)
                                ))
View(COLAB)
save(COLAB, file = "COLAB.RData")##load("COLAB.RData")

##Autoría extranjera
COLAB1<-filter(COLAB,PY!="2017")
View(COLAB1)
save(COLAB1, file = "COLAB1.RData")##load("COLAB1.RData")
CNT<-filter(COLAB1,TOT_EC_XA>0)
k1<-dim(CNT)
CIT<-filter(COLAB1,TOT_EX_XA>0)
k2<-dim(CIT)
Doc_solo_ec<-k1[1]-k2[1]
Doc_solo_ec
Doc_con_ex<-k2[1]
Doc_con_ex
Porcentaje_col_solo_ec_tot<-round((Doc_solo_ec)*100/k1[1],2)
Porcentaje_col_solo_ec_tot
Porcentaje_col_in_tot<-round((Doc_con_ex)*100/k1[1],2)
Porcentaje_col_in_tot
#
TY<-as.data.frame(rbind(Porcentaje_col_solo_ec_tot,Porcentaje_col_in_tot))
TY$V1<-round(TY$V1,2)
cxacom<-TY$V1
names(cxacom) <- c(paste("Colab.Solo.Ec:",TY$V1[1],"%"),paste("Colab.Ex:",
  TY$V1[2],"%"))
colors = c("green","blue")
gr11<-pie(cxacom, col =colors,
  main = "Colaboración EC-EX: 2010-2016.", radius= 1.5)
## Colaboración Internacional por año
CIPA<-select(COLAB,PY,TOT_EX_XA )
dim(CIPA)
View(CIPA)
CIPA1<-filter(CIPA,TOT_EX_XA>0)
View(CIPA1)

```

```

PCE<-as.data.frame(table(CIPA1$PY))
names(PCE)<-c("A", "Nro.artículos.con.ce")
View(PCE)
YCPF<-merge(x=PC, y = PCE, all.x = TRUE)
View(YCPF)
YCPF$Nro.artículos.sin.ce<-YCPF$PC-YCPF$Nro.artículos.con.ce
YCPF$"%CI"<-round(YCPF$Nro.artículos.con.ce*100/YCPF$PC,1)
YCPF$"%CN"<-round(YCPF$Nro.artículos.sin.ce*100/YCPF$PC,1)
#
names(YCPF)<-c("A", "PC", "P.CI", "P.CN", "%CI", "%CN")
#
YCPF<-as.data.frame(YCPF)
test_data_long <- melt(YCPF[1:7,1:4], id="A")
names(test_data_long)<-c("A", "Variable", "Total")
gr12<-ggplot(data=test_data_long,
              aes(x=A, y=Total, colour=Variable,group=Variable)) +
  geom_line()+geom_point()
gr12
#
TF<-as.data.frame(YCPF[1:7,1:4])
TF
nl<-dim(TF)[1]
TF$Tasa_CN[1]<-""
TF$Tasa_CE[1]<-""
for(i in 2:nl){
  TF$Tasa_CN[i]=round((TF$D.CN[i]-TF$D.CN[i-1])/TF$D.CN[i-1],2)
  TF$Tasa_CE[i]=round((TF$D.CI[i]-TF$D.CI[i-1])/TF$D.CI[i-1],2)
}
View(TF)
EQP_CN<-1+as.numeric(TF$Tasa_CN[2:7])
EQP_CE<-1+as.numeric(TF$Tasa_CE[2:7])
cn <-length(EQP_CN)
ce <-length(EQP_CE)
T_M_A_CN<-round(((prod(EQP_CN)^(1/cn))-1)*100,2)
T_M_A_CN
T_M_A_CE<-round(((prod(EQP_CE)^(1/ce))-1)*100,2)
T_M_A_CE
##Citación vs Colaboración
CICO<-cbind.data.frame(TF[1:7,],TCI[1:7,])

```

```

CICO<-cbind.data.frame(CICO$A,CICO$CT,
                      CICO$D.CI,CICO$D.CN)
names(CICO)<-c("A","CT",
             "D.CI","D.CN")
View(CICO)
test_data_long <- melt(CICO, id="A")
names(test_data_long)<-c("A","Variable","Total")
gr13<-ggplot(data=test_data_long,
            aes(x=A, y=Total, colour=Variable,group=Variable)) +
  geom_line()+geom_point()
gr13

#####Indicadores de Visibilidad e Impacto
#H INDEX POR IES (Cada IES es un autor) 2010-2017*
#load("M1F.RData")
#load("COLAB.RData")
View(M1F)
View(COLAB)
uni<-c("EPN","ESPE","ESPOCH","ESPOL","PUCE",
      "SEK","UAE","UASB","UCE","UCSG","UCUENCA","UDA",
      "UDLA","UEES","UG","ULEAM","UNL",
      "UPS","UPSE","USFQ","UTA","UTE","UTM","UTMACH",
      "UTN","UTPL","UNACH","YACHAY","FLACSO","IAEN","UIDE","UNIANDES","UTI
      ")
uniA<-c("EPN","ESPOL","USFQ","UCUENCA","ESPE","UEES","FLACSO","UASB")
uniB<-c("ESPOCH","PUCE","UCSG","UCE","UDA","UNL","SEK","UPS",
      "UTA","UTN","UTPL","UTE","UTI","UTM","UDLA","UIDE",
      "UG","UTMACH","IAEN")
uniC<-c("UNACH","UNIANDES","UPSE","ULEAM","UAE")
#
H_I<-function (MFA,CB,uni){
  cd <- paste(MFA$AU,"-",MFA$TI,"-",MFA$C1,"-",MFA$PY,"-",MFA$SO,"-",MFA$DT
    , "-",MFA$UT)
  MFA$codig <- cd
  cd2<-paste(CB$codigo,"-",CB$C1,"-",CB$PY,"-",CB$SO,"-",CB$DT,"-",CB$UT)
  CB$codig2 <- cd2
  P1<-select(MFA,codig, TI, PY, IES)
  P2<-select(CB,codig2, SU)

```

