

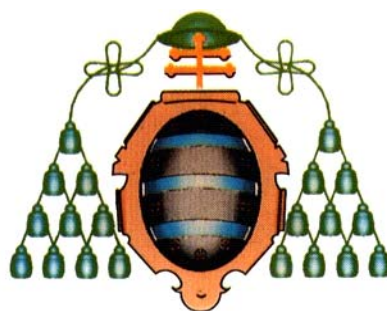
UNIVERSIDAD DE OVIEDO

Avances en Informática

Tesis Doctoral

**CARACTERIZACIÓN DE LA POPULARIDAD EN
SERVICIOS DE VÍDEO BAJO DEMANDA
APLICADA A PERIÓDICOS DIGITALES**

María Teresa González Aparicio



UNIVERSIDAD DE OVIEDO

Avances en Informática

**CARACTERIZACIÓN DE LA POPULARIDAD EN
SERVICIOS DE VÍDEO BAJO DEMANDA
APLICADA A PERIÓDICOS DIGITALES**

Tesis Doctoral

Autora:

María Teresa González Aparicio

Directores:

Roberto García Fernández

Xicu Xabiel García Pañeda

Jose Antonio López Brugos

DEPARTAMENTO DE INFORMÁTICA

Agradecimientos

En primer lugar, desearía expresar mi gratitud a los profesores Roberto García Fernández, Xicu Xabiel García Pañeda y Jose Antonio López Brugos, que fueron los máximos responsables de que esta Tesis Doctoral se hubiese podido llevar a cabo. Su colaboración y aportación de ideas han sido clave para el desarrollo de este trabajo. Especialmente y de forma muy entrañable, desearía resaltar la confianza que Jose Antonio López Brugos depositó en mí para que este proyecto llegase a buen término, siempre animándome a trabajar de forma incondicional, nunca tendré suficientes palabras de gratitud.

Al Vicerrector de Profesorado, Julio Antonio González García, por haberme ayudado a tener la oportunidad de seguir trabajando en mi carrera académica, sin desvincularme del ámbito universitario.

A todos los compañeros del Área de Ingeniería Telemática. Así como, a todas aquellas personas que siempre trataron de animarme constantemente. A todos ellos, muchísimas gracias por empujarme a seguir adelante.

A toda mi familia con el mayor cariño del mundo por apoyarme y ayudarme en todo momento, gracias por estar siempre ahí. En especial, a mis padres, excelentes personas, por su ayuda incondicional a lo largo de toda la vida; a mi marido, Jose, y a mis hijos, Jose Daniel y Manuel Enol, por ser siempre tan cariñosos conmigo.

Finalmente, me gustaría dedicar esta tesis a todas las mujeres que tienen que luchar día tras día por conciliar su vida familiar con su vida profesional.

Resumen

El contenido de tipo audio y vídeo presenta unas características especiales, que hacen que su distribución a través de Internet sea diferente respecto a otro tipo de objetos, como pueden ser las imágenes. El objetivo es conseguir una máxima eficacia y/o eficiencia en su distribución a través de la red. A lo largo de la literatura son muchas las técnicas que pretenden abordar el problema, desde redes específicas de distribución de contenido hasta una red de servidores, entre otras. Los factores que presentarían una notable influencia, para una recepción del contenido óptima, son variados, entre ellos cabría mencionar el tiempo de espera hasta que el usuario pueda visualizar el contenido, el ancho de banda de la línea que determinaría la calidad del vídeo que el usuario podría descargar, la proximidad del contenido solicitado al usuario, el tráfico en la red, etc. Por lo tanto, el control sobre cada uno de los factores mencionados con anterioridad se plantea necesario, aunque en algunas ocasiones pueda no resultar suficiente.

