



FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

Perspectiva de género en la investigación en Estadística e Investigación Operativa

Andrea García Martínez

2020/2021

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GRAO DE MATEMÁTICAS

Traballo Fin de Grao

Perspectiva de género en la investigación en Estadística e Investigación Operativa

Andrea García Martínez

Julio 2021

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Trabajo propuesto

Área de Coñecemento: Estadística e Investigación Operativa
Título: Perspectiva de género en la investigación en Estadística e Investigación Operativa
Breve descripción do contido
La promoción del papel de la mujer en diferentes ámbitos de la sociedad ha cobrado especial importancia en los últimos años. En este trabajo se estudiará el impacto de las políticas de perspectiva de género en la investigación en Estadística e Investigación Operativa.
Recomendacións
Outras observacións

Índice general

Resumen	VIII
Introducción	XI
1. El género en la investigación	1
2. Datos y metodología	7
2.1. Recogida de datos	7
2.2. Metodología	16
2.2.1. Modelo de Regresión Lineal Simple	16
2.2.2. Modelo de Regresión Lineal Múltiple	19
2.2.3. Modelo de Regresión Lineal Segmentada	19
2.2.4. R^2 o coeficiente de determinación	20
2.2.5. Diagnóstico del modelo	21
3. Resultados	23
3.1. Resultados de las revistas de Estadística	23
3.2. Resultados de las revistas de Investigación Operativa	28
3.3. Comparativa con el área de Matemática Aplicada	32
4. Conclusiones	37
A. Series Temporales	39
A.1. Estadística	39
A.1.1. Individuales	39
A.1.2. Total	41
A.2. Investigación Operativa	42
A.2.1. Individuales	42
A.2.2. Total	44

A.3. Matemática Aplicada	44
A.3.1. Individuales	44
A.3.2. Total	46
B. Resultados	47
B.1. Resultados en el área de Estadística	47
B.2. Resultados en el área de Investigación Operativa	49
B.3. Resultados de la comparativa con el área de Matemática Aplicada	51
Bibliografía	53

Resumen

El papel de las mujeres en los diferentes elementos de la sociedad actual y en particular, en la ciencia, ha tenido un fuerte impulso en los últimos años. Principalmente, esto se debe a la realización de relevantes estudios en el tema, así como la legislación de nuevas políticas de inclusión de perspectiva de género con el objetivo de alcanzar una sociedad más igualitaria en cuanto al género. En este estudio analizaremos la efectividad de tales políticas de perspectiva de género en la investigación en una disciplina matemática en expansión, como es la Estadística y el de la Investigación Operativa

Abstract

The role that women plays in the different parts of today's society and in particular, in science, has had a strong boost in recent years. Mainly, this is due to the realization of relevant studies on the subject, as well as the legislation of new policies for the inclusion of a gender perspective with the aim of achieving a more egalitarian society in terms of gender. In this study we will analyze the effectiveness of such gender perspective policies in research in an expanding mathematical discipline, such as Statistics and Operations Research.

Introducción

A pesar de que la sociedad está evolucionando continuamente, en la mayoría de sus ámbitos la igualdad de género sigue siendo una tarea pendiente. Hasta hace unos años no se contemplaba la posibilidad de que las mujeres accedieran a ciertas disciplinas, carreras y oficios. Con la ayuda de organizaciones que se han ido creando, las mujeres se han abierto paso en muchos de estos ámbitos.

Nacen así, instituciones a nivel mundial como la Comisión Jurídica y Social de la Mujer (CSW) [1]. Esta, se posiciona como la mayor organización cuyo único propósito es promover la igualdad de género y el empoderamiento de las mujeres. Se reúnen anualmente para debatir sobre lo conseguido, identificar los principales problemas y redactar tanto políticas, como leyes mundiales. Al final de dichas reuniones, publican las “conclusiones convenidas”, directrices que organismos tanto públicos como privados ponen en práctica.

También conviene mencionar a los Objetivos de Desarrollo del Milenio (ODM) y a sus sucesores, los de Desarrollo Sostenible (ODS), ambos gestados por las Naciones Unidas. Los primeros son ocho objetivos propuestos en el año 2000 cuya fecha de cumplimiento quedó fijada para 2015. Ahí, los ODS cogieron el relevo aumentando el número hasta un total de 17 objetivos cuyo final será en 2030. Ambos incluyen entre dichas metas, la igualdad de género y el empoderamiento de las mujeres, siendo el objetivo número tres en el caso de los ODM y el cinco en el caso de los ODS. Entre los avances conseguidos en las últimas décadas cabe destacar, como apuntan en [9] :

- Las leyes que se están reformulando para fomentar la igualdad de género.
- El número de niñas que están recibiendo una educación.
- El creciente número de mujeres en puestos de liderazgo.

A pesar de estos logros, sigue habiendo aspectos que mejorar como que las mujeres siguen estando infrarrepresentadas en altos cargos políticos o que su salario sigue siendo inferior al de los hombres.

Como mencionan en [8], los Objetivos de Desarrollo del Milenio consiguieron avances en muchos de los ámbitos propuestos como son la reducción de la pobreza extrema y una educación más accesible. Sin embargo otros la reducción del hambre o la mortalidad materna no se consiguieron. El esfuerzo invertido en que estos objetivos se alcanzasen ha desembocado en avances desiguales. A esto hay que sumarle que la atención puesta en las políticas necesarias para conseguir dichos objetivos no fue la suficiente.

A nivel Europeo tenemos al Instituto Europeo de la Igualdad de Género (EIGE). En el que se elaboran unas fichas personalizadas con información de cada país cuyas mediciones se hacen en base a los siguientes parámetros: poder, tiempo, conocimiento, salud, dinero y empleo. Otra de sus labores es la elaboración de un Índice de Igualdad de Género cuya escala va de 1 a 100, siendo 1 la desigualdad total y 100 la igualdad plena. Esta institución habla también de unos métodos y herramientas necesarios para que implementar la transversalización de género se convierta en una tarea menos abstracta. Pone por tanto, al alcance de todos, una serie de guías que seguir que organiza por ámbitos políticos.

No podemos obviar el papel que la pandemia de la COVID-19 ha tenido con respecto a la perspectiva de género. Lo cierto es que esta situación no solo no ha beneficiado a las mujeres y al papel de estas en la sociedad, sino que tristemente ha puesto en peligro todo lo conseguido hasta la fecha y va a dificultar esa meta que es la igualdad de género. Este efecto era de esperar y es que si antes de la pandemia la situación ya no era proclive, todos los efectos que nos ha dejado han amplificado esas carencias. Nos hemos visto obligados a quedarnos en casa durante meses y a nivel psicológico todo el mundo se ha visto afectado. Las mujeres, como parte fundamental de la sociedad, no han sido ajenas a esta situación. Por ejemplo, han aumentado los casos de mujeres que han tenido que dedicar más tiempo a cuidar de sus pequeños y mayores, teniendo que compaginar este duro trabajo no remunerado con el que algunas de ellas tenían y han conseguido mantener.

Este trabajo consta de 4 capítulos en los cuáles analizaremos el efecto de las diferentes políticas de género adoptadas desde diferentes estamentos (ver Figura 1.1), como por ejemplo en la investigación. En particular, se chequeará si se ha fomentado o no el papel de la mujer en esta tarea. Para ello, hemos realizado una labor de campo, de recogida de datos e información, y posteriormente, han sido analizados desde una perspectiva estadística. En particular, se ha hecho uso de modelos de regresión. En el Capítulo 1 introduciremos la situación ya mencionada pero centrándonos en el ámbito de las matemáticas y más aún en la investigación en Estadística e Investigación Operativa. En el Capítulo 2, presentaremos la muestra escogida y las herramientas estadísticas empleadas para el análisis. Una vez aplicados dichos métodos, se analizarán los resultados obtenidos en el Capítulo 3. Por último, en el Capítulo 4 sacaremos las conclusiones pertinentes de haber hecho el estudio.

Capítulo 1

El género en la investigación

En la investigación, la situación de las mujeres no es mejor que en el resto de los ámbitos. El objeto de este estudio, por tanto, es comprobar si las políticas implantadas a lo largo de las últimas dos décadas, algunas de las cuáles recogeremos en la Figura 1.1, han sido de ayuda a la hora de minimizar la brecha de género existente.

El desarrollo de técnicas específicas de la Estadística e Investigación Operativa en los últimos años ha hecho que el número de publicaciones en ambas áreas haya aumentado de forma considerable. Su aplicación en problemas de la vida real nos permite encontrar soluciones a nuevas cuestiones así como abordar problemáticas ya existentes desde un enfoque totalmente diferente. Entre ellas, destaca el análisis del papel que juega la mujer en la sociedad actual, como elemento clave, así como su papel en la ciencia. Los avances de la sociedad moderna y su progresiva incorporación al ámbito académico han favorecido el incremento de proyectos e investigaciones lideradas por mujeres. Este hecho se ha traducido en un aumento exponencial de las autorías de mujeres de artículos científicos a la vez que también ha cobrado interés el estudio de temáticas orientadas específicamente a la mujer. A pesar de que todos estos indicadores reflejan un aumento progresivo del papel de la mujer en la ciencia, aún están lejos de alcanzarse los niveles conseguidos por los hombres en cuanto al volumen de autorías, acceso a cátedras o liderazgo de nuevas investigaciones. Así lo reflejan los datos recogidos en artículos como [4]. Una de las partes del estudio era una encuesta en la que preguntaban a hombres y mujeres, cuántos artículos habían enviado a las revistas mejor consideradas en sus campos en los últimos 5 años. La media de los dos grupos, era de 6 artículos, lo cuál deja ver, como apuntan en [7], que el trabajo realizado está siendo el mismo por ambas partes. La diferencia radica por tanto, en la selección que hacen las revistas de los artículos que publican, y no en que las mujeres estén investigando o escribiendo menos que los hombres.

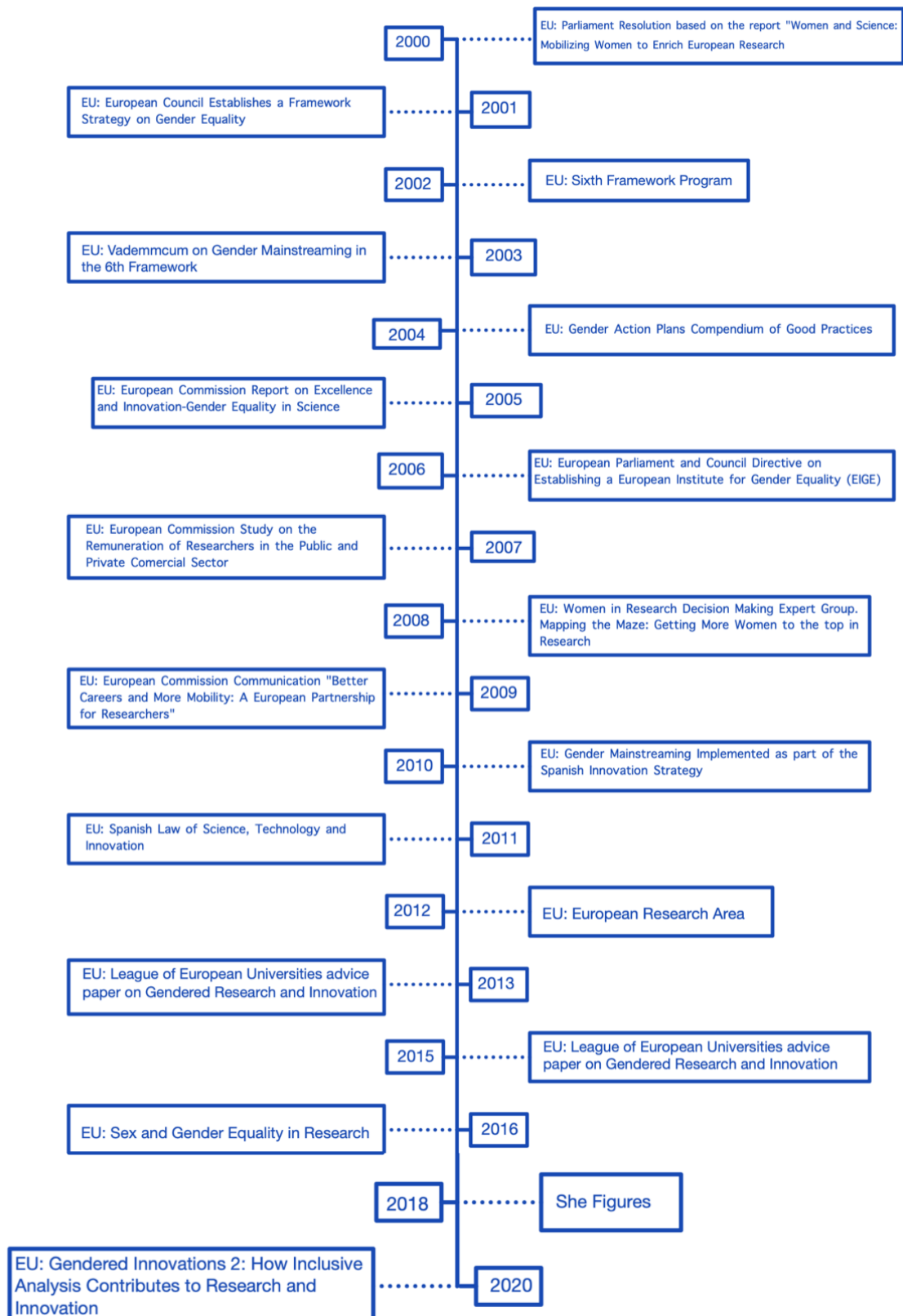


Figura 1.1: Línea Temporal de políticas recogidas en la página de Gendered Innovations [13]

Desde hace unos años hacia aquí, en España se han creado una serie de organizaciones como la Comisión de la Mujer en la Estadística y en la Investigación Operativa (CMEIO) o la Comisión de Mujeres y Matemáticas, la primera de ellas en Febrero de 2010 por el consejo ejecutivo de la Sociedad de Estadística e Investigación Operativa (SEIO) y la segunda por la Real Sociedad Matemática Española (RSME). Presentan objetivos y propósitos muy similares entre los cuáles se encuentran:

- El análisis de la situación actual a la que se enfrentan tanto las niñas en enseñanzas medias como las mujeres universitarias.
- La elaboración de guías sobre estereotipos y sesgos inconscientes entre el profesorado.
- Planes de visibilización de mujeres matemáticas ya consolidadas.

Entre estos objetivos destacamos el que propone la Comisión de Mujeres y Matemáticas para que en las webs de las revistas aparezcan las estadísticas de artículos presentados y aceptados desgregados por sexo, permitiendo así el análisis del sesgo existente.