```

P3<-as.data.frame(sqldf("select * from P1 A left join P2 B on A.codig=B.
    codig2"))
ART<-select (P3, IES, TI, SU, PY)
ART$IES<-as.character (ART$IES)
unih<-uni
ART1<-filter (ART, IES %in% unih)
H_IND<-as.data.frame (unih)
names (H_IND) <-c ("IES")
H<-numeric ()
for (i in 1:length (unih)) {
  nom <- unih [i]
  ART1<-filter (ART, IES %in% nom)
  ART2<-as.data.frame (sqldf ("select * from ART1 order by SU desc"))
  d_a<-dim (ART2)
  ART2$INDICE<-c (1:d_a [1])
  ART2$ver <- ifelse (ART2$SU >= ART2$INDICE , 1, 0)
  ART3<-filter (ART2, ART2$ver==1)
  hindex<-dim (ART3) [1]
  H [i] <-hindex }
H_IND$H<-H
H_IND<-as.data.frame (sqldf ("select * from H_IND order by H desc"))
return (H_IND)
}
#
H_TOTAL<-H_I (M1F, COLAB, uni)
H_TOTAL

### IAEN, UIDE, UNIANDES NO TIENEN VALOR H INDEX
H_TOTAL_F<-filter (H_TOTAL, H!=0)
H_TOTAL_F
# Impacto
d1<-quantile (H_TOTAL_F$H, 0.25)
d1
d2<-quantile (H_TOTAL_F$H, 0.50)
d2
d3<-quantile (H_TOTAL_F$H, 0.75)
d3
H_TOTAL_F$Nivel.Impacto<-ifelse (H_TOTAL_F$H<=d1, "Bajo", ifelse ((d1<H_TOTAL_F
    $H & H_TOTAL_F$H<=d2), "Moderado", ifelse ((d2<H_TOTAL_F$H & H_TOTAL_F$H<=

```



```

    d3), "Medio", "Alto"))))
View(H_TOTAL_F)
#
HA<-arrange(filter(H_TOTAL, IES %in% uniA), H)
HB<-filter(H_TOTAL, IES %in% uniB)
HC<-filter(H_TOTAL, IES %in% uniC)
HA<-as.data.frame(sqldf("select * from HA order by H desc"))
HB<-as.data.frame(sqldf("select * from HB order by H desc"))
HC<-as.data.frame(sqldf("select * from HC order by H desc"))
View(HA)
View(HB)
View(HC)

##Indice H por categoría: 2010-2016
#load("M1F.RData")
#load("COLAB1.RData")
uniA<-c("EPN", "ESPOL", "USFQ", "UCUENCA", "ESPE", "UEES", "FLACSO", "UASB")
uniB<-c("ESPOCH", "PUCE", "UCSG", "UCE", "UDA", "UNL", "SEK", "UPS",
        "UTA", "UTN", "UTPL", "UTE", "UTI", "UTM", "UDLA", "UIDE",
        "UG", "UTMACH", "IAEN")
uniC<-c("UNACH", "UNIANDES", "UPSE", "ULEAM", "UAE")
a1<-2010:2011
a2<-2010:2012
a3<-2010:2013
a4<-2010:2014
a5<-2010:2015
a6<-2010:2016
## Categoría A
uni<-uniA
M1h<-filter(M1F, PY %in% a1)
H1<-H_I(M1h, COLAB1, uni)
M2h<-filter(M1F, PY %in% a2)
H2<-H_I(M2h, COLAB1, uni)
M3h<-filter(M1F, PY %in% a3)
H3<-H_I(M3h, COLAB1, uni)
M4h<-filter(M1F, PY %in% a4)
H4<-H_I(M4h, COLAB1, uni)
M5h<-filter(M1F, PY %in% a5)
H5<-H_I(M5h, COLAB1, uni)

```

```

M6h<-filter(M1F,PY %in% a6)
H6<-H_I(M6h,COLAB1,uni)
H_A<-cbind.data.frame(H1,H2,H3,H4,H5,H6)
names(H_A)<-c("IES1","H1","IES2","H2","IES3","H3",
             "IES4","H4","IES5","H5","IES6","H6")
View(H_A)
## Categoría B
uni<-uniB
M3h<-filter(M1F,PY %in% a3)
H3<-H_I(M3h,COLAB1,uni)
M4h<-filter(M1F,PY %in% a4)
H4<-H_I(M4h,COLAB1,uni)
M5h<-filter(M1F,PY %in% a5)
H5<-H_I(M5h,COLAB1,uni)
M6h<-filter(M1F,PY %in% a6)
H6<-H_I(M6h,COLAB1,uni)
H_B<-cbind.data.frame(H3,H4,H5,H6)
names(H_B)<-c("IES3","H3",
             "IES4","H4","IES5","H5","IES6","H6")
View(H_B)
## Categoría C
uni<-uniC
M3h<-filter(M1F,PY %in% a3)
H3<-H_I(M3h,COLAB1,uni)
M4h<-filter(M1F,PY %in% a4)
H4<-H_I(M4h,COLAB1,uni)
M5h<-filter(M1F,PY %in% a5)
H5<-H_I(M5h,COLAB1,uni)
M6h<-filter(M1F,PY %in% a6)
H6<-H_I(M6h,COLAB1,uni)
H_C<-cbind.data.frame(H3,H4,H5,H6)
names(H_C)<-c("IES3","H3",
             "IES4","H4","IES5","H5","IES6","H6")
View(H_C)
# Impacto A
uni<-uniA
M1Fh<-filter(M1F,PY!="2017")
H_A<-H_I(M1Fh,COLAB1,uni)
View(H_A)

```