Esta tesis se enmarca en el contexto de un servicio de vídeo bajo demanda (VoD), que tiene como fin proporcionar vídeos relativos a noticias de distinta naturaleza, presentes en una serie de periódicos digitales de ámbito autonómico, los cuales ya son conocidos a nivel de prensa escrita en la comunidad correspondiente, como son "La Nueva España" (Asturias), "La Opinión A Coruña" (A Coruña), "Faro de Vigo" (Vigo), "La Provincia" (Islas Canarias), "Levante-EMV" (Valencia) y "Superdeporte" (Valencia). Sin embargo, a pesar de ser periódicos regionales comparten una serie de características que los hacen destacar en su ámbito, como es la diversidad de usuarios que los visitan, introducción diaria de nuevo contenido y una amplia variedad de vídeos relativos a noticias correspondientes a distinta temática y clasificados visualmente para el usuario en función de la misma. Por consiguiente, en particular esta tesis estaría enmarcada en un servicio de noticias bajo demanda "*News-on-Demand*" (NoD). Dado el incipiente uso de Internet, la consulta de noticias vía online es cada vez mayor, y por lo tanto una distribución eficiente y eficaz del contenido hacia el usuario se plantea como una necesidad cada vez más eminente.

La temática que se pretende abordar es una parte específica dentro de esta problemática tan compleja, como es la caracterización de la demanda sobre los vídeos presentes en periódicos digitales, o dicho de otra forma lo que se ha dado a conocer en

gran parte de la literatura como la "*popularidad*" de los vídeos. A lo largo de los diferentes artículos que han sido analizados, las principales distribuciones estadísticas aplicadas para su caracterización han sido variadas, tales como Zipf-like, two Zipf-like, Mandelbrot y Stretched. Una aportación de este trabajo frente a investigaciones previas es la caracterización de la popularidad mediante una distribución denominada "*transformación de Box-Cox*", donde en casos concretos aportó una mejora frente a las distribuciones anteriores. Cada una de las distribuciones que han sido mencionadas fueron aplicadas sobre la demanda de los vídeos en cada uno de los seis periódicos mencionados anteriormente, y se realizó una comparativa de los resultados. El uso de técnicas estadísticas básicas como la estadística descriptiva, la aplicación de las distribuciones pertinentes, el cálculo de la bondad de ajuste y los contrastes de hipótesis se hacen necesarios tanto para el cálculo de los modelos como para su validación.

Sin embargo, esta tesis se inició con el estudio de la popularidad de los vídeos en uno de los primeros diarios digitales de noticias en España, Asturias.com, siendo el más popular en términos absolutos entre los periódicos digitales que ofrecían información en lengua asturiana. La frecuencia de incorporación de nuevos vídeos aproximadamente era de uno por semana, característica que lo hacía singular, pues el hecho de introducir contenidos de forma tan espaciada en el tiempo hacía que vídeos con una gran demanda siguiesen siendo solicitados pese a la introducción de nuevos contenidos. En este periódico, el valor del parámetro θ de la distribución Zipf-like resultó ser un buen estimador de la popularidad en periodos de estabilidad del contenido.

Otra aportación importante ha sido el análisis y modelado de patrones de redundancia presentes en secuencias de peticiones, es decir, en que medida se concentran las peticiones sobre los diferentes vídeos. Así, cuando las peticiones se concentran sobre unos pocos vídeos se trataría de una redundancia fuerte, en caso contrario si las peticiones se encuentran más dispersas se trataría de una redundancia débil. El grado de redundancia ha sido calculado mediante un método de redundancia global (gR) y otro parcial (pR). La aplicación de estos métodos es nueva en este tipo de servicios. La principal contribución de este análisis ha sido la obtención de un modelo de redundancia parcial global para cada periódico digital, el cual permite predecir el grado de peticiones de los vídeos que tiene una alta probabilidad de repetirse en un futuro cercano. Una aportación importante es que el modelo resultó ser independiente tanto del usuario como del tiempo.

Abstract

The audio/video content have special characteristics, which implies a different treatment from the rest of the objects delivered on the Internet, like images. The goal is to reach the maximum effectiveness and/or efficiency in its travel through the network. Along the literature, there are many techniques oriented to tackle this problem, from specific content distribution networks to a server network among others. There are many factors, which could have a big influence on the quality of the service, such as the packet delay variation (jitter), the proximity of the content to the user or the Internet traffic. Therefore, it is necessary to have each factor under control, although it could not be enough in many cases.