De forma paralela, como se muestra en la Figura 1.1, se ha creado legislación específica en materia de género en la ciencia. Por ejemplo, la Ley de la Ciencia de 2011 a la cuál hacemos referencia en dicha figura, se habla de nuevos métodos y criterios en los que la persona evaluadora no conozca ningún aspecto personal de la evaluada. Se pretende así, intentar evitar que consciente o inconscientemente, la elección de personal de los organismos públicos se vea influenciada por condiciones ya sean sociales o personales como el género. Se pretende a su vez no considerar negativamente el tiempo que los candidatos hayan estado “ausentes”, ya sea por maternidad, paternidad, incapacidad temporal, etc. Para conseguir esto, lo que se pide es que las personas que evalúan, sean expertas en género o profesionales sociales.

Otro de los ejemplos es el She Figures [3]. En este documento se recogen datos como las todavía existentes diferencias entre las distintas áreas de estudio. Habiendo un predominio de los hombres en los masters y doctorados de las ciencias naturales y la tecnología, mientras que las mujeres son más numerosas en las ciencias de la salud. Los estudios realizados al respecto, muestran que este hecho es resultado de los estereotipos, así como de las divisiones de trabajo por género, las limitaciones de tiempo, el sesgo en las prácticas o la elección del campo de estudio entre otras.

Como apuntan desde el *European Research Area and Innovation Committee*, muchos estudios realizados hasta la fecha, muestran una notable predilección por los hombres a la hora de contratar o elegir al personal científico en cualquier proyecto. El trabajo de estos es considerado mejor que el de sus compañeras femeninas por su simple condición de hombre.

Estos hechos deberían penalizarse e intentar corregirse por parte de toda la sociedad, ya que este prejuicio no existe solo en la cabeza de los hombres, sino por desgracia, también en el de muchas mujeres. Las mujeres necesitan por tanto, esforzarse más y conseguir mayores méritos, y aún así ante un hombre siguen teniendo elevadas posibilidades de salir perdiendo. A esto, debemos sumarle otras injusticias como la distorsionada visión en cuanto a puestos de liderazgo. En estos, los hombres son indudablemente los favoritos, tachando a las mujeres que reúnen las condiciones adecuadas para dichos puestos. Además las mujeres no suelen ser invitadas ni a tantos congresos, ni a formar parte de comités científicos o equipos de investigación. La Comisión de Mujeres y Matemáticas en [7] recoge una serie de propuestas cara a las futuras convocatorias de financiación, y lo hablado en este párrafo.

El *Standing Working Group on Gender in Research and Innovation* hizo una serie de recomendaciones que recogen en [2] con el propósito de asegurar que los investigadores tengan las mismas oportunidades en el momento de ser evaluados. Dichas recomendaciones van dirigidas tanto a las *Research Funding Organisations* como a las autoridades nacionales. Esperan así que el proceso sea justo, transparente e imparcial y que los resultados no se vean comprometidos por condiciones sociales como el género de los investigadores. Algunas de estas recomendaciones son:

- Las RFOs, deberían llevar a cabo y posteriormente hacer públicas, las estadísticas sobre la proporción de investigadores por género que se encuentran entre los solicitantes y beneficiarios de subvenciones.
- La formación en materia de perspectiva de género debería estar entre los requisitos de evaluadores y personal.
- Se deberían incluir expertos de género en los paneles de evaluación con el objetivo de asegurar que el tema de perspectiva de género es correctamente evaluado en los proyectos.
- Los criterios de evaluación deben estar claros y al alcance de todo el mundo, teniendo en cuenta las posibles interrupciones de los evaluados a lo largo de su carrera.
- En la medida de lo posible, se optará por la revisión doble-ciego debido a sus ventajas con respecto a los posibles prejuicios de género, aumentando las posibilidades de que las mujeres reciban subvenciones.

Paralelamente a lo mencionado hasta ahora, es relevante hablar sobre el lenguaje sexista en el que las investigadoras encuentran dificultades a la hora de rellenar documentos pensados para hombres. Muchas veces no solo son invisibilizadas por el lenguaje sino que

directamente sus compañeros hombres se llevan gran parte, o incluso todo el reconocimiento. Este es el conocido como efecto Matilda, concepto definido inicialmente por Matilda Joslyn Gage en su ensayo La mujer inventora, y descrito por Rossiter [11].

Finalmente, no deberíamos descuidar el papel que juegan las universidades y los principales centros de investigación en este tipo de competencias transversales. Debería por tanto, ser objeto de estudio el hecho de incluirlas, ya que los futuros profesionales que van a acceder a diversos trabajos y puestos dentro de la sociedad, es ahí donde se forman no solo a nivel profesional sino a nivel personal. La Universidad, como motor de la sociedad en general, no puede presentarse exclusivamente, como una entidad de formación, sino como una entidad que eduque en valores a mujeres y hombres que busquen la igualdad real entre ambos sexos en la sociedad. En este trabajo, se pretende comprobar la efectividad de las diferentes políticas de género adoptadas desde el ámbito político o académico hasta el momento, en aras de reducir la brecha de género existente entre hombres y mujeres en la ciencia. Tal y como se ha mencionado, el predominio del género masculino sobre el femenino queda patente, aunque en los últimos años dicha distancia se haya reducido. Con ello, pretendemos evaluar las debilidades de las medidas adoptadas en el contexto de la perspectiva de género con el objetivo de que, en el futuro, estas, realmente favorezcan a la mujer.

Capítulo 2

Datos y metodología

En este capítulo hablaremos de los datos escogidos para esta investigación así como de las herramientas estadísticas empleadas para el análisis que vamos a desempeñar. En la Sección 2.1 veremos los datos sobre los que vamos a trabajar y el criterio de selección de la muestra escogida. A su vez, introduciremos las series temporales de las revistas haciendo una pequeña parada en algunas de las que más han llamado nuestra atención. En la Sección 2.2 veremos los modelos de regresión utilizados. La regresión es muy útil en este tipo de estudios por brindarnos la posibilidad de ver como varía el número de publicaciones a lo largo de los años. También nos ayuda a predecir como se comportará en el futuro y ver cuánto de útiles están siendo las medidas tomadas para acabar con las diferencias de género en la investigación.

2.1. Recogida de datos

Para este estudio, hemos acudido a los buscadores de las revistas previamente seleccionadas introduciendo en ellos las palabras claves mencionadas en la siguiente tabla:

Gender	Female and Male	Woman and Man	Gender and Perspective
---------------	------------------------	----------------------	-------------------------------

que fueron cuidadosamente seleccionadas como indicadores de la inclusión de perspectiva de género.

Mostraremos en las Tablas 2.1 y 2.2 las listas de revistas seleccionadas, las pertenecientes al campo de la Estadística y las de Investigación Operativa, respectivamente. En concreto, hemos seleccionado una muestra de las revistas científicas más representativas en ambas áreas. El criterio de elección de las revistas se basa en el impacto de los artículos que contiene dicha área, en términos de las citas que reciben. Las revistas elegidas, han sido

por tanto, revistas de InCites Journal Citation Reports, con distintos factores de impacto.

El factor de impacto de una revista mide la importancia y repercusión de una revista en su campo. Por ser de los índices bibliométricos más usados, apuntaremos algunos datos que el lector debería conocer. Este valor se obtiene al dividir el número de citas recibidas en un año, de artículos publicados los dos años anteriores, entre el número total de publicaciones en esos dos años. Ejemplifiquemos dicho cálculo con la revista *Extremes*, perteneciente al área de Estadística, y el año más reciente disponible en el Journal Citation Report, que es el 2019:

$$JIF = \frac{\text{Citas en 2019 a publicaciones en 2017(47) + 2018(38)}}{\text{Total de artculos citados en 2017(34) + 2018(32)}}$$

Algunos de los aspectos influyentes a la hora de calcular dicho factor son:

- No se consideran las autocitas, ni las citas negativas
- Las áreas que mayor popularidad tienen en el momento, son las que mayor posibilidades de tener un factor de impacto más elevado tienen. En la actualidad, podríamos decir por tanto, que las revistas de Medicina que incluyan estudios sobre la COVID-19 probablemente sean las que mayor factor de impacto tengan.
- No tiene sentido comparar factores de impacto de distintas áreas temáticas, pero también hay que tener en cuenta que una revista puede estar asignada a más de un área.
- Obviamente lo óptimo es estar dentro del primer cuartil en el ranking.

De esta forma, las revistas pertenecientes al área de Estadística son las siguientes:

Revistas Estadística

Journal of Statistical Software (JSS)

Computational Statistics and Data Analysis (CSDA)

Journal of Machine Learning Research (JMLR)

Quality Technology and Quantitative Management (QTQM)

Biostatistics (B)

Journal of Statistical Planning and Inference (JSPI)

Test (T)

Econometrica (E)

Chemometrics and Intelligent Laboratory Systems (CILS)

Statistical Methods in Medical Research (SMMR)

Finance and Stochastics (FS)

Journal of Business and Economic Statistics (JBES)
 Statistical Analysis and Data Mining (SADM)
 Pharmaceutical Statistics (PS)
 Insurance: Mathematics and Economics (IME)
 Stata Journal (SJ)
 Extremes (EX)
 Statistica Neerlandica (SN)
 Astin Bulletin (AB)
 Lifetime Data Analysis (LDA)
 Statistical Modelling (SM)
 Fuzzy Sets and Systems (FSS)
 Spatial Statistics (SS)
 Stochastic Environmental Research and Risk Assessment (SERRA)
 Advances in Data Analysis and Classification (ADAC)

Tabla 2.1: Lista de revistas de Estadística

La siguiente lista corresponde a las revistas del área de Investigación Operativa:

Revistas Investigación Operativa

Annals of Operation Research (AOR)
 European Journal of Operational Research (EJOR)
 International Journal of Game Theory (IJGT)
 Omega (O)
 Journal of Optimization Theory and Applications (JOTA)
 Technovation (T)
 International Journal of Production Economics (IJPE)
 Socio-Economic Planning Sciences (SEPS)
 International Transactions in Operational Research (ITOR)
 Journal of Scheduling (JS)
 Optimization Methods and Software (OMS)
 OR Spectrum (ORS)
 Applied Stochastic Models in Business and Industry (ASMBI)
 Engineering Economist (EE)
 Networks and Spatial Economics (NSE)
 Memetic Computing (MC)

Mathematical Programming(MP)
 Fuzzy Optimation and Decision Making (FODM)
 Decision Support Systems (DSS)
 Expert Systems with Applications (ESA)
 Flexible Services and Manufacturing Journal (FSMJ)
 Queueing Systems (QS)
 Journal of Simulation (JOS)
 Production Planning and Control (PPC)
 International Journal of Information Technology and Decision Making (IJITDM)

Tabla 2.2: Lista de revistas de Investigación Operativa

Una vez decidido el listado definitivo, el siguiente paso en este estudio es hacer un recuento del número de publicaciones por año que contiene al menos una de las palabras clave para cada una de las revistas. La representación de las series temporales resultantes se muestran en las Figuras 2.1 y 2.3. Para su obtención se ha utilizado el código R creado específicamente y que se detalla en el Apéndice A.

Las 25 primeras, que aparecen en la Figura 2.1, corresponden a las revistas del área de Estadística. Las hemos posicionado según el orden seguido previamente en la lista. También haremos un breve comentario sobre algunas de las que más nos han llamado la atención. Estas aparecerán rodeadas con recuadros azules como se puede observar en la Figura 2.2.

Conviene mencionar que para este estudio, hemos considerado las publicaciones que contenían al menos una de las palabras clave desde el 2000 hasta el 2021. Debido a causas obvias, los datos del 2021 solo contemplan las publicaciones hasta el mes de Abril.

En la Figura 2.2, las series temporales de la parte superior, de izquierda a derecha corresponden a las revistas *Journal of Statistical Software* y *Biostatistics*. Sobre ellas conviene señalar que son las únicas de la muestra elegida para este área, que son prácticamente nulas. Y las dos inferiores, nuevamente de izquierda a derecha pertenecen a las revistas *Journal of Business and Economic Statistics* y *Pharmaceutical Statistics*. En este caso se observa una tendencia creciente.