```

d1<-quantile (H_A$H,0.25)
d1
d2<-quantile (H_A$H,0.50)
d2
d3<-quantile (H_A$H,0.75)
d3
H_A$Nivel.Impacto<-ifelse (H_A$H<=d1, "Bajo", ifelse ((d1<H_A$H & H_A$H<=d2), "
    Moderado", ifelse ((d2<H_A$H & H_A$H<=d3), "Medio", "Alto")))
View(H_A)
# Impacto B
uni<-uniB
M1Fh<-filter (M1F,PY!="2017")
H_B<-H_I (M1Fh, COLAB1, uni)
View(H_B)
### IAEN, UIDE NO TIENEN VALOR H INDEX
H_B<-filter (H_B,H!=0)
View(H_B)
d1<-quantile (H_B$H,0.25)
d1
d2<-quantile (H_B$H,0.50)
d2
d3<-quantile (H_B$H,0.75)
d3
H_B$Nivel.Impacto<-ifelse (H_B$H<=d1, "Bajo", ifelse ((d1<H_B$H & H_B$H<=d2), "
    Moderado", ifelse ((d2<H_B$H & H_B$H<=d3), "Medio", "Alto")))
View(H_B)
# Impacto C
uni<-uniC
M1Fh<-filter (M1F,PY!="2017")
H_C<-H_I (M1Fh, COLAB1, uni)
View(H_C)
### UNIANDES NO TIENEN VALOR H INDEX
H_C<-filter (H_C,H!=0)
View(H_C)
d1<-quantile (H_C$H,0.25)
d1
d2<-quantile (H_C$H,0.50)
d2
d3<-quantile (H_C$H,0.75)

```

```

d3
H_C$Nivel.Impacto<-ifelse (H_C$H<=d1, "Bajo", ifelse ((d1<H_C$H & H_C$H<=d2), "
  Moderado", ifelse ((d2<H_C$H & H_C$H<=d3), "Medio", "Alto")))
View(H_C)
#
gr14<-ggplot (data=H_A, aes (x=reorder (IES, +H), y=H, fill=H)) +
  geom_bar (stat="identity", position=position_dodge ()) +
  labs (x = "IES", y = "H")+ coord_flip ()
gr14
#
gr15<-ggplot (data=H_B[1:17, ], aes (x=reorder (IES, +H), y=H, fill=H)) +
  geom_bar (stat="identity", position=position_dodge ()) +
  labs (x = "IES", y = "H")+ coord_flip ()
gr15
#
gr16<-ggplot (data=H_C[1:4, ], aes (x=reorder (IES, +H), y=H, fill=H)) +
  geom_bar (stat="identity", position=position_dodge ()) +
  labs (x = "IES", y = "H")+ coord_flip ()
gr16

View(MC)
##### Áreas de Producción
dfo<-select (Scopus_SourceList, Title, ASJC)
dfo$Title<-toupper (dfo$Title)
View(dfo)
names (Scopus_SourceList)
View(Scopus_ASJC)
View(Scopus_SourceList)
#load("MC.RData")
MCp<-filter (MC, PY!="2017")
S<-gsub ("", "", MCp$SO, fixed = TRUE)
S<-sort (table (S), decreasing = TRUE)
View(S)
S<-as.data.frame (S)
sum (S$Freq)
ds <- read.csv ( "SOUR.csv", sep = ";", head = T )
View(ds)
d<- as.character (ds$J)
resaf<- strsplit (d, ";")

```

```

ds$N<-as.numeric(summary(resaf)[,1])
View(ds)
so<-as.character(ds$SO)
SO<-as.character()
n <- as.numeric(summary(resaf)[,1])
for(i in 1:length(n)){
  SO<-c(SO,rep(so[i],each = n[i]))
}
Af<-data.frame(AF=unlist(resaf),SO)
View(Af)
x<-dim(Af)
#
dfo<-select(Scopus_ASJC,ASJC,Description)
View(dfo)
df1<-as.data.frame(sqldf("select * from Af A left join dfo B on A.AF=B.ASJC
"))
View(df1)
ARI<-as.data.frame(sqldf("select * from S A left join df1 B on A.S=B.SO"))
View(ARI)
#
ARI1<-select(ARI,SO,Freq,Description)
names(ARI1)<-c("SO","Nro.articulos_so","AREA" )
View(ARI1)
patron <-"([all()])"
ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"",ARI1,ignore.case=T))
patron <-"([miscellaneous()])"
ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"",ARI1,ignore.case=T))
patron <-"([medical()])"
ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"",ARI1,ignore.case=T))
patron <-"([social science()])"
ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"",ARI1,ignore.case=T))
patron <-"([miscellaneous()])"
ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"Veterinary",ARI1,ignore.
case=T))
patron <-"([clinical()])"
ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"Veterinary",ARI1,ignore.
case=T))
patron <-"([nursing()])"

```