This thesis has been developed in a Video-on-Demand (VoD) context. The aim of this specific type of service is to provide videos related to pieces of news from different Spanish digital newspapers with a regional scope, which are well-known in the region they belong to. In our work, six Spanish digital newspapers, such as “La Nueva España” (Asturias), “La Opinion A Coruña” (A Coruña), “Faro de Vigo” (Vigo), “La Provincia” (Islas Canarias), “Levante-EMV” (Valencia) and “Superdeporte” (Valencia), have been analyzed. However, although they are linked to a specific region from Spain, they share some common and remarkable features, which are the diversity of users, a daily introduction of new content, and a wide variety of videos related to different topics with a visual classification for the user according to the type of news they are associated to (national, international, sports, etc). Therefore, specifically this thesis has been focused on a News-on-demand (NoD) service.

Due to the growth of Internet, the queries of pieces of videos related to news have increased. Therefore, an effectiveness and efficiency content delivery is required to reach an optimal coverage of user needs. The theme to deal with, it is a specific part inside VoD service, which is the video demand characterization in digital newspapers, which is known in the literature as “video popularity”. Along the different research papers, some statistical distributions have been applied, such as Zipf-like, two Zipf-like, Mandelbrot and Stretched. A new contribution of this thesis is the video popularity characterization with a statistical distribution called “Box-Cox transformation”, where in some cases provided better results than the previous distributions. All of them have been analyzed in the six digital newspapers and a comparison of results has been done.

The use of basic statistical techniques like the descriptive statistics, the application of some distributions, the goodness of fit and hypothesis tests were analyzed as the model calculation as the validation process.

However, this thesis begins with the study of video popularity characterization in one of the first digital newspapers from Spain in Asturian language, “Asturies.com”. The content introduction frequency was one per week. Therefore, videos with a high demand were still highly requested despite of the arrival of new videos. In this service, the parameter θ turned out a good estimator in periods where the content remained stable.

Another important contribution has been the analysis and modeling of redundancy patterns in sequences of video requests, that is how the requests are spread along the different videos. Therefore, when requests are concentrated on a few videos reflects a strong redundancy, on the contrary when requests are more dispersed show a weak redundancy. The level of redundancy has been measured with a global redundancy method (gR) and another partial redundancy method (pR). The application of these methods is new in this type of services. The main contribution of this analysis has been a global partial redundancy model for each of the six digital newspapers. This model would allow to forecast the level of redundancy for videos with a high probability of being requested again in a near future. The model was independent of the user and time.

ÍNDICES

ÍNDICE GENERAL

Capítulo 1. INTRODUCCIÓN	1
1.1. Introducción.....	3
1.2. Entorno y motivación de la tesis	3
1.3. Objetivos de la tesis.....	6
1.4. Estructura de la tesis	9
Capítulo 2. ESTADO DEL ARTE	11
2.1. Introducción.....	13
2.2. Caracterización de la popularidad	14
2.2.1. Distribución Zipf-like	15
2.2.2. Distribución Zipf-like generalizada.....	25
2.2.3. Combinación de dos Zipf-like (two Zipf-like)	26
2.2.4. Distribución Mandelbrot	30
2.2.5. Distribución Stretched	30
2.3. Concentración de la demanda.....	32
2.4. Redundancia global y parcial	34
Capítulo 3. ANÁLISIS DEL PERIÓDICO DIGITAL Asturias.com.....	37
3.1. Introducción.....	39
3.2. Análisis de la popularidad	40
3.2.1. Baja popularidad en el vídeo anterior y en el nuevo	41
3.2.2. Baja popularidad en el vídeo anterior y alta en el nuevo.....	43
3.2.3. Alta popularidad en el vídeo anterior y baja en el nuevo	45
3.2.4. Alta popularidad en el vídeo anterior y alta en el nuevo	47
3.3. Conclusiones.....	48
Capítulo 4. CARACTERÍSTICAS DE LOS PERIÓDICOS DIGITALES BAJO ANÁLISIS.....	51