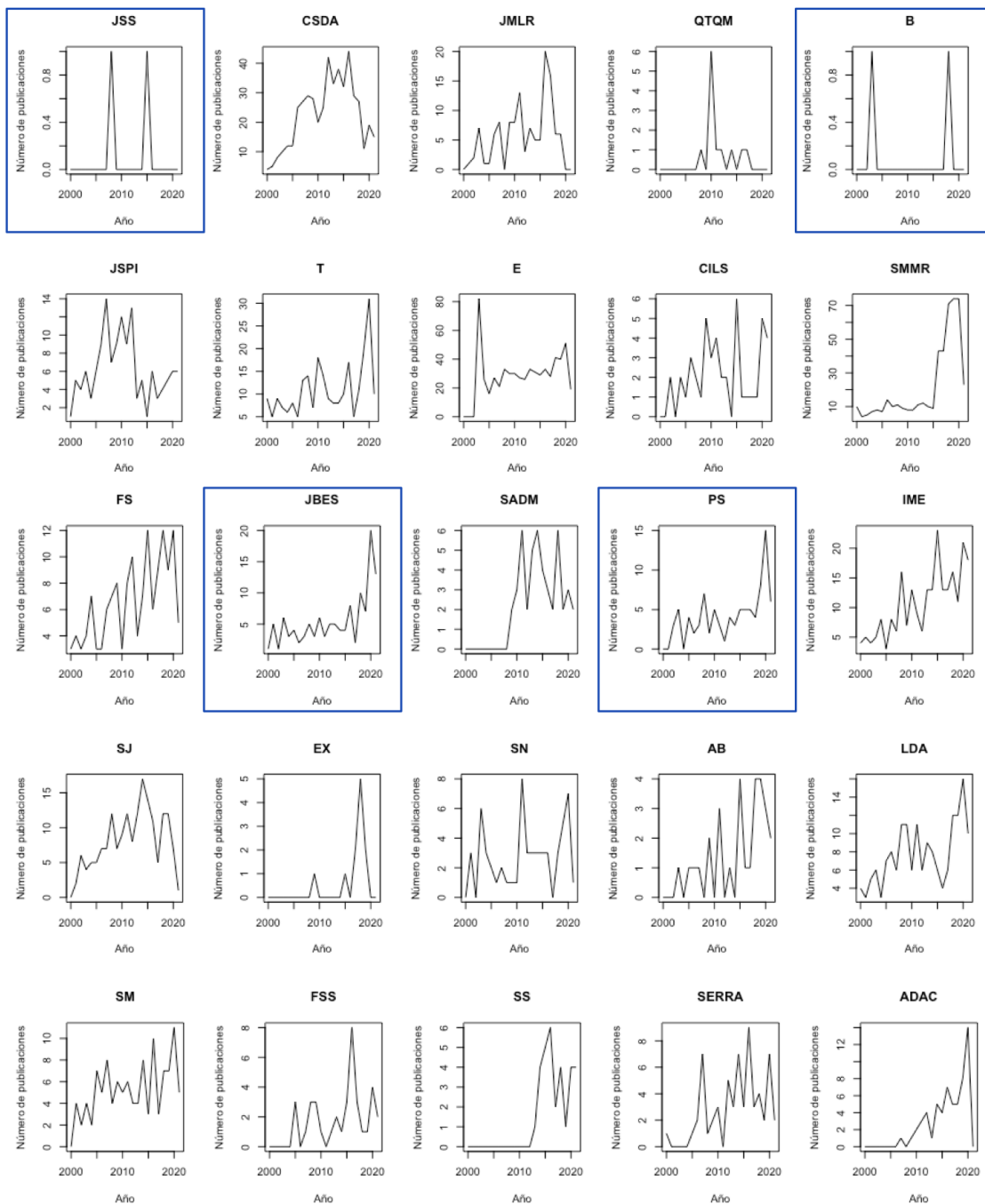


Figura 2.1: Series Temporales de las revistas de Estadística

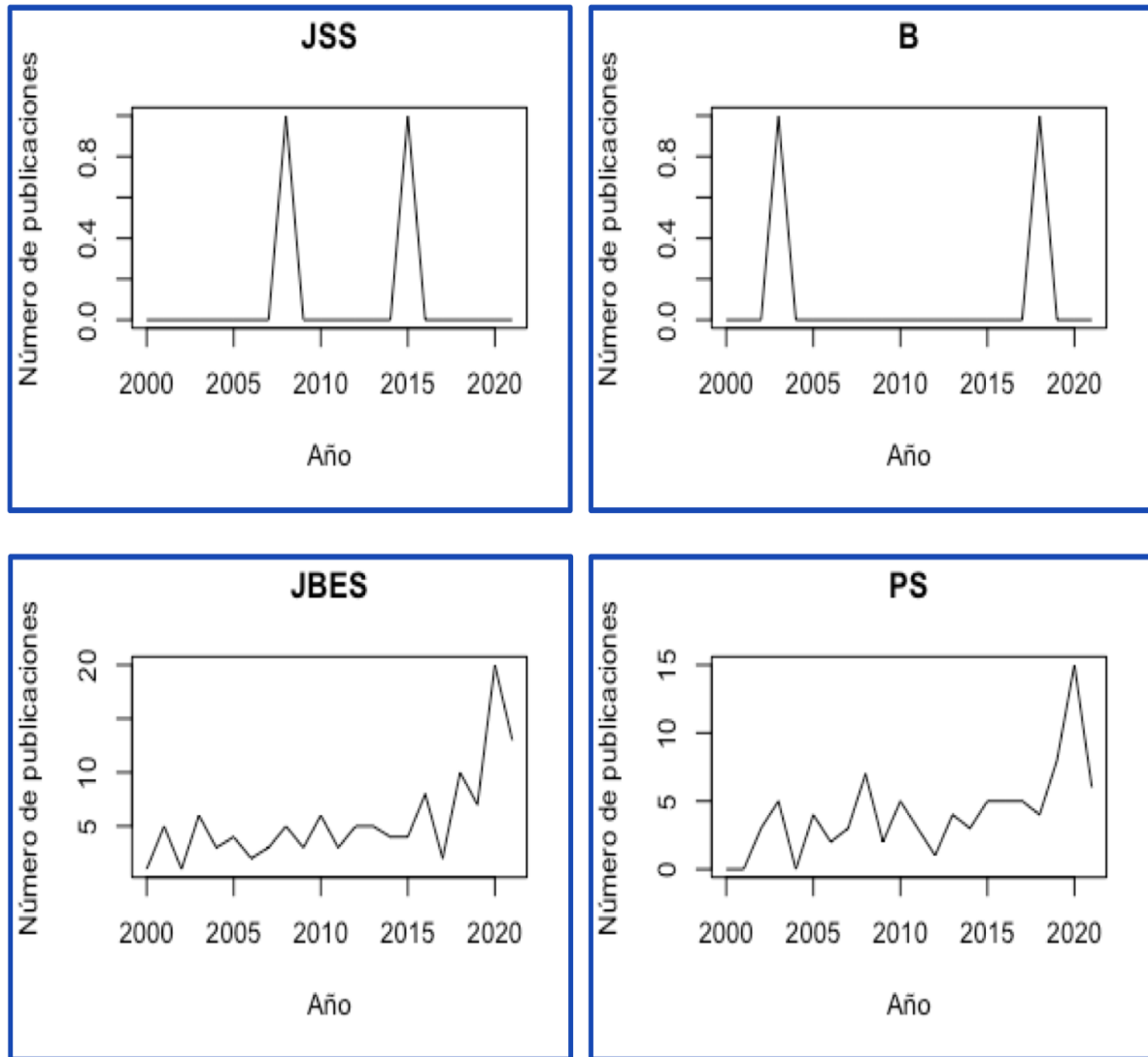


Figura 2.2: Series temporales pertenecientes a las revistas: Journal of Statistical Software (JSS), Biostatistics (B), Journal of Business and Economic Statistics (JBES), Pharmaceutical Statistics (PH)

En la Figura 2.3 se muestran las series temporales de las 25 revistas del área de Investigación Operativa. Marcamos nuevamente con un recuadro azul algunas de las más representativas a las que posteriormente les haremos un zoom.

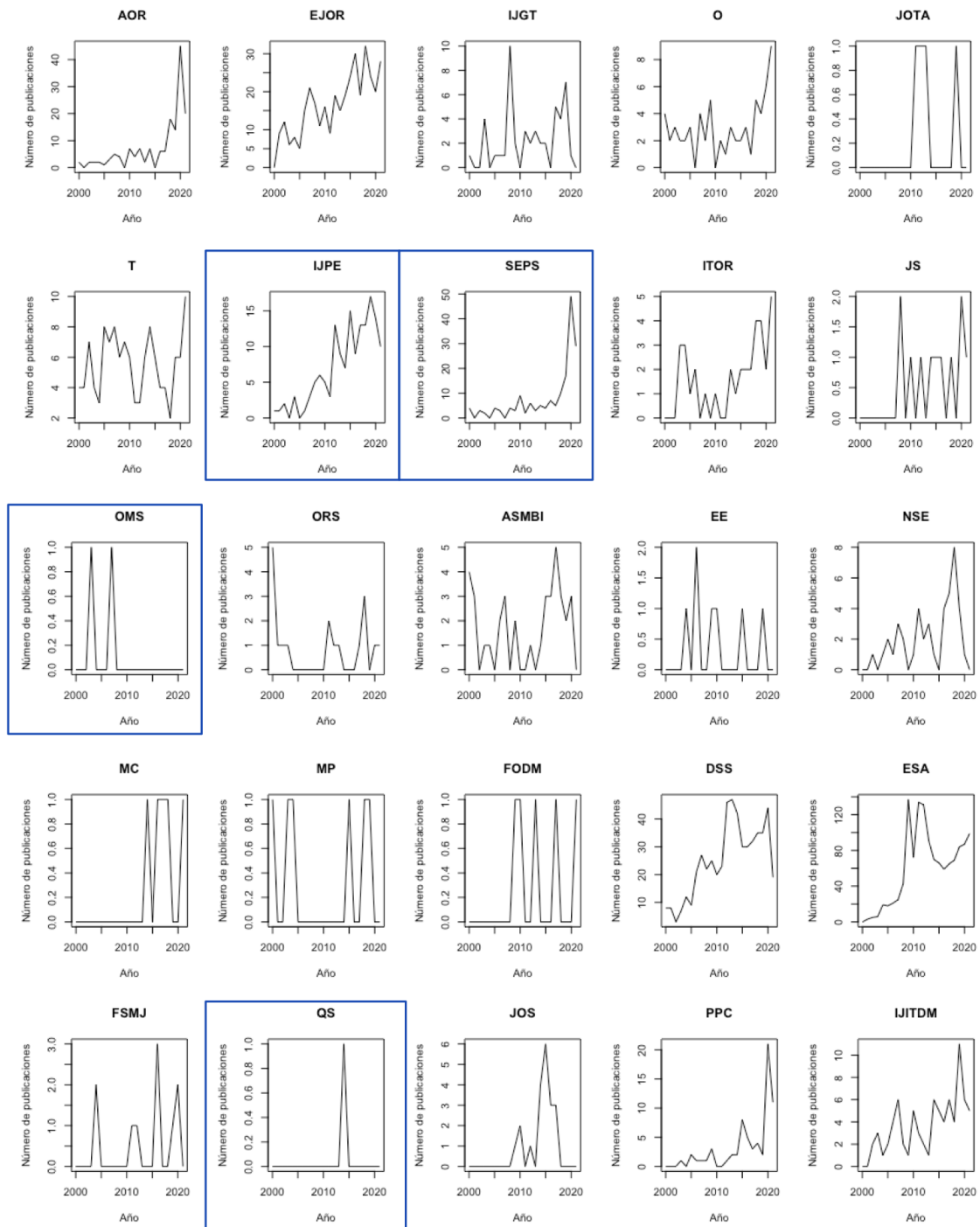


Figura 2.3: Series Temporales de las revistas de Investigación Operativa

En la Figura 2.4, correspondiente a dicho zoom, podemos observar que las series temporales superiores, son unas de las que mayor tendencia creciente presentan. Estas gráficas pertenecen a las revistas *International Journal of Production Economics* y *Socio-Economic Planning Sciences*, de izquierda a derecha. Por el contrario, las series temporales inferiores que corresponden a las revistas *Optimization Methods and Software* y *Queueing Systems*, son prácticamente nulas.

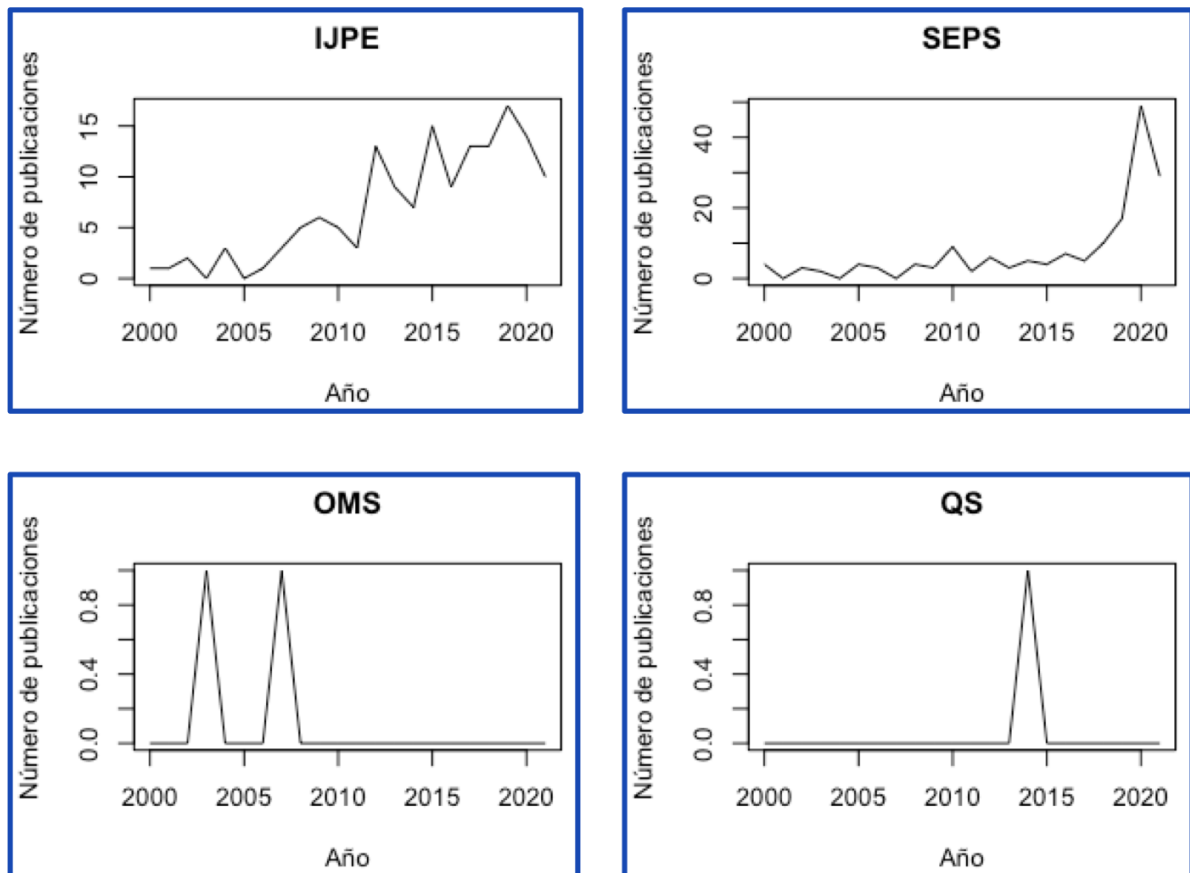


Figura 2.4: Series temporales pertenecientes a las revistas: *International Journal of Production Economics* (IJPE), *Socio-Economic Planning Sciences* (SEPS), *Optimization Methods and Software* (OMS), *Queueing Systems* (QS)

Finalmente, tenemos las series temporales del total de publicaciones que contienen al menos una de las palabras clave. La Figura 2.5 se corresponde con la del área de Estadística y la de la Figura 2.6 con la de Investigación Operativa. Aunque en las

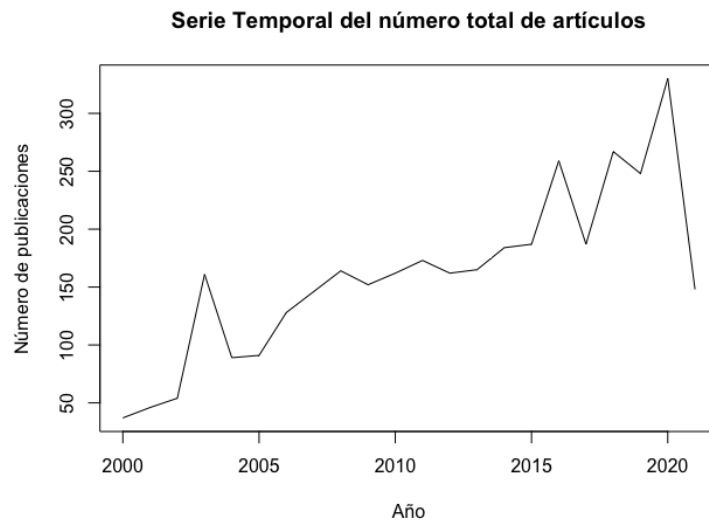


Figura 2.5: Serie Temporal del número total de publicaciones que contienen al menos una de las palabras clave del área de Estadística

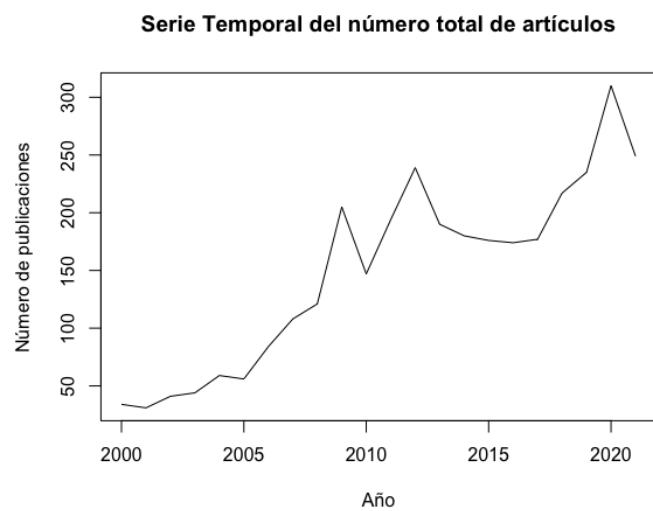


Figura 2.6: Serie Temporal del número total de publicaciones que contienen al menos una de las palabras clave del área de Investigación Operativa

2.2. Metodología

Para este estudio emplearemos modelos de regresión. El propósito de estos es explicar la dependencia existente entre un conjunto de variables respuesta (o dependientes) y un conjunto de variables explicativas (o independientes). Además de tener en cuenta tanto posibles errores de medida como la aleatoriedad de las variables.