```

ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"Veterinary",ARI1,ignore.
  case=T))
patron <-"veterinary"
ARI1<-apply(ARI1, 2, function(ARI1) gsub(patron,"Veterinary",ARI1,ignore.
  case=T))
ARI1<-as.data.frame(ARI1)
ARI1$Nro.articulos_so<-as.numeric(ARI1$Nro.articulos_so)
#Composición en porcentaje
CEP<- as.data.frame(ARI1 %>%
  group_by (AREA) %>%
  summarise ( TOT=sum(Nro.articulos_so), TOT_AP=n()))
##TOT_AP es las veces que apareció en la tabla ARI1
CEP$POR<-round( (CEP$TOT/sum(CEP$TOT) ) *100, 2)
CEP<-arrange (CEP, desc (POR) )
CEP$POR.A<-cumsum(CEP$POR)
View(CEP)
CEP1<-select (CEP, AREA, TOT, POR, POR.A)
View(CEP1)
CEP2<-filter(CEP1, TOT>130)
cxacom<-as.numeric(as.character(CEP2$POR) )
names(cxacom) <- paste(CEP2$AREA, CEP2$POR, "%")
gr17<-pie(cxacom, col = rainbow(length(names(cxacom))),
  main = "Áreas de producción: 2010-2016.", radius= 1.8)

##### MODELOS #####

#####CRECIMIENTO DE LA PRODUCCIÓN:PRICE
View(PC)
x<-PC$A[1:7]
t<-0:6
P<-as.numeric(PC$PC[1:7])
Datos <- data.frame(t,P)
View(Datos)
GP<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y=PC)
GP
Po<-P[1]

```

```

#Regresión
regresion1<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresion1
sr<-summary(regresion1)
sr
#Prueba KS
FR<-as.data.frame(fitted(regresion1))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad
#
a<-sr$coefficients[1]
a
names(TH)<-c("t","Val.Real","Val.Ajust")
test_data_long <- suppressWarnings((melt(TH, id="t")))
names(test_data_long)<-c("A","Variable","Total")
gr18<-ggplot(data=test_data_long,
             aes(x=A, y=Total, colour=Variable,group=Variable)) +
  geom_line()
gr18
##PROYECCIÓN AL 2020
#Como se obtuvo el parametro a, entonces:
A<-2017:2020
PC_P<-as.data.frame(A)
sr<-summary(regresion1)
a<-sr$parameters[1,1]
a
for(i in 8:11){
  PC_P$PCP[i-7]<-round(Po*exp(a*(i-1)),0)
}

```

```

PC_P
View(PC_P)
##### Price: Categorías
uniA_p<-c("EPN","ESPOL","USFQ","UCUENCA","ESPE","UEES")
PRI_A<-filter(PC_I,IES %in% uniA_p,A !="2017")
View(PRI_A)
PT1<- as.data.frame(PRI_A %>% group_by (A) %>%
                    summarise (P=sum(Nro.artículos)))
P_A_T<-ggplot(data=PT1, aes(x=A, y=P, group=1)) +
  geom_line(color="blue")+
  geom_point()
P_A_T
View(PT1)
###
t<-0:6
P<-as.numeric(PT1$P)
Datos <- data.frame(t,P)
View(Datos)
Po<-P[1]
###
regresion1<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresion1
sr<-summary(regresion1)
sr
#Prueba KS
FR<-as.data.frame(fitted(regresion1))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad
#####

```



```

uniB_p<-c("ESPOCH", "PUCE", "UCSG", "UCE", "UDA", "UNL", "SEK", "UPS",
          "UTA", "UTN", "UTPL", "UTE", "UTI", "UTM", "UDLA", "UIDE",
          "UG", "UTMACH", "IAEN")
PRI_B<-filter(PC_I, IES %in% uniB_p, A != "2017")
View(PRI_B)
PT2<- as.data.frame(PRI_B %>% group_by (A) %>%
          summarise (P=sum(Nro.artículos)))
P_B_T<-ggplot(data=PT2, aes(x=A, y=P, group=1)) +
  geom_line(color="blue")+
  geom_point()
P_B_T
View(PT2)
###
t<-0:6
P<-as.numeric(PT2$P)
Datos <- data.frame(t,P)
View(Datos)
Po<-P[1]
###
regresion1<-nls(P~ Po*exp(a*t), Datos, start=list(a=1.1))
regresion1
sr<-summary(regresion1)
sr
#Prueba KS
FR<-as.data.frame(fitted(regresion1))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P, FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t, Datos$P, round(FR, 0))
names(TH)<-c("t", "P", "FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad
#####
uniC_p<-c("UNACH", "UNIANDES", "UPSE", "ULEAM", "UAE")

```

```

PRI_C<-filter(PC_I, IES %in% uniC_p, A !="2017")
View(PRI_C)
PT3<- as.data.frame(PRI_C %>% group_by (A) %>%
      summarise (P=sum(Nro.artículos)))
P_C_T<-ggplot(data=PT3, aes(x=A, y=P, group=1)) +
  geom_line(color="blue")+
  geom_point()
P_C_T
View(PT3)
###
t<-0:6
P<-as.numeric(PT3$P)
Datos <- data.frame(t,P)
View(Datos)
Po<-P[1]
###
regresion1<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresion1
sr<-summary(regresion1)
sr
#Prueba KS
FR<-as.data.frame(fitted(regresion1))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad

####PRICE: Categoría A
uniA_p<-c("EPN","ESPOL","USFQ","UCUENCA","ESPE","UEES")
PRI_A<-filter(PC_I, IES %in% uniA_p, A !="2017")
View(PRI_A)

```

```

GP_A<-ggplot (data=PRI_A,
              aes(x=A, y=Nro.artículos, colour=IES,group=IES)) +
  geom_line()+labs (x="A",y="PC")
GP_A
#### EPN ####
AUX<-filter(PRI_A, IES %in% "EPN")
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr19<-ggplot (data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y=PC)
gr19
Po<-P[1]
#Regresión
regresionla<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresionla
sr<-summary(regresionla)
sr
a<-sr$parameters[1,1]
a
#Prueba KS
FR<-as.data.frame(fitted(regresionla))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresionla$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad
##PROYECCIÓN AL 2020
#Como se obtuvo el parametro a, entonces:
A<-2017:2020

```

```

PC_P1<-as.data.frame(A)
sr<-summary(regresion1a)
a<-sr$parameters[1,1]
a
for(i in 8:11){
  PC_P1$PCP[i-7]<-round(Po*exp(a*(i-1)),0)
}
PC_P1
View(PC_P1)
#### ESPOL ####
AUX<-filter(PRI_A, IES %in% "ESPOL")
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr20<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y=PC)
gr20
Po<-P[1]
#Regresión
regresion1b<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresion1b
sr<-summary(regresion1b)
sr
a<-sr$parameters[1,1]
a
#Prueba KS
FR<-as.data.frame(fitted(regresion1b))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1b$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)

```

```

R.cuad<-1-(SSR/SST)
R.cuad
##PROYECCIÓN AL 2020
#Como se obtuvo el parametro a, entonces:
A<-2017:2020
PC_P2<-as.data.frame(A)
sr<-summary(regresion1b)
a<-sr$parameters[1,1]
a
for(i in 8:11){
  PC_P2$PCP[i-7]<-round(Po*exp(a*(i-1)),0)
}
PC_P2
View(PC_P2)
#### USFQ ####
AUX<-filter(PRI_A, IES %in% "USFQ")
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr21<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y=PC)
gr21
Po<-P[1]
#Regresión
regresion1c<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresion1c
sr<-summary(regresion1c)
sr
a<-sr$parameters[1,1]
a
#Prueba KS
FR<-as.data.frame(fitted(regresion1c))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado

```

```

TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1c$m$deviance() #### desviacion del modelo ajustado
SST<-sum((TH$P-mean(TH$P))^2) #### variabilidad total
R.cuad<-1-(SSR/SST)
R.cuad
##PROYECCIÓN AL 2020--El modelo no ajusta satisfactoriamente.
#### UCUENCA ####
AUX<-filter(PRI_A,IES %in% "UCUENCA")
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr22<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y=PC)
gr22
Po<-P[1]
#Regresión
regresion1d<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresion1d
sr<-summary(regresion1d)
sr
a<-sr$parameters[1,1]
a
#Prueba KS
FR<-as.data.frame(fitted(regresion1d))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1d$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)

```

```

R.cuad
##PROYECCIÓN AL 2020
#Como se obtuvo el parametro a, entonces:
A<-2017:2020
PC_P3<-as.data.frame(A)
sr<-summary(regresion1d)
a<-sr$parameters[1,1]
a
for(i in 8:11){
  PC_P3$PCP[i-7]<-round(Po*exp(a*(i-1)),0)
}
PC_P3
View(PC_P3)
#### ESPE ####
AUX<-filter(PRI_A, IES %in% "ESPE")
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr23<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y=PC)
gr23
Po<-P[1]
#Regresión
regresionle<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresionle
sr<-summary(regresionle)
sr
a<-sr$parameters[1,1]
a
#Prueba KS
FR<-as.data.frame(fitted(regresionle))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))

```

```

names (TH) <-c ("t", "P", "FT")
#R.cuad=1-(SSR/SST)
SSR<-regresionle$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad
##PROYECCIÓN AL 2020
#Como se obtuvo el parametro a, entonces:
A<-2017:2020
PC_P4<-as.data.frame(A)
sr<-summary(regresionle)
a<-sr$parameters[1,1]
a
for(i in 8:11){
  PC_P4$PCP[i-7]<-round(Po*exp(a*(i-1)),0)
}
PC_P4
View(PC_P4)
#### UEES ####
AUX<-filter(PRI_A, IES %in% "UEES") ## En este caso, en el 2010 es 0.
AUX
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr24<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y=PC)
gr24
#Se considera t0=2011
Datos<-Datos[2:7,]
Datos$t<-0:5
Po<-Datos$P[1]
#Regresión
regresionlf<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresionlf
sr<-summary(regresionlf)
sr
a<-sr$parameters[1,1]

```