En esta sección del capítulo haremos un pequeño recorrido por los modelos de regresión que utilizaremos para el presente estudio. Empezaremos por el más básico de todos, el modelo de regresión lineal simple (Sección 2.2.1), donde también hablaremos de la inferencia sobre los parámetros. Haremos una parada en el modelo lineal múltiple (Sección 2.2.2) y acabaremos con el modelo de regresión lineal segmentada (Sección 2.2.3). Finalmente, comentaremos la importancia del R^2 (Sección 2.2.4) y de la diagnosis del modelo (Sección 2.2.5).

2.2.1. Modelo de Regresión Lineal Simple

En el caso del modelo lineal simple, trataremos de explicar la relación que existe entre la variable respuesta y una única variable explicativa. Formalmente, esto se puede escribir:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

donde Y será nuestra variable respuesta, X nuestra variable explicativa, n el número de observaciones de Y y X , β_0 el intercepto y β_1 la pendiente de la recta de regresión que explica la tasa de crecimiento de la respuesta con respecto a la variable explicativa.

La recta de regresión se va a usar para predecir el valor de la variable respuesta a partir de un valor de la explicativa. Tendremos por tanto, que si consideramos $\hat{\beta}_0$ y $\hat{\beta}_1$, obtenemos la predicción $\hat{\beta}_0 + \hat{\beta}_1 x$ de la variable respuesta. En términos de los datos muestrales tenemos la predicción $\hat{\beta}_0 + \hat{\beta}_1 x_i$ para la observación x_i . Obtenemos por tanto, como errores de predicción:

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad \text{para } i \in \{1, \dots, n\} \quad (2.2)$$

que son los residuos de la regresión. Una de las formas de estimar los parámetros del modelo 2.1 es el empleo de mínimos cuadrados. Lo que queremos es coger los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ que den lugar a los residuos más pequeños. De aquí sacamos los valores que minimizan la suma de los residuos al cuadrado, que viene dada por la expresión $\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \sum_{i=1}^n \varepsilon_i^2$.

Inferencia sobre los parámetros

En este apartado hablaremos de los intervalos de confianza, que nos dan una medida de la precisión de las estimaciones, y de los contrastes de hipótesis, que nos indican si los valores obtenidos pueden corresponder a los valores de los parámetros.

- **Inferencia sobre β_0**

Tanto $\hat{\beta}_0$ como $\hat{\beta}_1$, son estimadores insesgados y simétricos. Empezaremos hablando de $\hat{\beta}_0$.

Como pivote para la construcción de intervalos de confianza o para la realización de contrastes de hipótesis, podemos hacer uso de la estandarización de $\hat{\beta}_0$:

$$\frac{\hat{\beta}_0 - \beta_0}{\sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{nS_x^2}}} \in N(0, 1)$$

El problema de esto es que tendríamos que tener el valor de σ^2 y no lo sabemos. Para subsanar este inconveniente no hay más que estimar el error típico, obteniendo:

$$\widehat{\text{Error Típico}}(\hat{\beta}_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{nS_x^2}}$$

Y así, utilizamos el pivote:

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{nS_x^2}}} \in T_{n-2}$$

Cambiando la distribución normal estándar por una T de Student con $n-2$ grados de libertad. Lo que nos indica este pivote es que el intervalo de confianza para β_0 con nivel de confianza $(1 - \alpha)$, está centrado en $\hat{\beta}_0$, y su radio será el producto del cuantil de la T de Student de orden $1 - \alpha/2$ con $n-2$ grados de libertad, $t_{n-2, \alpha/2}$

Para el contraste de hipótesis en cuanto a β_0 :

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_a : \beta_0 \neq 0 \end{cases}$$

La región de rechazo de H_0 es: Y así, utilizamos el pivote:

$$\frac{\hat{\beta}_0}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{nS_x^2}}} < -t_{n-2, \alpha}$$

■ **Inferencia sobre β_1**

Al igual que nos pasaba con el intercepto, tenemos dos opciones de pivote para la pendiente:

- si σ es conocida:

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma/(S_x\sqrt{n})} \in N(0, 1)$$

- si σ es desconocida:

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/(S_x\sqrt{n})} \in T_{n-2}$$

El intervalo de confianza para β_1 , con nivel de confianza $(1 - \alpha)$, es:

$$\left(\hat{\beta}_1 - t_{n-2, \alpha/2} \frac{\hat{\sigma}}{S_x\sqrt{n}}, \hat{\beta}_1 + t_{n-2, \alpha/2} \frac{\hat{\sigma}}{S_x\sqrt{n}} \right)$$

En cuánto al contraste, tendríamos:

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{cases}$$

Donde la región de rechazo de H_0 es:

$$\frac{|\hat{\beta}_1|}{\hat{\sigma}/(S_x\sqrt{n})} > T_{n-2}$$

■ **Inferencia sobre σ^2**

En el caso de la varianza del error, tenemos que el pivote sería:

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \in \chi_{n-2}^2$$

lo cuál nos da un intervalo de confianza para la varianza del error, con nivel de confianza $(1 - \alpha)$, de:

$$\left(\frac{(n-2)\hat{\sigma}^2}{\chi_{n-2, \alpha/2}^2}, \frac{(n-2)\hat{\sigma}^2}{\chi_{n-2, 1-\alpha/2}^2} \right)$$

A diferencia de los dos casos anteriores, el estimador $\hat{\sigma}^2$ no es simétrico en torno a σ^2 . De hecho, los intervalos tampoco son simétricos en torno al estimador. En cuanto al contraste:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_a : \sigma^2 \neq \sigma_0^2 \end{cases}$$

2.2.2. Modelo de Regresión Lineal Múltiple

Visto el modelo lineal simple, conviene hablar sobre el modelo lineal múltiple, que no es más que una extensión del simple. En este caso, en vez de una única variable explicativa X , tendremos varias variables explicativas X_1, X_2, \dots, X_{p-1} . Obtendremos dicho modelo como resultado de hacer una combinación lineal de las variables explicativas:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + \varepsilon \quad (2.3)$$

Donde $\beta_0, \beta_1, \dots, \beta_{p-1}$ son los coeficientes que acompañan a las variables y ε el error. La razón por la cual las variables explicativas van de 1 a $(p-1)$, es para obtener un total de p coeficientes, entre los cuales se encuentra el intercepto, β_0 , y los $(p-1)$ restantes que se corresponden con las respectivas $(p-1)$ variables explicativas. Al igual que pasa con el modelo simple, en el caso del modelo múltiple podemos considerar un diseño fijo o un diseño aleatorio de las variables explicativas. Una muestra bajo diseño fijo de este modelo se puede expresar de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \quad (2.4)$$

donde Y_i es la variable respuesta del i -ésimo individuo, $x_{i,1}, \dots, x_{i,p-1}$ las variables explicativas del mismo y ε_i el error asociado a dicho individuo. En el modelo lineal simple, había varias suposiciones sobre el error. Las hipótesis de homocedasticidad, normalidad e independencia, también se supondrán para este modelo, de forma que: $\varepsilon_1, \dots, \varepsilon_n \in N(0, \sigma^2)$ y son independientes. Teniendo que la varianza de los errores es denotada por σ^2 y será un parámetro a estimar.

2.2.3. Modelo de Regresión Lineal Segmentada

Como apuntan en [12], el problema que nos presenta el modelo simple en (2.1) es que asume que nuestra pendiente no depende de la variable explicativa X , es decir, es constante, lo cual no se suele corresponder con la realidad. Esto nos lleva a considerar un modelo alternativo donde si se considere una posible variación de la pendiente.

Entran aquí los modelos de regresión segmentada (o regresión por partes), se caracterizan por la existencia de un valor en la variable explicativa X al que denominaremos break-point o punto de cambio. Este punto, implica un cambio en los parámetros de tendencia o pendiente para la función de regresión. En particular, emplearemos los modelos de regresión segmentada lineal, en los que la relación entre la variable respuesta y la variable explicativa dentro de los distintos intervalos (o segmentos), se obtiene por regresión simple.

Escribiremos este último modelo de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_\psi (X_i - \psi)_+ + \varepsilon_i, \quad i = 1, \dots, n \quad (2.5)$$

donde $(X_i - \psi)_+ = (X_i - \psi)I(X_i > \psi)$ siendo $I(\cdot)$ la función indicadora. Aparecen como resultado aquí, dos partes, que de hecho son dos rectas, una antes del break-point ψ y otra después del mismo. Tenemos ahora que β_1 es la pendiente del modelo cuando $X_i \leq \psi$ y $\beta_1 + \beta_\psi$ es la pendiente cuando $X_i > \psi$. A su vez, podríamos decir que β_ψ es la diferencia entre la pendiente del primer modelo y la del segundo. En otro caso estaríamos ante dos modelos iguales y es el hecho de que existe un break-point, es decir, $\psi \neq 0$.

Con respecto al modelo 2.5, la literatura propone varios métodos de estimación. Uno de ellos asume la existencia de un punto de cambio y emplea mínimos cuadrados. Hinkley (1969) desarrolla un método en el que se intenta encontrar el intervalo que maximiza la función de verosimilitud del modelo y luego determina el punto de corte de dos rectas dentro del intervalo. Otro posible método es el que minimiza la suma de residuos al cuadrado. Pero en 2003, como podemos observar en [5], Muggeo propuso un algoritmo basado en las aproximaciones de Taylor para estimar los parámetros del modelo incluyendo uno o de forma general varios puntos de cambio. Aquí nos puede surgir un pequeño contratiempo y es que estamos asumiendo que existe al menos un break-point y este hecho no siempre está garantizado. Necesitamos así un test previo que nos asegure dicha existencia. Podemos emplear por ejemplo los test de Davies o Score cuya hipótesis nula para el modelo 2.5 puede ser escrita como $H_0 = \beta_\psi = 0$ donde $\beta_\psi = 0$ indica que no hay ningún break-point.

Llegados a este punto cabe destacar que al haber reunido los datos a lo largo del tiempo, podemos encontrar ciertas dificultades al estimar los modelos mencionados por la existencia de correlaciones temporales. Nace de aquí la necesidad de comprobar si esto ocurre, para impedir que nuestras conclusiones del ajuste centrado en observaciones independientes no se vean afectadas. Tenemos para ello herramientas como el test de Breusch-Godfrey que testea la correlación serial de orden superior.

2.2.4. R^2 o coeficiente de determinación

Vamos a hablar del coeficiente de determinación ya que podemos tomarlo para medir como de bueno es el ajuste de regresión. Asociando que este sea alto con una buena regresión, pero teniendo en cuenta que un bajo valor del mismo no supone que nuestro modelo no sea válido. Esto viene de que aunque se cumplan las suposiciones del modelo, este, puede tener una mucha varianza en la variable de error y hacer así que nuestro coeficiente de determinación sea pequeño. Lo que si podremos decir en este caso, es que

aunque el modelo lo tenga por que ser malo o incorrecto, va a ser menos útil que otro que presente un coeficiente de determinación más elevado.

Se representa por R^2 y se define como la proporción de varianza explicada de un modelo. Se puede calcular con la expresión:

$$R^2 = 1 - \frac{RSS}{TSS} \quad (2.6)$$

donde RSS es la suma residual de cuadrados y viene dada por:

$$RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.7)$$

siendo \hat{Y}_i las predicciones en base al modelo. Y donde TSS es la suma total de cuadrados y se calcula:

$$TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (2.8)$$

Podemos observar por tanto que su valor oscila entre cero y uno, informando de una mayor proximidad entre las observaciones y la recta ajustada, cuanto más próximo a 1 esté su valor.

2.2.5. Diagnósis del modelo

En este apartado haremos un repaso por las hipótesis que debe cumplir un modelo de regresión y como ver si se cumplen. Dichas hipótesis son las de linealidad, homocedasticidad, normalidad e independencia. Aunque las tres últimas hipótesis se refieren a los errores de la regresión, como no los observamos, los aproximamos a través de los residuos. Tenemos así que lo importante está en el cómputo global del conjunto de residuos y no en que haya una correspondencia exacta entre los mismos y sus respectivos errores. La hipótesis de linealidad no está relacionada con los errores, pero los residuos si que nos permiten ver si se cumple o no.

Hipótesis de Linealidad

En el caso del modelo lineal simple, para que se cumpla este supuesto, los residuos no deben presentar un tendencia respecto de X, es decir, la función de regresión es una línea recta. Para estudiar el cumplimiento de esta hipótesis emplearemos tests como el de Ramsey o el de Harvey-Collier. También nos será útil, en el caso simple, hacer un gráfico de dispersión.

Hipótesis de Homocedasticidad

Que se cumpla el supuesto de homocedasticidad quiere decir que la varianza del error se mantenga constante para todo el recorrido de la variable explicativa. Para comprobar su cumplimiento también tenemos varios tests como el de Breusch-Pagan o el de Harrison-McCabe. En cuanto gráficos, en el caso simple, podemos representar los residuos frente a la explicativa, y en el caso múltiple, un gráfico de los residuos frente a los ajustes.

Hipótesis de Normalidad

Para ver si se cumple la hipótesis de normalidad tenemos varios tests entre los que se encuentran el de Kolmogorov-Smirnov, el test de Shaphiro-Wilk o un test ji-cuadrado. En este caso, gráficamente nos podemos apoyar en un histograma o un gráfico QQ.

Hipótesis de Independencia

Esta suposición, nos dice que los n errores obtenidos de las n observaciones bajo el modelo de regresión, son independientes entre sí.

Capítulo 3

Resultados

En este capítulo recogeremos y analizaremos los resultados obtenidos del estudio. Ajustaremos el modelo de regresión lineal simple, haremos las transformaciones pertinentes y en caso de que sea necesario, segmentaremos el modelo. Este capítulo consta de tres secciones, la Sección 3.1 en el cuál realizaremos lo mencionado para el área de Estadística, acompañándonos de gráficas y tablas donde visualizar los resultados para un análisis más cómodo y adecuado de los mismos. En la Sección 3.2, se llevará a cabo un procedimiento similar pero para el correspondiente campo, que es el de Investigación Operativa. Finalmente, en la Sección 3.3 presentamos un estudio comparativo con el área de Matemática Aplicada, en el cuál, aunque no con tanta profundidad, veremos en que posición se encuentra esta disciplina con respecto a la cuestión de género.

3.1. Resultados de las revistas de Estadística

Antes de profundizar un poco más, hagamos un análisis descriptivo de los datos de las revistas escogidas para el área de Estadística.

Revistas	Mínimo	Primer Cuartil	Mediana	Media	Tercer Cuartil	Máximo
JSS	0	0	0	0.9	0	1
CSDA	4	12	25	22.5	29	44
JMLR	0	1	5.5	5.6	7.8	20
QTQM	0	0	0	0.55	1	6
B	0	0	0	0.1	0	1
JSPI	1	4	6	6.2	8.5	14
T	5	7.3	9	11.1	13.75	31
E	0	22.3	28.5	28.3	33	82

CILS	0	1	2	2.1	3	6
SMMR	4	8	10	21.4	20.8	74
FS	3	4	6.5	6.6	8.8	12
JBES	1	3	4.5	5.5	6	20
SADM	0	0	2	2.1	3	6
PS	0	2.3	4	4.1	5	15
IME	3	6	10	10.7	13	23
SJ	0	5	7	7.96	12	17
EX	0	0	0	0.5	0	5
SN	0	1	3	2.68	3	8
AB	0	0	1	1.32	2	4
LDA	3	6	6.5	7.73	10.75	16
SM	0	4	5	5.23	7	11
FSS	0	0	1	1.68	3	8
SS	0	0	0	1.41	3.5	6
SERRA	0	1	2	2.82	3.75	9
ADAC	0	0	1	2.73	4.75	14

Tabla 3.1: Análisis descriptivo del área de Estadística

Sobre los resultados recogidos en la Tabla 3.1 destacaremos los correspondientes a la revista *Journal of Statistical Software*. Su factor de impacto es de 13.642, y en las dos categorías que aparece (estadística y probabilidad y *computer science, interdisciplinary applications*) está dentro del primer cuartil. Sus resultados dejan bastante que desear, ya que que una revista tan influyente tenga una media de 0.9 artículos en los cuáles se incluye la perspectiva de género no es demasiado alentador. En *Computational Statistics and Data Analysis* tienen una media de 22.5 artículos que incluyen al menos una de las palabras clave. El caso de esta revista se aleja bastante del de la anterior, ya que su factor de impacto es de 1.186 y en las categorías en las que aparece, el mejor cuartil conseguido es el Q2 en Estadística y Probabilidad. La que tiene mayor media, con 28.3, es la revista *Technovation*. Tiene un factor de impacto de 5.729, el cuál es bastante más bajo que el de *Journal of Statistical Software*, que como vimos antes, tiene la tercera peor media con un 0.9.

Ahora nos paramos a mirar los datos resultantes del análisis descriptivo del total de artículos que contienen al menos una de las palabras clave desde el año 2000 hasta el 2021:

Mínimo	Primer Cuartil	Mediana	Media	Tercer Cuartil	Máximo
37	132.5	162	160.9	186.2	330

Tabla 3.2: Análisis descriptivo del número total de revistas

Una vez hecho el análisis inicial de nuestros datos, tanto de las revistas a nivel individual, como del número total de publicaciones por año (Tabla 3.2), desde el 2000 hasta el 2021, llega el turno de ajustar el modelo de regresión lineal simple. Hemos considerado como variable respuesta Y , la serie temporal anual de la transformación de Box-Cox para el número de artículos en las revistas de Estadística que contienen al menos una de las palabras clave, y como variable explicativa X , el año.

Antes de estimar los parámetros, es necesario comprobar si hay una relación lineal entre la variable respuesta y la explicativa, o no. Para ello nos apoyaremos en el coeficiente de correlación lineal de Pearson que tiene un valor de 0.85, el cuál está muy proximo a 1 y por tanto nos indica que existe una relación lineal positiva muy fuerte.

En caso de que al hacer la diagnosis del modelo para comprobar que se cumplen las hipótesis, obtengamos que alguna de ellas no lo hace, tenemos dos opciones. Una es cambiar el modelo por uno que no necesite que se cumplan esas hipótesis y otra, transformar los datos.

Normalmente, las hipótesis no se cumplen por el carácter asimétrico de sus variables, que es resultado de que alguna de ellas o las dos sean positivas. La hipótesis de normalidad no es la única perjudicada por la asimetría. Hipótesis como la linealidad y la homocedasticidad pueden verse comprometidas a causa de la modificación que sufren la función de regresión y la de la varianza condicional, respectivamente. Tenemos así, que realizar una transformación de nuestros datos nos puede solventar más de un problema de golpe.

Para el presente estudio, emplearemos la transformación de Box-Cox. Otras bastante comunes, que de hecho resultan ser dos casos particulares de las de Box-Cox, son la raíz cuadrada y el logaritmo. Dicha transformación se define:

$$\forall y \in (0, +\infty) \quad t_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln(y) & \text{si } \lambda = 0 \end{cases} \quad (3.1)$$

tenemos por tanto que la transformación depende del λ que elijamos, de modo que escojemos el valor de λ que nos de la transformación óptima para nuestro caso.

Tenemos además que dicha transformación se aplicará en una variable positiva ya que está definida en $(0, +\infty)$. Podemos distinguir varios casos en cuanto a la elección del λ :

- Si $\lambda > 1$, el efecto de la transformación hará que los valores menores que 1 se con-

traigan y que los mayores 1, por el contrario, se expandan según la potencia λ que hayamos seleccionado.

- Para $\lambda \in (0, 1)$, obtenemos el caso contrario al arriba mencionado, obteniendo la raíz correspondiente.
- Cuando $\lambda = 0$, lo que obtenemos es una contracción extrema de los valores grandes, mientras que los más cercanos a 0 se expanden hacia $-\infty$. Este caso es el del logaritmo.
- Finalmente, si $\lambda < 0$, lo que obtenemos es la variable invertida.

Debemos fijarnos en la naturaleza de las variables y en los diagramas de dispersión como criterio de selección de la transformación. Si no conseguimos el modelo lineal deseado, siempre nos quedará buscar modelos más flexibles o ver la influencia de otras explicativas.

Empleando el paquete de R mencionado en [6] el parámetro λ seleccionado para la transformación es $\lambda = 0.6$

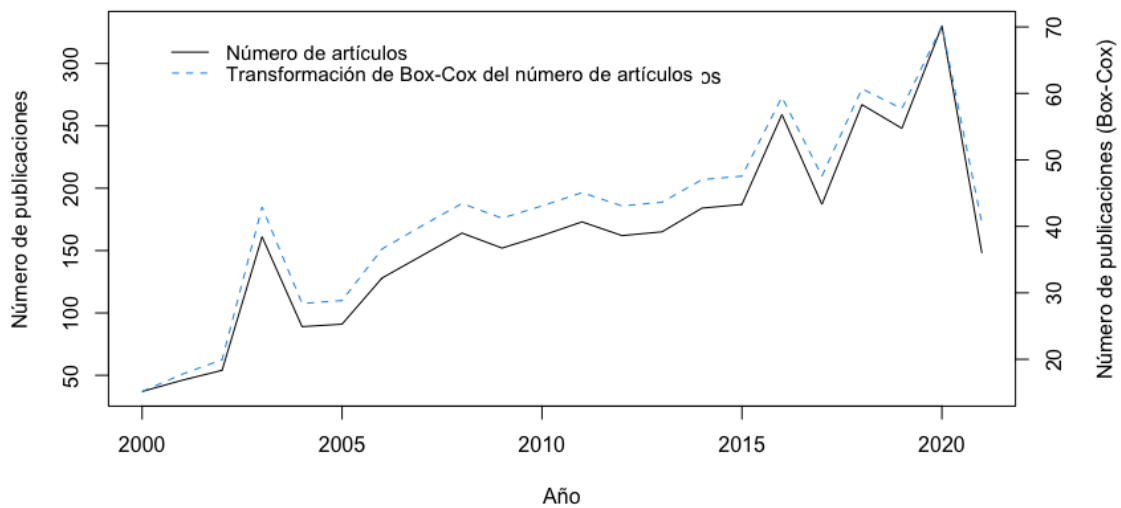


Figura 3.1: Serie Temporal del número de palabras clave (línea azul discontinua) y de la transformación de Box-Cox del número de palabras clave (línea negra)

	Estimación	Error Standard	Valor T	p-valor
$\hat{\beta}_0$	-3565.02	508.90	-7.01	8.50e-07
$\hat{\beta}_1$	1.79	0.25	7.09	7.18e-07

Tabla 3.3: Ajuste del modelo de regresión lineal simple donde la variable Y es la serie temporal de la transformación Box-Cox del número total de artículos que contienen las palabras claves desde el año 2000 hasta el 2021, y la variable X es el año

En el summary de nuestro modelo, podemos observar que el valor del error estándar residual es de 7.532, y el de R^2 de 0.7152. Además el valor del test F es de 50.23 y su correspondiente p valor, $7.18 \cdot 10^{-9}$. Tenemos así un valor del R^2 próximo a 1 y que junto a el bajo valor crítico para el F test, nos indican que la variable explicativa considerada para explicar la transformación Box-Cox para el número de artículos que contienen las palabras clave es buena. Además de que los coeficientes del modelo son significativos.

El modelo de regresión lineal sobre series temporales nos lleva a hacer implícitamente algunas suposiciones con respecto a las variables y los residuos. No solo asumimos que hay una relación lineal entre la variable dependiente y las explicativas, sino que los errores tienen media cero y siguen una distribución normal con varianza constante. También damos por hecho que no están autocorrelados y son independientes de las variables explicativas.

El test de Breusch-Godfrey, explora si se cumple el supuesto de independencia de los errores en regresión lineal. Sus hipótesis son:

$$\begin{cases} H_0 : \text{los errores son independientes} \\ H_a : \text{los errores no son independientes} \end{cases}$$

En este caso el p-valor obtenido al realizar dicho test es 0.1433 lo cuál nos da razones para decir que los residuos no están autocorrelados.

En este caso no hemos obtenido break-points, lo cuál nos indica que la tendencia no cambia. Nos quedamos con un modelo de regresión simple cuya recta se ajusta (Figura 3.2) mediante:

$$Y_i = -3565,022 + 1,794X_i$$

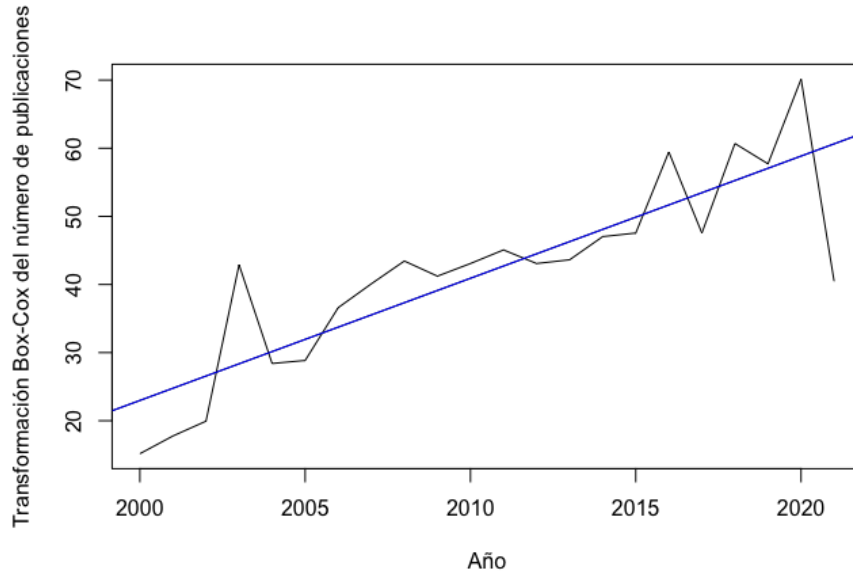


Figura 3.2: Transformación Box-Cox y ajuste del modelo segmentado (línea azul)

3.2. Resultados de las revistas de Investigación Operativa

Hagamos ahora un análisis descriptivo de las revistas correspondientes al área de Investigación Operativa. Nos pararemos a comentar los resultados obtenidos en cuanto a la inclusión de la perspectiva de género en las revistas del campo. Para ello, tendremos en cuenta la relevancia de estas revistas dentro de este campo y por tanto la repercusión de las mismas en la sociedad. Para esto, nos apoyaremos en los factores de impacto de las mismas, que como explicamos en la Sección 2.1 del Capítulo 2.

Sobre los datos recogidos en la Tabla 3.4, comenzaremos destacando que la revista que mayor media de publicaciones que contengan al menos una de las palabras clave tiene, es *Expert Systems with Applications*, con una media de 59.27. Esta revista solo tiene un factor de impacto de 5.452. Aunque es verdad que en las categorías en las que aparece, que son tres, está en todas en el Q1. La de menor media es *Queueing Systems*, cuya serie temporal destacamos en la Figura 2.4 por ser casi nula. Su factor de impacto de 1.114, que es todavía más bajo que el de la revista anterior y solo aparece en dos categorías en las cuáles está en el Q4. De hecho la diferencia entre el factor de impacto de la revista con mayor media y la revista con menor media es solo de 4.338.

Revistas	Mínimo	Primer Cuartil	Mediana	Media	Tercer Cuartil	Máximo
AOR	0	2	4	7.14	7	45
EJOR	0	9.50	16.50	16.32	20.75	32
IJGT	0	0.25	1.50	2.23	3	10
O	0	2	2.5	2.96	4	9
JOTA	0	0	0	0.18	0	1
T	2	4	6	5.55	7	10
IJPE	0	2.25	5.50	6.82	12.25	17
SEPS	0	3	4	7.68	6.75	49
ITOR	0	0	1.5	1.59	2	5
JS	0	0	0	0.5	1	2
OMS	0	0	0	0.09	0	1
ORS	0	0	0.5	0.82	1	5
ASMBI	0	0	1.5	1.68	3	5
EE	0	0	0	0.32	0.75	2
NSE	0	0.25	1	1.96	3	8
MC	0	0	0	0.23	0	1
MP	0	0	0	0.27	0.75	1
FODM	0	0	0	0.23	0	1
DSS	3	13.75	24	24.77	34.25	47
ESA	0	19.50	65.50	59.27	86.25	137
FSMJ	0	0	0	0.45	0.75	3
QS	0	0	0	0.05	0	1
JOS	0	0	0	0.91	1	6
PPC	0	0.25	1.5	3.09	3	21
IJITDM	0	2	3.5	3.59	5	11

Tabla 3.4: Análisis descriptivo del área de Investigación Operativa

Finalmente hacemos un análisis descriptivo del total de publicaciones que contienen alguna de las palabras clave desde 2000 hasta 2021:

Mínimo	Primer Cuartil	Mediana	Media	Tercer Cuartil	Máximo
31	65.25	175	148.68	202.25	310

Tabla 3.5: Análisis descriptivo Inicial del número total de artículos

Observando las tablas de los totales (Tablas 3.2 y 3.5), podemos concluir que en media, el área de estadística incluye la perspectiva de género en su contenido en mayor medida que el área de Investigación Operativa.