```

a
#Prueba KS
FR<-as.data.frame(fitted(regresion1f))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1f$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad
##PROYECCIÓN AL 2020--El modelo no ajusta satisfactoriamente.
#### PROYECCIONES MODELOS EXPONENCIALES
PR_A<-cbind.data.frame(PC_P1,PC_P2,PC_P3,PC_P4)
names(PR_A)<-c("A","EPN","A1","ESPOL","A2","UCUENCA","A3","ESPE")
PR_AF<-select(PR_A,"A","EPN","ESPOL","UCUENCA","ESPE")
View(PR_AF)
#### ALTERNATIVAS PARA USFQ Y UEES
#### UEES ####
AUX<-filter(PRI_A,IES %in% "UEES")
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr25<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y="PC")
gr25
# Tomar una regr. lineal permite considerar como t0=0, a diferencia de
  Price.
regresion1<-lm(P~t,Datos)
regresion1
sr<-summary(regresion1)
sr
# El intercepto es no significativo

```

```

regresion2<-lm(P~t-1,Datos)
regresion2
sr<-summary(regresion2)
sr
#Prueba F
F.T<-sr$fstatistic
F.T
#R cuadrado
R.cuad<-sr$r.squared
R.cuad
##PROYECCIÓN AL 2020
### Como se obtuvo el coeficiente asociado al t, entonces:
A<-2017:2020
PC_P5<-as.data.frame(A)
c<-sr$coefficients[1]
c
for(i in 8:11){
  PC_P5$PCP[i-7]<-round(c*(i-1),0)
}
View(PC_P5)
#### USFQ ####
AUX<-filter(PRI_A, IES %in% "USFQ")
t<-0:6
P<-as.numeric(AUX$Nro.artículos)
Datos <- data.frame(t,P)
View(Datos)
gr26<-ggplot(data=Datos, aes(x=t, y=P, fill=P)) +
  geom_line(color="blue")+
  geom_point()+labs(y="PC")
gr26
Po<-P[1]
#Subperíodo
Datos<-Datos[4:7,]
Datos$t<-0:3
Po<-Datos$P[1]
#Regresión
regresion1<-nls(P~ Po*exp(a*t),Datos, start=list(a=1.1))
regresion1
sr<-summary(regresion1)

```

```

sr
a<-sr$parameters[1,1]
a
#Prueba KS
FR<-as.data.frame(fitted(regresion1))
names(FR)<-c("fitted.reg")
View(FR)
KS<-suppressWarnings(ks.test(Datos$P,FR$fitted.reg))
KS
#R cuadrado
TH<-cbind.data.frame(Datos$t,Datos$P,round(FR,0))
names(TH)<-c("t","P","FT")
#R.cuad=1-(SSR/SST)
SSR<-regresion1$m$deviance()
SST<-sum((TH$P-mean(TH$P))^2)
R.cuad<-1-(SSR/SST)
R.cuad
##PROYECCIÓN AL 2020
#Como se obtuvo el parametro a, entonces:
A<-2017:2020
PC_P6<-as.data.frame(A)
sr<-summary(regresion1)
a<-sr$parameters[1,1]
a
for(i in 8:11){
  PC_P6$PCP[i-7]<-round(Po*exp(a*(i-1)),0)
}
View(PC_P6)

#####LOTKA: ÍNDICE DE PRODUCTIVIDAD
#load("M1F.RData")
#View(M1F)
ART<-select(M1F,IES,AU)
View(ART)
# Lotka del paquete Bibliometrix
L<-function (ART1) {
  listAU = strsplit(as.character(ART1$AU), ";")
  listAU = lapply(listAU, function(l) trim.leading(l))
  listAU = lapply(listAU, function(l) {

```

```

l = trim.leading(l)
l = sub(" ", ",", l, fixed = TRUE)
l = sub(",,", ",", l, fixed = TRUE)
l = gsub(" ", "", l, fixed = TRUE)
})
AU = gsub(" ", "", unlist(listAU), fixed = TRUE)
Authors = sort(table(AU), decreasing = TRUE)
AUdf = data.frame(Authors)
AuthorProd = aggregate(AUdf, by = list(AUdf$Freq), "length")
AuthorProd[, 2] = as.numeric(AuthorProd[, 2])
AuthorProd[, 3] = AuthorProd[, 2]/sum(AuthorProd[, 2])
names(AuthorProd) = c("N.Articles", "N.Authors", "Freq")
LOTKA = (lm(log10(Freq) ~ log10(N.Articles), data = AuthorProd))
Yt = LOTKA$coeff[1] - 2 * log10(AuthorProd[, 1])
p = suppressWarnings(ks.test(AuthorProd[, 3], 10^Yt, exact = FALSE)$p.
  value)
L = list(Beta = abs(as.numeric(LOTKA$coeff[2])), C = 10^(as.numeric(LOTKA
  $coeff[1])),
  R2 = summary(LOTKA)$r.squared, fitted = 10^LOTKA$fitted.values,
  p.value = p, AuthorProd = AuthorProd)
return(L[5])
}
#Categoría A
uniA_C<-c("EPN", "ESPOL", "USFQ", "UCUENCA", "ESPE", "UEES", "UASB", "FLACSO")
A<-character()
LA<-numeric()
for(i in 1:length(uniA_C)){
  nom<-uniA_C[i]
  ART1<-filter (ART, IES %in% nom )
  A[i]<-nom
  LA[i]<-as.numeric(L(ART1))
}
LA<-round(LA, 3)
LOTKA_A<-as.data.frame(cbind(A, LA))
LOTKA_A<-arrange(LOTKA_A, desc(LA))
View(LOTKA_A)
### Categoría B
uniB_C<-c("ESPOCH", "PUCE", "UCSG", "UCE", "UDA", "UNL", "SEK", "UPS",

```

```

      "UTA", "UTN", "UTPL", "UTE", "UTM", "UDLA", "UG", "UTMACH", "UTI") # Sin "
      UIDE", "IAEN"
B<-character()
LB<-numeric()
for(i in 1:length(uniB_C)){
  nom<-uniB_C[i]
  ART1<-filter (ART, IES %in% nom )
  B[i]<-nom
  LB[i]<-as.numeric(L(ART1))
}
LB<-round(LB,3)
LOTKA_B<-as.data.frame(cbind(B, LB))
LOTKA_B<-arrange(LOTKA_B, desc(LB))
View(LOTKA_B)
### Categoría C
uniC_C<-c("UNACH", "UPSE", "ULEAM", "UAE", "UNIANDES")
C<-character()
LC<-numeric()
for(i in 1:length(uniC_C)){
  nom<-uniC_C[i]
  ART1<-filter (ART, IES %in% nom )
  C[i]<-nom
  LC[i]<-as.numeric(L(ART1))
}
LC<-round(LC,3)
LOTKA_C<-as.data.frame(cbind(C, LC))
LOTKA_C<-arrange(LOTKA_C, desc(LC))
View(LOTKA_C)
### YACHAY
uni_Y<-c("YACHAY")
C<-character()
LC<-numeric()
for(i in 1:length(uni_Y)){
  nom<-uni_Y[i]
  ART1<-filter (ART, IES %in% nom )
  C[i]<-nom
  LC[i]<-as.numeric(L(ART1))
}
LC<-round(LC,3)

```