Nuevamente comprobaremos si hay una relación lineal entre la variable respuesta y la explicativa. En este caso, el coeficiente de correlación lineal de Pearson tiene un valor de 0.923, que indica una relación lineal positiva muy fuerte. Una vez hecho el análisis de los datos, fue necesario considerar una transformación de Box-Cox de parámetro $\lambda = 0.626$. Dicho valor del parámetro fue nuevamente obtenido con el paquete de R [6].

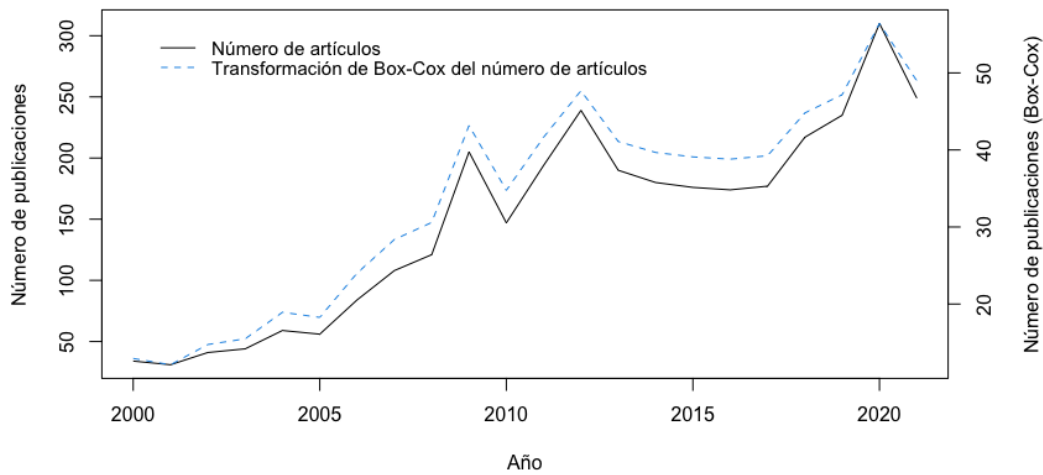


Figura 3.3: Serie Temporal del número de palabras clave (línea azul discontinua) y de la transformación de Box-Cox del número de palabras clave (línea negra)

	Estimación	Error Standard	Valor T	p-valor
$\hat{\beta}_0$	-18661.28	2748.98	-6.79	1.33e-06
$\hat{\beta}_1$	9.36	1.37	6.85	1.18e-06

Tabla 3.6: Ajuste del modelo de regresión lineal simple donde la variable Y es la serie temporal de la transformación Box-Cox del número total de artículos que contienen las palabras claves desde el año 2000 hasta el 2021, y la variable X es el año

Otros datos a destacar obtenidos al hacer el resumen del modelo, son el error estándar residual que tiene un valor de 40.69. El valor del F test es de 46.88 y su p-valor de $1.18 \cdot 10^{-6}$.

El coeficiente de determinación es de 0.7, valor próximo a 1, que junto con el bajo valor para el F test, nos indica que la variable explicativa empleada es útil para explicar la transformación de Box-Cox. Por último, cabe destacar, la significación de los coeficientes del modelo que es elevada.

Nuevamente tenemos que tener en cuenta que el modelo de regresión lineal sobre series temporales nos lleva a hacer implícitamente algunas suposiciones con respecto a las variables y los residuos. Para el supuesto de independencia de los errores volvemos a usar el test de Breusch-Godfrey, que nos da un p-valor de 0.02. Tenemos así, que podemos suponer que los residuos no son autocorrelados.

Una vez ajustado y validado, segmentamos el modelo. El break-point estimado al emplear el método de Muggeo corresponde al año 2011 (Figura 3.4). Tenemos por tanto que:

- Antes del 2011:

$$Y_i = -5658,778 + 2,834X_i$$

- Después del 2011:

$$Y_i = -5658,778 + 2,834X_i - 1,924(X_i - 2011)$$

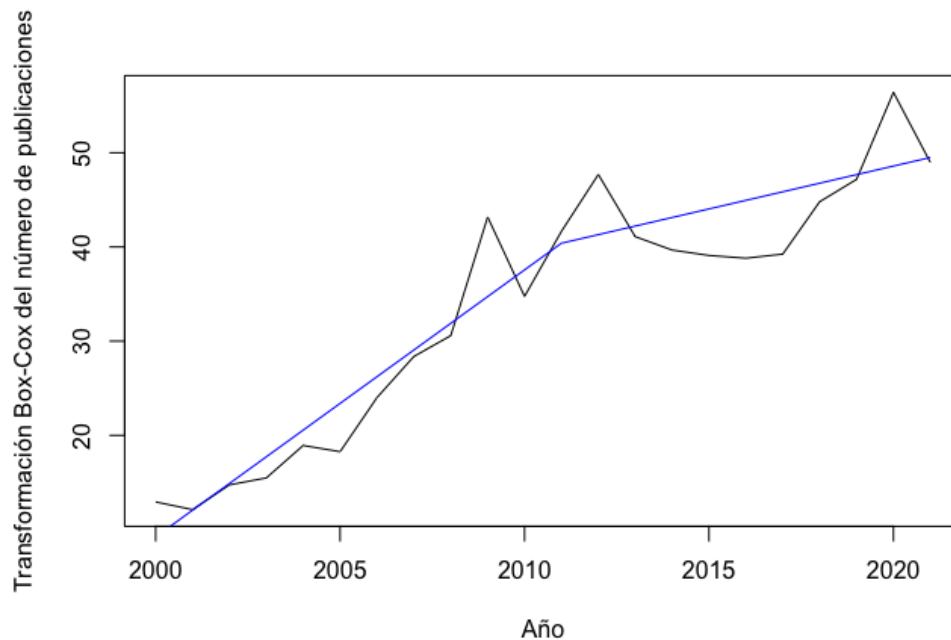


Figura 3.4: Transformación Box-Cox y ajuste del modelo segmentado (línea azul)

3.3. Comparativa con el área de Matemática Aplicada

En esta sección realizaremos un análisis para el área de Matemática Aplicada, análogo al realizado para las dos anteriores. Aunque se podía haber hecho para cualquier otra rama de las matemáticas (incluso para otras disciplinas científicas), la elección se justifica por la aplicabilidad de sus teorías sobre problemas de la vida real. La selección de la muestra de revistas escogidas para este caso está hecha siguiendo el criterio mencionado en la Sección 2.1. Esta vez, como solo es para comparar, ha sido conveniente trabajar con un total de 20 revistas. Las recogemos en la siguiente tabla:

Revistas Investigación Operativa

Nonlinear Analysis: Hybrid Systems (NAHS)
 Forum of Mathematics Pi (FMP)
 Journal of Cryptology (JC)
 Chaos (C)
 Applied Mathematics Letters (AML)
 Applied and Computational Harmonic Analysis (ACHA)
 Archive for Rational Mechanics and Anal (ARMA)
 SIAM Journal on Imaging Science (JIS)
 SIAM Journal on Optimization (JO)
 BIT Numerical Mathematics (BITNM)
 Finite Elements in Analysis and Design (FEAD)
 Communications on Pure and Applied Mathematics (CPAM)
 Numerical Algorithms (NA)
 Journal of Scientific Computing (JSC)
 Nonlinear Analysis: Real World Applications (RWA)
 Analysis and Mathematical Physics (AMP)
 SIAM Review (SR)
 Physica D: Nonlinear Phenomena (PDNP)
 Boundary Value Problems (BVP)
 Communications in Nonlinear Science and Numerical Simulation (CNSNS)
 Journal of Fourier Analysis and Applications (JFAA)
 Qualitative Theory of Dynamical Systems (QTDS)
 Stochastics and Partial Differential Equations: Analysis and Computations (SPDEAC)
 Journal of Mathematical Imaging and Vision (JMIV)
 Journal of Complexity (JOC)

Tabla 3.7: Lista de revistas de Investigación Operativa

Nuevamente, como se muestra en el Apéndice A, hemos convertido los datos a series temporales y las hemos representado en la Figura 3.5.

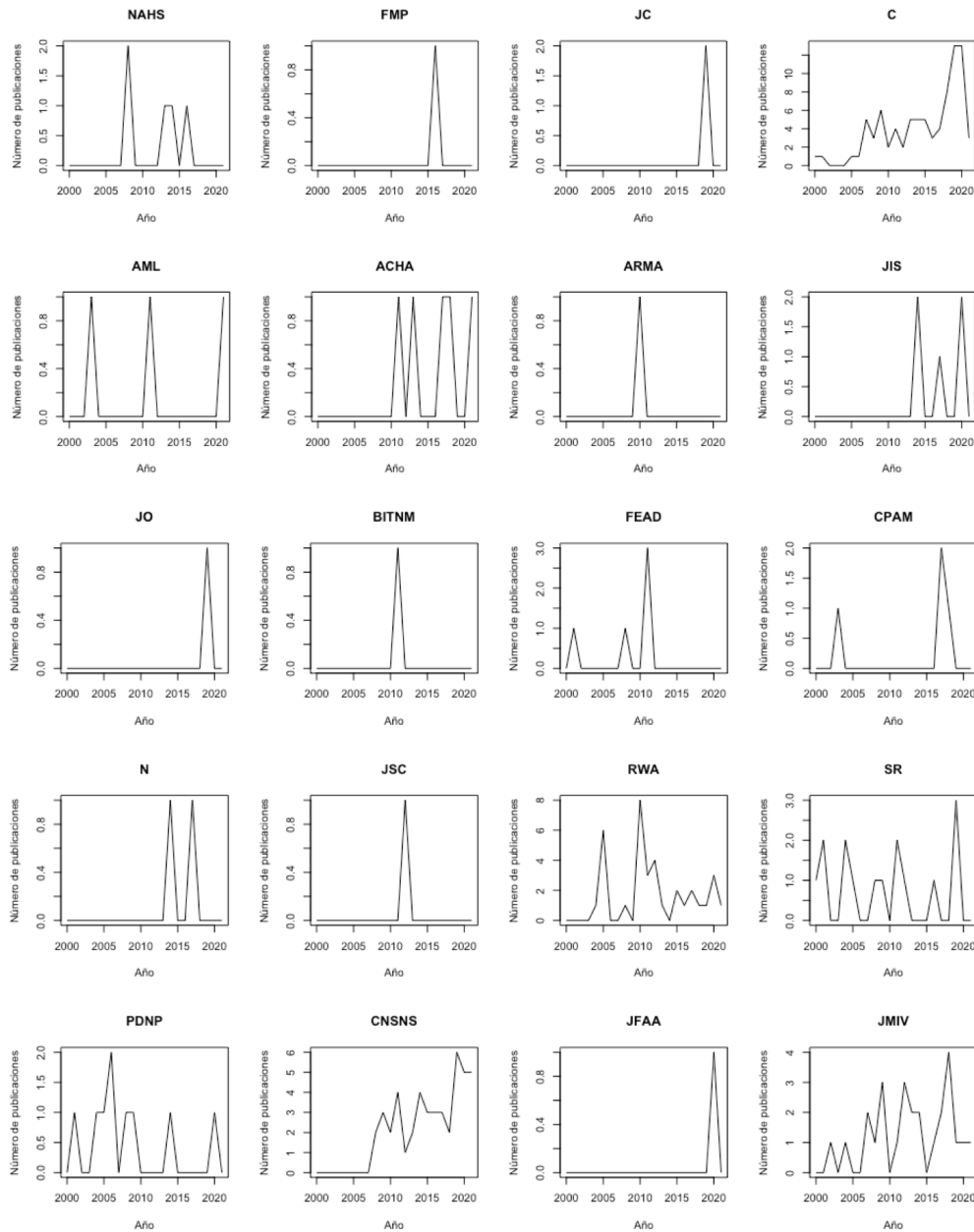


Figura 3.5: Series Temporales de las revistas de Matemática Aplicada

En este caso no haremos zoom, pero si merece la pena comentar que el número de revistas cuyas series temporales son prácticamente nulas, es más elevado que en las otras dos áreas consideradas.

Mínimo	Primer Cuartil	Mediana	Media	Tercer Cuartil	Máximo
1	5.5	12	11.5	15.5	28

Tabla 3.8: Análisis descriptivo del número total de publicaciones que contienen al menos una de las palabras clave

También podemos observar el análisis descriptivo del total de revistas del área de Matemática Aplicada (Tabla 3.8), en el cuál obtenemos unos datos peores que los obtenidos anteriormente por las áreas de Estadística e Investigación Operativa. De hecho donde antes obteníamos unas medias de 160.9 y de 148.68, para las áreas de Estadística e Investigación Operativa respectivamente, ahora obtenemos una media de 11.5.

El coeficiente de correlación lineal de Pearson tiene un valor de 0.82, que vuelve a indicar una relación lineal positiva muy fuerte debido a su proximidad a 1. Una vez hecho el análisis de los datos, fue necesario considerar una transformación de Box-Cox de parámetro $\lambda = 0.505$

	Estimación	Error Standard	Valor T	p-valor
$\hat{\beta}_0$	-1787.85	302.99	-5.90	9.02e-06
$\hat{\beta}_1$	0.89	0.15	5.94	8.29e-06

Tabla 3.9: Ajuste del modelo de regresión simple

Nuevamente, al hacer el resumen del modelo, nos encontramos con más datos que los mencionados en la Tabla 3.9. Destacaremos que el error estándar residual tiene un valor de 1.36. El valor del F test es de 42.34 y su p valor de $2.42 \cdot 10^{-6}$. El coeficiente de determinación es de 0.68, valor no excesivamente próximo a 1, pero que junto con el bajo valor para el F test, nos indica que la variable explicativa empleada es útil para explicar la transformación de Box-Cox. Finalmente destacamos, que la significación de los coeficientes del modelo es elevada.

Por último, en la Figura 3.6, podemos observar la comparación de las series temporales, la original y la de la transformación.

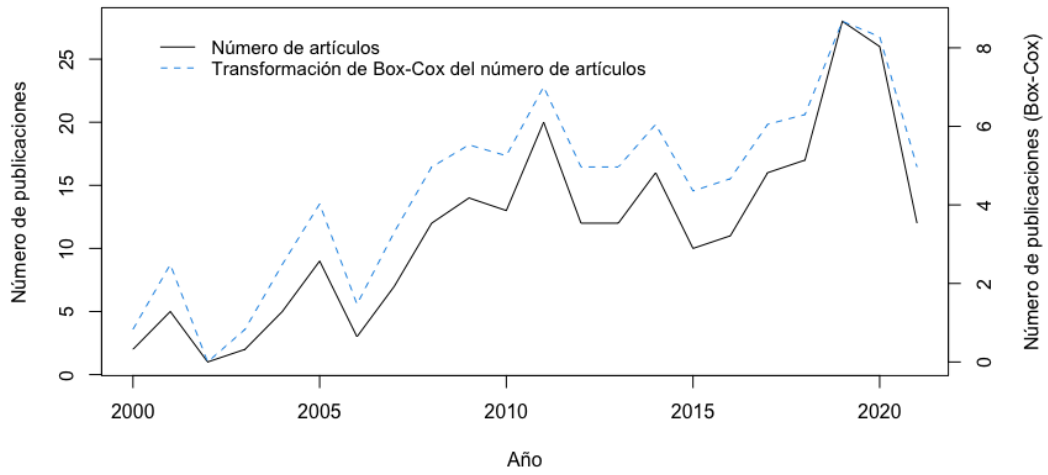


Figura 3.6: Serie Temporal del número de palabras clave (línea azul discontinua) y de la transformación de Box-Cox del número de palabras clave (línea negra)

Capítulo 4

Conclusiones

Aunque en los últimos años el papel de la mujer en los diferentes ámbitos de la sociedad ha sufrido un importante impulso, aún no se ha conseguido una igualdad real mujer vs. hombre. La educación e investigación universitarias no son ajenas a esta situación a pesar de que en los últimos años las principales organizaciones y entidades educativas han hecho hincapié en el fortalecimiento de la mujer en la ciencia. Sin embargo, la docencia en Matemáticas parece no verse favorecida por este impulso. En un primer paso de nuestra investigación, hemos analizado el impacto de la perspectiva de género en la docencia en el Grado de Matemáticas de la USC, basándonos en las directrices que da Rebollar en [10]. Frente a otras facultades españolas que imparten docencia en Matemáticas, al menos las guías docentes de las materias aquí impartidas no reflejan dicha distinción. Sin duda, si vemos la etapa universitaria como un período de crecimiento y desarrollo personal clave para el alumnado, cabe pensar que los futuros egresados han perdido la oportunidad de educarse en perspectiva de género en matemáticas.

Al inicio de este trabajo nos encontrábamos con una gran cantidad de datos recogidos, bastante poco alentadores, sobre la inclusión de la perspectiva de género en la investigación. Queríamos ver si las políticas implantadas a lo largo de estos últimos años estaban marcando una diferencia, o si por el contrario, están cayendo en saco roto.

Para este análisis nos ayudamos de los modelos de regresión, tanto del simple, como del segmentado. Gracias a estos modelos nos ha sido posible. Ahora ya, con los resultados sobre el papel, toca sacar conclusiones, y por desgracia, no son las que nos gustaría leer. Es verdad que, en general, hay una tendencia creciente en el número de artículos que incluyen la perspectiva de género. Pero en el área de Investigación Operativa, podemos observar un cambio en la tendencia en el año 2011. Este cambio no es positivo, sino que nos indica, que el número de publicaciones que incluyen la perspectiva de género no crece tanto como lo hacía antes de ese año.

Estos resultados muestran la insuficiencia de las propuestas hechas hasta la fecha. Tanto las organizaciones mundiales, a título global, como cada uno de nosotros a título personal, tenemos que intentar seguir haciendo ver que todavía no somos iguales. Tenemos que seguir analizando que está fallando y como solucionarlo, sacando nuevos proyectos y nuevas leyes que nos ayuden poco a poco a que la mujer tenga el sitio que se merece en la ciencia. Tenemos que estar concienciados de que si colaboramos avanzamos más rápido y de que si dejamos atrás, a la mitad de la población, por el simple hecho de ser mujeres, estamos desperdiciando un potencial que nos podría ayudar a crecer como sociedad.

Recogiendo un poco lo visto a lo largo de este trabajo, puede que el problema no radique exclusivamente en la cantidad de leyes. Quizás debería priorizarse la cantidad de gente a la que se llega, ya que muchas propuestas y proyectos son desconocidos para la mayoría de la población. El primer escalón a superar, debería ser la propia inseguridad experimentada por niñas y mujeres que muchas veces hace que no se decanten por carreras de ciencias o proyectos de investigación. Y eso tenemos que atajarlo en el ámbito de la enseñanza, incluyendo la perspectiva de género tanto en colegios como universidades. Esta situación podría verse facilitada si conseguimos subsanar los obstáculos a los que se enfrentan. Es por eso que hay que seguir trabajando en materia de género dentro del ámbito de la investigación y de la sociedad. Deberíamos conseguir que las revistas no hagan esas distinciones a la hora de elegir a quién le publican los artículos, que las becas sean concedidas a quien las merece sin que el género sea un obstáculo. Sobre todo, no deberíamos perder lo ya conseguido.

Apéndice A

Series Temporales

A.1. Estadística

En esta sección del Apéndice, recogeremos el código empleado para representar las series temporales del área de Estadística. En la Subsección A.1.1 pondremos el código correspondiente a las series temporales de cada una de las 25 revistas y en la Subsección A.1.2, el del número total de artículos.

A.1.1. Individuales

Primero hemos hecho la lectura de los datos correspondientes a las revistas de Estadística, que previamente hemos recogido en un documento .csv. Después hemos convertido dichos datos en series temporales y finalmente los hemos representado.

```
#Lectura de datos
estadistica=read.csv2("tablaest.csv",encoding="UTF-8",stringsAsFactors=TRUE)
head(estadistica)
attach(estadistica)
```

```
#Convertimos las variables a series temporales
pub_JSS=ts(JSS,freq=1,start=c(2000,1))
pub_CSDA=ts(CSDA,freq=1,start=c(2000,1))
pub_JMLR=ts(JMLR,freq=1,start=c(2000,1))
pub_QTQM=ts(QTQM,freq=1,start=c(2000,1))
pub_B=ts(B,freq=1,start=c(2000,1))
pub_JSPI=ts(JSPI,freq=1,start=c(2000,1))
pub_T=ts(T,freq=1,start=c(2000,1))
```

```
pub_E=ts(E,freq=1,start=c(2000,1))
pub_CILS=ts(CILS,freq=1,start=c(2000,1))
pub_SMMR=ts(SMMR,freq=1,start=c(2000,1))
pub_FS=ts(FS,freq=1,start=c(2000,1))
pub_JBES=ts(JBES,freq=1,start=c(2000,1))
pub_SADM=ts(SADM,freq=1,start=c(2000,1))
pub_PS=ts(PS,freq=1,start=c(2000,1))
pub_IME=ts(IME,freq=1,start=c(2000,1))
pub_SJ=ts(SJ,freq=1,start=c(2000,1))
pub_EX=ts(EX,freq=1,start=c(2000,1))
pub_SN=ts(SN,freq=1,start=c(2000,1))
pub_AB=ts(AB,freq=1,start=c(2000,1))
pub_LDA=ts(LDA,freq=1,start=c(2000,1))
pub_SM=ts(SM,freq=1,start=c(2000,1))
pub_FSS=ts(FSS,freq=1,start=c(2000,1))
pub_SS=ts(SS,freq=1,start=c(2000,1))
pub_SERRA=ts(SERRA,freq=1,start=c(2000,1))
pub_ADAC=ts(ADAC,freq=1,start=c(2000,1))
```

#Gráficas de las series Temporales

```
par(mfrow=c(2,5))
plot(pub_JSS,main="JSS",xlab="Año",ylab="Número de publicaciones")
plot(pub_CSDA,main="CSDA",xlab="Año",ylab="Número de publicaciones")
plot(pub_JMLR,main="JMLR",xlab="Año",ylab="Número de publicaciones")
plot(pub_QTQM,main="QTQM",xlab="Año",ylab="Número de publicaciones")
plot(pub_B,main="B",xlab="Año",ylab="Número de publicaciones")
plot(pub_JSPI,main="JSPI",xlab="Año",ylab="Número de publicaciones")
plot(pub_T,main="T",xlab="Año",ylab="Número de publicaciones")
plot(pub_E,main="E",xlab="Año",ylab="Número de publicaciones")
plot(pub_CILS,main="CILS",xlab="Año",ylab="Número de publicaciones")
plot(pub_SMMR,main="SMMR",xlab="Año",ylab="Número de publicaciones")
plot(pub_FS,main="FS",xlab="Año",ylab="Número de publicaciones")
plot(pub_JBES,main="JBES",xlab="Año",ylab="Número de publicaciones")
plot(pub_SADM,main="SADM",xlab="Año",ylab="Número de publicaciones")
plot(pub_PS,main="PS",xlab="Año",ylab="Número de publicaciones")
plot(pub_IME,main="IME",xlab="Año",ylab="Número de publicaciones")
```

```
plot(pub_SJ,main="SJ",xlab="Año",ylab="Número de publicaciones")
plot(pub_EX,main="EX",xlab="Año",ylab="Número de publicaciones")
plot(pub_SN,main="SN",xlab="Año",ylab="Número de publicaciones")
plot(pub_AB,main="AB",xlab="Año",ylab="Número de publicaciones")
plot(pub_LDA,main="LDA",xlab="Año",ylab="Número de publicaciones")
plot(pub_SM,main="SM",xlab="Año",ylab="Número de publicaciones")
plot(pub_FSS,main="FSS",xlab="Año",ylab="Número de publicaciones")
plot(pub_SS,main="SS",xlab="Año",ylab="Número de publicaciones")
plot(pub_SERRA,main="SERRA",xlab="Año",ylab="Número de publicaciones")
plot(pub_ADAC,main="ADAC",xlab="Año",ylab="Número de publicaciones")

#Zoom de las series temporales
par(mfrow=c(2,2))
plot(pub_JSS,main="JSS",xlab="Año",ylab="Número de publicaciones")
plot(pub_B,main="B",xlab="Año",ylab="Número de publicaciones")
plot(pub_JBES,main="JBES",xlab="Año",ylab="Número de publicaciones")
plot(pub_PS,main="PS",xlab="Año",ylab="Número de publicaciones")
```

A.1.2. Total

En esta subsección hemos hecho lo análogo para el número total de las revistas, cuyos datos recogimos previamente en un documento .txt.

```
#Lectura de datos
totalest=read.table("revistasest_total.txt",header=TRUE)
head(totalest)
attach(totalest)
#Convertimos a serie temporal
st_est=ts(Número,freq=1,start=c(2000,1))
#Representamos
plot(st_est,main="Serie Temporal del número total de artículos",xlab="Año",
ylab="Número de publicaciones")
```

A.2. Investigación Operativa

Volvemos a dividir la sección en dos subsecciones con los mismos apartados. Una para las series de todas las revistas y otra para la del total, pero en este caso para el área de Investigación Operativa.

A.2.1. Individuales

Como en la Subsección A.1.1, recogimos los datos de las revistas de Investigación Operativa en un documento .csv. Procedimos a la lectura de datos y posteriormente convertimos dichos datos en series temporales que finalmente representamos.

```
#Lectura de datos
inv_op=read.csv2("tablaio.csv",encoding="UTF-8",stringsAsFactors=TRUE)
head(inv_op)
attach(inv_op)

#Convertimos las variables a series temporales
pub_AOR=ts(AOR,freq=1,start=c(2000,1))
pub_EJOR=ts(EJOR,freq=1,start=c(2000,1))
pub_IJGT=ts(IJGT,freq=1,start=c(2000,1))
pub_O=ts(O,freq=1,start=c(2000,1))
pub_JOTA=ts(JOTA,freq=1,start=c(2000,1))
pub_T=ts(T,freq=1,start=c(2000,1))
pub_IJPE=ts(IJPE,freq=1,start=c(2000,1))
pub_SEPS=ts(SEPS,freq=1,start=c(2000,1))
pub_ITOR=ts(ITOR,freq=1,start=c(2000,1))
pub_JS=ts(JS,freq=1,start=c(2000,1))
pub_OMS=ts(OMS,freq=1,start=c(2000,1))
pub_ORs=ts(ORS,freq=1,start=c(2000,1))
pub_ASMBI=ts(ASMBI,freq=1,start=c(2000,1))
pub_EE=ts(EE,freq=1,start=c(2000,1))
pub_NSE=ts(NSE,freq=1,start=c(2000,1))
pub_MC=ts(MC,freq=1,start=c(2000,1))
pub_MP=ts(MP,freq=1,start=c(2000,1))
pub_FODM=ts(FODM,freq=1,start=c(2000,1))
pub_DSS=ts(DSS,freq=1,start=c(2000,1))
pub_ESA=ts(ESA,freq=1,start=c(2000,1))
```

```
pub_FSMJ=ts(FSMJ,freq=1,start=c(2000,1))
pub_QS=ts(QS,freq=1,start=c(2000,1))
pub_JOS=ts(JOS,freq=1,start=c(2000,1))
pub_PPC=ts(PPC,freq=1,start=c(2000,1))
pub_IJITDM=ts(IJITDM,freq=1,start=c(2000,1))

#Gráficas de las series Temporales
par(mfrow=c(2,5))
plot(pub_AOR,main="AOR",xlab="Año",ylab="Número de publicaciones")
plot(pub_EJOR,main="EJOR",xlab="Año",ylab="Número de publicaciones")
plot(pub_IJGT,main="IJGT",xlab="Año",ylab="Número de publicaciones")
plot(pub_O,main="O",xlab="Año",ylab="Número de publicaciones")
plot(pub_JOTA,main="JOTA",xlab="Año",ylab="Número de publicaciones")
plot(pub_T,main="T",xlab="Año",ylab="Número de publicaciones")
plot(pub_IJPE,main="IJPE",xlab="Año",ylab="Número de publicaciones")
plot(pub_SEPS,main="SEPS",xlab="Año",ylab="Número de publicaciones")
plot(pub_ITOR,main="ITOR",xlab="Año",ylab="Número de publicaciones")
plot(pub_JS,main="JS",xlab="Año",ylab="Número de publicaciones")
plot(pub_OMS,main="OMS",xlab="Año",ylab="Número de publicaciones")
plot(pub_ORs,main="ORS",xlab="Año",ylab="Número de publicaciones")
plot(pub_ASMBI,main="ASMBI",xlab="Año",ylab="Número de publicaciones")
plot(pub_EE,main="EE",xlab="Año",ylab="Número de publicaciones")
plot(pub_NSE,main="NSE",xlab="Año",ylab="Número de publicaciones")
plot(pub_MC,main="MC",xlab="Año",ylab="Número de publicaciones")
plot(pub_MP,main="MP",xlab="Año",ylab="Número de publicaciones")
plot(pub_FODM,main="FODM",xlab="Año",ylab="Número de publicaciones")
plot(pub_DSS,main="DSS",xlab="Año",ylab="Número de publicaciones")
plot(pub_ESA,main="ESA",xlab="Año",ylab="Número de publicaciones")
plot(pub_FSMJ,main="FSMJ",xlab="Año",ylab="Número de publicaciones")
plot(pub_QS,main="QS",xlab="Año",ylab="Número de publicaciones")
plot(pub_JOS,main="JOS",xlab="Año",ylab="Número de publicaciones")
plot(pub_PPC,main="PPC",xlab="Año",ylab="Número de publicaciones")
plot(pub_IJITDM,main="IJITDM",xlab="Año",ylab="Número de publicaciones")

#Zoom de las series temporales
par(mfrow=c(2,2))
```

```
plot(pub_IJPE,main="IJPE",xlab="Año",ylab="Número de publicaciones")
plot(pub_SEPS,main="SEPS",xlab="Año",ylab="Número de publicaciones")
plot(pub_OMS,main="OMS",xlab="Año",ylab="Número de publicaciones")
plot(pub_QS,main="QS",xlab="Año",ylab="Número de publicaciones")
```

A.2.2. Total

Después de hacer el recuento del total de publicaciones que contienen al menos una de las palabras clave desde el año 2000 hasta el 2021, hemos recogido los datos en un archivo .txt. Tras la lectura de datos, hemos convertido los mismos en series temporales y posteriormente los hemos representado.

```
#Lectura de datos
totalio=read.table("revistasio_total.txt",header=TRUE)
head(totalio)
attach(totalio)
#Convertimos a serie temporal
st_io=ts(Número,freq=1,start=c(2000,1))
#Representamos
plot(st_io,xlab="Año",ylab="Número de publicaciones")
```

A.3. Matemática Aplicada

Por último, adjuntamos el código para la representación de las series temporales de las revistas, tanto individualmente como para el total, del área de Matemática Aplicada.