```

LOTKA_Y<-as.data.frame(cbind(C,LC))
LOTKA_Y
##ÍNDICE PRODUCTIVIDAD PERSONAL PARA LA EPN
View(ART)
AR1<-filter (ART, IES %in% "EPN")
PP<-function (ART1) {
  listAU = strsplit(as.character(ART1$AU), ";")
  listAU = lapply(listAU, function(l) trim.leading(l))
  listAU = lapply(listAU, function(l) {
    l = trim.leading(l)
    l = sub(" ", ",", l, fixed = TRUE)
    l = sub(",,", ",", l, fixed = TRUE)
    l = gsub(" ", "", l, fixed = TRUE)
  })
  AU = gsub(" ", "", unlist(listAU), fixed = TRUE)
  Authors = sort(table(AU), decreasing = TRUE)
  ##View(Authors)
  AUdf = data.frame(Authors)
  return(AUdf) }
IPP<-PP (AR1)
names(IPP)<-c("AU","Nro.Documentos")
IPP$IP<-log10(IPP$Nro.Documentos)
IPP_F<-sqldf("select A.*,CASE WHEN IP>=1 THEN 'Grandes Productores'
              WHEN IP==0 THEN 'Productores Transitorios'
              ELSE 'Productores Intermedios 'END TIPO FROM
              IPP A")
View(IPP_F)
PG<-as.data.frame(IPP_F %>% group_by (TIPO)
                  %>% summarise (N.au=n()))
cxacom<- PG$N.au
names(cxacom) <- paste(PG$TIPO, "(",PG$N.au, ")")
gr27<-pie(cxacom, col = rainbow(length(names(cxacom))),
          main = "Composición EPN", radius=1.5)
pr<-round(PG$N.au*100/sum(PG$N.au),2)
pr
#####

```

Bibliografía

- [1] CEAACES, 2017,
<http://www.ceaaces.gob.ec/sitio/>
- [2] *Luventicus*, Recuperado en 2016 de <http://www.luventicus.org/articulos/02R011/index.html>
- [3] ROMANI F., HUAMANI C., GONZALES-ALCAIDE G., (2011), *Estudios Bibliométricos como línea de investigación en las ciencias biomédicas: una aproximación para el pregrado*, Lima, Perú.
- [4] GLANZEL W., (2003), *Bibliometrics as a research field*.
- [5] ESCOBAR DE MOREL M., VELÁZQUEZ DE MALDONADO H., (2011), *Productividad científica en la Universidad Nacional de Asunción*.
- [6] *USFQ*, Recuperado en 2016 de <http://noticias.usfq.edu.ec/2016/05/0034-de-la-ciencia-se-publica-en-ecuador.html>
- [7] CASTILLO L., *Introducción a la información científica y técnica. Curso 2001-2002*.
- [8] PRAT A., *Working Paper: Módulo de capacitación para la recolección y el análisis de indicadores de producto de las actividades de ciencia y tecnología.*, BID.
- [9] ECURED, Recuperado en 2017 de https://www.ecured.cu/Bibliometría#Ley_de_Zipf
- [10] ELSEVIER, Recuperado en 2017 de https://service.elsevier.com/app/answers/detail/a_id/11274/kw/content%20policy%20and%20selection/supporthub/scopus/related/1/.
- [11] CEAACES, Recuperado en 2017 de <http://www.ceaaces.gob.ec/sitio/acreditacion-y-categorizacion/>.

- [12] KOSKINEN J., MIETTUNEN J., PAAJALA H., LAURONEN E., NIEMINEN P., KOPONEN H., TIENARI P., ISOHANNI M., (2008), *How to use bibliometric methods in evaluation of scientific research? An example from Finnish schizophrenia research.*
- [13] SISA I., ESPINEL M., FORNASINI M., MANTILLA G. , (2011), *Scientific output in the health sciences in Ecuador.*
- [14] DÍAZ T., RODRÍGUEZ M., (2017), *Indicadores bibliométricos de las tesis de pregrado en el repositorio digital de la Facultad de Ciencias Humanísticas y Sociales de la Universidad Técnica de Manabí, Ecuador, período 2015 – 2016.*
- [15] CELI L., MUÑOZ DE CAMACHO S. , (2017), *Posicionamiento de los Investigadores del Ecuador en Webometrics.*
- [16] URBIZAGASTEGUI R., (2016), *Bibliometrics, informetrics, scientometrics and other "Metricin Brazil.*
- [17] QUINDEMIL E., CEDEÑO R. , (2016), *Análisis bibliométrico de la literatura científica publicada en la revista La Técnica (2010 – 2016).*
- [18] PEÑA D., (2011), *Some reasons from bibliometrics to evaluate the Venezuelan scientific research.*
- [19] PAZMIÑO M. , (2014), *Análisis Bibliométrico de la Producción Científica de Influenza A-H1N1 en América del Sur desde abril del 2009 a abril del 2013: Identificando vacíos de conocimiento en Salud Pública sobre A-H1N1 en la región.*
- [20] SISA I., (2010), *Análisis Bibliométrico de la Producción Científica en Ciencias de la Salud en Ecuador 1999 – 2009: Hacia una Política de Investigación y Desarrollo en Función de las Prioridades de Salud del País.*
- [21] BRUQUE S., (2013), *La producción científica en Ecuador en el contexto latinoamericano. Un análisis bibliométrico comparativo.*
- [22] FINCH A., *Metrics*, CSIRO.
- [23] *BIBLIOMETRIX*, 2017, <http://www.bibliometrix.org/>
- [24] LOTKA ALFRED J., (1926), *The frequency distribution of scientific productivity*, Journal of the Washington Academy of Science.

- [25] GORBEA PORTAL S., (2005), *El modelo matemático de Lotka: su aplicación a la producción científica latinoamericana en ciencias bibliotecológica y de la información*, UNAM: Centro Universitario de Investigaciones Bibliotecológicas, México.
- [26] RUIZ BAÑOS R., *Ley de Lotka*, Universidad de Granada: Departamento de Biblioteconomía y Documentación.
- [27] ÁLVAREZ-MUÑOZ P., PÉREZ-MONTORO M., *Análisis de la producción y de la visibilidad científica de Ecuador en el contexto andino (2000-2013)*, El profesional de la información.
- [28] GAGOLEWSKI M., *Bibliometric Impact Assessment With R and the CITAN Package*, Polonia.
- [29] CASTERÁ V.J.T., (2013), *Estudio bibliométrico de la producción científica y de consumo de las revistas sobre nutrición indizadas en la Red SciElo*, Universidad de Alicante.
- [30] *CSUScopusNarrative*, Chicago State University, Recuperado en 2017 de <http://www.csu.edu/CTRE/pdf/CSUScopusNarrative.pdf>
- [31] *Manejo de cadenas y expresiones regulares.*, Universitat de Barcelona, Recuperado en 2017 de <http://www.ub.edu/stat/docencia/Cursos-R/Radvanced/materials/CadenasenR.pdf>
- [32] MIRÓ Ò., BURBANO P. , (2013), *El factor de impacto, el índice h y otros indicadores bibliométricos.*, Grupo de investigación “Urgencias: procesos y patologías”, IDIBAPS, Barcelona.
- [33] RUSSELL J. M., MADERA JARAMILLO M. J., AINSWORTH S. , (2009), *El análisis de redes en el estudio de la colaboración científica.*
- [34] GONZÁLEZ DE DIOS J., MOYA M., MATEOS HERNÁNDEZ M. A. , (1997), *Indicadores bibliométricos: Características y limitaciones en el análisis de la actividad científica*, Anales Españoles de Pediatría.
- [35] ESCORCIA T. , (2008), *El análisis bibliométrico como herramienta para el seguimiento de publicaciones científicas, tesis y trabajos de grado.*, Pontificia Universidad Javeriana, Bogotá.
- [36] ESCALONA FERNANDEZ M. I., LAGAR BARBOSA P., Y PULGARÍN GUERRERO A. , (2010), *WEB OF SCIENCE VS. SCOPUS: Un estudio cuantitativo en Ingeniería Química.*, Anales de Documentación, 159-175.

- [37] RUIZ BAÑOS R., *El crecimiento de la Ciencia.*, Universidad de Granada: Departamento de Biblioteconomía y Documentación.
- [38] PUCHE R. C., (2015), *Editorial: El crecimiento exponencial del conocimiento médico y la actualización permanente de nuestros graduados.* , Rev. Méd. Rosario 81: 102-105.
- [39] TÚÑEZ LÓPEZ M., (2013), *El 'índice h' de la investigación en Comunicación en España, Portugal y Latinoamérica: Web of Knowledge (WoK), Scopus y Google Scholar Metrics.*, Communication&Society/Comunicación y Sociedad, Vol.26, n.4, pp.53-75.
- [40] CEAACES, *Instrumentos de Cuantificación del Desempeño Institucional.*
- [41] GONZÁLEZ ALCAIDE. G., GÓMEZ FERRI. J., (2014), *La colaboración científica: principales líneas de investigación y retos de futuro.*, Revista Española de Documentación Científica, 37.
- [42] HUGGETT S., GURNEY T., JUMELET A., (2016), *INDICADORES BIBLIOMÉTRICOS DE LA ACTIVIDAD CIENTÍFICA ESPAÑOLA 2005-2014.*, Fundación Española para la Ciencia y la Tecnología.