A.3.1. Individuales

El procedimiento llevado a cabo en esta subsección es el análogo al de las Subsecciones A.1.1 y A.2, solo que en este caso los datos corresponden a las revistas del área de Matemática aplicada.

```
#Lectura de datos
mat_apl=read.csv2("tablapl.csv",encoding="UTF-8",stringsAsFactors=TRUE)
head(mat_apl)
attach(mat_apl)

#Convertimos las variables a series temporales
```

```

pub_NAHS=ts(NAHS,freq=1,start=c(2000,1))
pub_FMP=ts(FMP,freq=1,start=c(2000,1))
pub_JC=ts(JC,freq=1,start=c(2000,1))
pub_C=ts(C,freq=1,start=c(2000,1))
pub_AML=ts(AML,freq=1,start=c(2000,1))
pub_ACHA=ts(ACHA,freq=1,start=c(2000,1))
pub_ARMA=ts(ARMA,freq=1,start=c(2000,1))
pub_JIS=ts(JIS,freq=1,start=c(2000,1))
pub_JO=ts(JO,freq=1,start=c(2000,1))
pub_BITNM=ts(BITNM,freq=1,start=c(2000,1))
pub_FEAD=ts(FEAD,freq=1,start=c(2000,1))
pub_CPAM=ts(CPAM,freq=1,start=c(2000,1))
pub_N=ts(N,freq=1,start=c(2000,1))
pub_JSC=ts(JSC,freq=1,start=c(2000,1))
pub_RWA=ts(RWA,freq=1,start=c(2000,1))
pub_SR=ts(SR,freq=1,start=c(2000,1))
pub_PDNP=ts(PDNP,freq=1,start=c(2000,1))
pub_CNSNS=ts(CNSNS,freq=1,start=c(2000,1))
pub_JFAA=ts(JFAA,freq=1,start=c(2000,1))
pub_JMIV=ts(JMIV,freq=1,start=c(2000,1))

#Gráficas de las series Temporales
par(mfrow=c(2,4))
plot(pub_NAHS,main="NAHS",xlab="Año",ylab="Número de publicaciones")
plot(pub_FMP,main="FMP",xlab="Año",ylab="Número de publicaciones")
plot(pub_JC,main="JC",xlab="Año",ylab="Número de publicaciones")
plot(pub_C,main="C",xlab="Año",ylab="Número de publicaciones")
plot(pub_AML,main="AML",xlab="Año",ylab="Número de publicaciones")
plot(pub_ACHA,main="ACHA",xlab="Año",ylab="Número de publicaciones")
plot(pub_ARMA,main="ARMA",xlab="Año",ylab="Número de publicaciones")
plot(pub_JIS,main="JIS",xlab="Año",ylab="Número de publicaciones")
plot(pub_JO,main="JO",xlab="Año",ylab="Número de publicaciones")
plot(pub_BITNM,main="BITNM",xlab="Año",ylab="Número de publicaciones")
plot(pub_FEAD,main="FEAD",xlab="Año",ylab="Número de publicaciones")
plot(pub_CPAM,main="CPAM",xlab="Año",ylab="Número de publicaciones")
plot(pub_N,main="N",xlab="Año",ylab="Número de publicaciones")

```

```
plot(pub_JSC,main="JSC",xlab="Año",ylab="Número de publicaciones")
plot(pub_RWA,main="RWA",xlab="Año",ylab="Número de publicaciones")
plot(pub_SR,main="SR",xlab="Año",ylab="Número de publicaciones")
plot(pub_PDNP,main="PDNP",xlab="Año",ylab="Número de publicaciones")
plot(pub_BVP,main="BVP",xlab="Año",ylab="Número de publicaciones")
plot(pub_CNSNS,main="CNSNS",xlab="Año",ylab="Número de publicaciones")
plot(pub_JFAA,main="JFAA",xlab="Año",ylab="Número de publicaciones")
plot(pub_JMIV,main="JMIV",xlab="Año",ylab="Número de publicaciones")
```

A.3.2. Total

Nuevamente tras el recuento total, hemos convertido los datos y representado las series temporales.

```
#Lectura de datos
totalapl=read.table("revistasma_total.txt",header=TRUE)
head(totalapl)
attach(totalapl)
#Convertimos a series temporales
st_apl=ts(Número,freq=1,start=c(2000,1))
#Representamos
plot(st_apl,xlab="Año",ylab="Número de publicaciones")
```

Apéndice B

Resultados

Presentaremos el código empleado para la obtención de los resultados comentados en el Capítulo 3.

B.1. Resultados en el área de Estadística

Empezamos por el área de Estadística. Incluimos lo hecho a lo largo de este estudio para obtener los resultados comentados en la Sección 3.1.

```
#Cargamos las librerías necesarias:
library(MASS)
library(segmented)
library(lmtest)

#Lectura de los datos
totalest=read.table("revistasest_total.txt",header=TRUE)
head(totalest)
attach(totalest)

#Veamos que lambda cogemos
#Transformación Box-Cox:
trans.boxcox=boxcox(modelo_est)
lambda_est=trans.boxcox$x[which.max(trans.boxcox$y)]
lambda_est

#Cambiamos la variable por la transformada:
```

```
Num_est=((Número^lambda_est)-1)/lambda_est

#Hacemos la serie temporal de la transformación
stnew_est=ts(Num_est,freq=1,start=c(2000,1))

#Representamos la serie temporal original y la de la transformación
par(mar = c(5, 4, 4, 4) + 0.5)
plot(st_est,xlab="Año", ylab="Número de publicaciones")
par(new=TRUE)
plot(stnew_est, col=4,lty=2,axes = FALSE,bty = "n", xlab="", ylab="")
axis(4)
mtext("Número de publicaciones (Box-Cox)", side = 4,line=3)
legend("topleft",inset=0.05,legend = c("Número de artículos",
"Transformación de Box-Cox del número de artículos"), lty = c(1, 2),
col = c(1, 4),box.lty = 0,cex = 0.95)

#modelo transformado
#Modelo transformado
mod_est_new<-lm(stnew_est~Año)
mod_est_new
summary(mod_est_new)

#Diagnosis del modelo:
#-Homocedasticidad
hmctest(modelo_est_new)
bptest(mod_est_new)
#-Normalidad
res_io_new<-rstudent(modelo_est_new)
shapiro.test(res_est_new)
#-Linealidad
reset(modelo_est_new)

#Test de Breush-Godfrey
bgtest(mod_est_new)
```

B.2. Resultados en el área de Investigación Operativa

En esta sección, adjuntamos el código correspondiente al área de Investigación Operativa. Los resultados obtenidos son analizados y comentados en la Sección 3.2.

```
#Cargamos las librerías necesarias:
library(MASS)
library(segmented)
library(lmtest)

#Lectura de datos
totalio=read.table("revistasio_total.txt",header=TRUE)
head(totalio)
attach(totalio)

#Análisis descriptivo inicial
summary(totalio)

#Convertimos a serie temporal
st_io=ts(Número,freq=1,start=c(2000,1))

#Primero ajustamos el modelo de regresión lineal simple
modelo_io<-lm(st_io~Año)
summary(modelo_io)

#Para ver que lambda cogemos
#Transformación Box-Cox:
trans.boxcox=boxcox(modelo_io)
lambda_io=trans.boxcox$x[which.max(trans.boxcox$y)]
lambda_io

#Cambiamos la variable por la transformada:
Num_io=((Número^lambda_io)-1)/lambda_io

#Hacemos la serie temporal de la transformación
```

```
stnew_io=ts(Num_io,freq=1,start=c(2000,1))

#Representamos la serie temporal original y la de la transformación
stnew_io=ts(Num_io,freq=1,start=c(2000,1))
par(mar = c(5, 4, 4, 4) + 0.5)
plot(st_io,xlab="Año", ylab="Número de publicaciones")
par(new=TRUE)
plot(stnew_io, col=4,lty=2,axes = FALSE,bty = "n", xlab="", ylab="")
axis(4)
mtext("Número de publicaciones (Box-Cox)", side = 4,line=3)
legend("topleft",inset=0.05,legend = c("Número de artículos",
"Transformación de Box-Cox del número de artículos"), lty = c(1, 2),
col = c(1, 4),box.lty = 0,cex = 0.95)

#Modelo con la transformación Box-Cox
mod_io_new<-lm(stnew_io~Año)
mod_io_new
summary(mod_io_new)

#Diagnosis del modelo:
#-Homocedasticidad
hmctest(modelo_io_new)
bptest(mod_io_new)
#-Normalidad
res_io_new<-rstudent(modelo_io_new)
shapiro.test(res_io_new)
#-Linealidad
reset(modelo_io_new)
#Test de Breush-Godfrey
bgttest(mod_io_new)

#Segmentamos el modelo
segmented(mod_io_new)
```

B.3. Resultados de la comparativa con el área de Matemática Aplicada

Finalmente, en esta tercera y última sección, proporcionamos el código correspondiente al análisis comparativo de área de Matemática Aplicada.

```
#Cargamos las librerías necesarias:
library(MASS)
library(segmented)
library(lmtest)

#Lectura de datos
totalapl=read.table("revistasma_total.txt",header=TRUE)
head(totalapl)
attach(totalapl)

#Convertimos a serie temporal
st_apl=ts(Número,freq=1,start=c(2000,1))

#Primero ajustamos el modelo de regresión lineal simple
modelo_apl<-lm(st_apl~Año)
summary(modelo_apl)

#Transformación Box-Cox:
trans.boxcox=boxcox(modelo_apl)
lambda_apl=trans.boxcox$x[which.max(trans.boxcox$y)]
lambda_apl

#Cambiamos la variable por la transformada:
Num_apl=((Número^lambda_apl)-1)/lambda_apl

#Hacemos la serie temporal de la transformación
stnew_apl=ts(Num_apl,freq=1,start=c(2000,1))

#Representamos la serie temporal original y la de la transformación
par(mar = c(5, 4, 4, 4) + 0.5)
```

```
plot(st_apl,xlab="Año", ylab="Número de publicaciones")
par(new=TRUE)
plot(stnew_apl, col=4,lty=2,axes = FALSE,bty = "n", xlab="", ylab="")
axis(4)
mtext("Número de publicaciones (Box-Cox)", side = 4,line=3)
legend("topleft",inset=0.05,legend = c("Número de artículos",
"Transformación de Box-Cox del número de artículos"), lty = c(1, 2),
col = c(1, 4),box.lty = 0,cex = 0.95)

#Modelo con la transformación Box-Cox
mod_apl_new<-lm(Num_apl~Año)
mod_apl_new

#Diagnosis del modelo:
#-Homocedasticidad
hmctest(modelo_apl_new)
bptest(mod_apl_new)
#-Normalidad
res_apl_new<-rstudent(modelo_apl_new)
shapiro.test(res_apl_new)
#-Linealidad
reset(modelo_apl_new)

#Test de Breush-Godfrey
bgtest(mod_apl_new)
```

Bibliografía

- [1] Comisión Jurídica y Social de la Mujer. (s.f.). <https://www.unwomen.org/es/csw/csw-snapshot>
- [2] ERAC Standing Working Group on Gender in Research and Innovation. (28 de Marzo de 2019). ST-1204-2019-INI <https://data.consilium.europa.eu/doc/document/ST-1204-2019-INIT/en/pdf>
- [3] Figures, S. (2018). Gender Equality in Research and Innovation in the European Union: Doctoral Graduates. 2018.
- [4] Mihaljević, H., & Santamaría, L. (2020). Authorship in top-ranked mathematical and physical journals: Role of gender on self-perceptions and bibliographic evidence. *Quantitative Science Studies*, 1-25.
- [5] Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19), 3055-3071.
- [6] Muggeo, V. M., & Muggeo, M. V. M. (2017). Package ‘segmented’. *Biometrika*, 58(525-534), 516.
- [7] Mujeres y Matemáticas. (30 de julio de 2020). Stop Discriminación. <https://mym.rsme.es>
- [8] ONU Mujeres. (2013). Un objetivo transformador e independiente para lograr la igualdad de género, los derechos y el empoderamiento de las mujeres: imperativos y componentes clave. <https://www.unwomen.org/es/news/in-focus/mdg-momentum>
- [9] Organización de las Naciones Unidas. (2020). gender equality womens rights in review key facts and figures <https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2020/gender-equality-womens-rights-in-review-key-facts-and-figures-en.pdf?la=en&vs=935>

- [10] Rebollar Sánchez, E. M. (2013). El género en los planes de estudio de los grados de educación en las universidades públicas españolas. Universitat Autònoma de Barcelona,.
- [11] Rossiter, M. W. (1993). The Matthew Matilda effect in science. *Social studies of science*, 23(2), 325-341.
- [12] Saavedra-Nieves, P. & Saavedra-Nieves, A. (2021). A perspectiva de xénero na investigación en Estatística e Investigación de Operacións, 57, 2-5.
- [13] Schiebinger, L., Klinge, I., Sánchez de Madariaga, I., Paik, H. Y., Schraudner, M., & Stefanick, M. (2011-2019). Gendered innovations in science, health & medicine, engineering, and environment. <https://genderedinnovations.stanford.edu>