4.1. Introducción.....	53
4.2. Ámbito de aplicación de la popularidad.....	55
4.3. Características de los periódicos digitales.....	57
4.3.1. Desde el punto de vista de las secciones.....	57
4.3.2. Desde el punto de vista de los días de la semana.....	63
4.4. Conclusiones.....	68
Capítulo 5. RESULTADOS DEL AJUSTE MEDIANTE DISTINTAS DISTRIBUCIONES ESTADÍSTICAS.....	71
5.1. Introducción.....	73
5.2. Distribución Zipf-like.....	75
5.2.1. Periódico "La Opinión A Coruña".....	76
5.2.2. Periódico "Faro de Vigo".....	77
5.2.3. Periódico "La Provincia".....	78
5.2.4. Periódico "Levante-EMV".....	79
5.2.5. Periódico "La Nueva España".....	80
5.2.6. Periódico "Superdeporte".....	81
5.3. Combinación de dos Zipf-like (two Zipf-Like).....	82
5.3.1. Periódico "La Opinión A Coruña".....	83
5.3.2. Periódico "Faro de Vigo".....	84
5.3.3. Periódico "La Provincia".....	85
5.3.4. Periódico "Levante-EMV".....	85
5.3.5. Periódico "La Nueva España".....	87
5.3.6. Periódico "Superdeporte".....	89
5.4. Distribución Mandelbrot.....	90
5.4.1. Periódico "La Opinión A Coruña".....	91
5.4.2. Periódico "Faro de Vigo".....	91
5.4.3. Periódico "La Provincia".....	91

5.4.4. Periódico "Levante-EMV"	92
5.4.5. Periódico "La Nueva España"	93
5.4.6. Periódico "Superdeporte"	96
5.5. Distribución Stretched	97
5.5.1. Periódico "La Opinión A Coruña"	97
5.5.2. Periódico "Faro de Vigo"	98
5.5.3. Periódico "La Provincia"	99
5.5.4. Periódico "Levante-EMV"	100
5.5.5. Periódico "La Nueva España"	101
5.5.6. Periódico "Superdeporte"	102
5.6. Transformación de Box-Cox	102
5.6.1. Periódico "La Opinión A Coruña"	104
5.6.2. Periódico "Faro de Vigo"	106
5.6.3. Periódico "La Provincia"	108
5.6.4. Periódico "Levante-EMV"	110
5.6.5. Periódico "La Nueva España"	111
5.6.6. Periódico "Superdeporte"	113
5.7. Variabilidad del periodo	113
5.7.1. Distribución Zipf-like	114
5.7.2. Distribución Mandelbrot	116
5.7.3. Distribución Stretched	119
5.8. Concentración de la demanda	122
5.9. Conclusiones	128
Capítulo 6. VALIDACIÓN DE LOS MODELOS	131
6.1. Introducción	133
6.2. Validación de los modelos	133

6.2.1. Generación de poblaciones sintéticas	133
6.2.2. Tests de bondad de ajuste	134
6.2.3. Validación de las poblaciones sintéticas	137
6.3. Validación de los modelos.....	139
6.4. Conclusiones.....	143
Capítulo 7. CARACTERIZACIÓN DE LA POPULARIDAD MEDIANTE REDUNDANCIAS PARCIALES.....	145
7.1. Introducción.....	147
7.2. Métodos para el estudio de redundancias	147
7.3. Análisis de la redundancia.....	151
7.4. Validación.....	158
7.5. Conclusiones.....	160
Capítulo 8. CONCLUSIONES Y LINEAS DE TRABAJO FUTURAS	161
8.1. Introducción.....	163
8.2. Conclusiones.....	163
8.3. Trabajos futuros.....	165
8.4. Publicaciones derivadas de esta tesis	166
8.5. Otras publicaciones	167
Capítulo 9. REFERENCIAS	169
9.1. Referencias	171
ANEXO I. Datos específicos de la distribución Zipf-like	175
ANEXO II. Datos específicos de la distribución Two Zipf-like	195
ANEXO III. Datos específicos de la distribución Mandelbrot.....	209
ANEXO IV. Datos específicos de la distribución Stretched.....	227
ANEXO V. Datos específicos de la distribución Box-Cox.....	241

ÍNDICE FIGURAS

Figura 1. Pares de secuencias involucradas en el cálculo de pR.....	35
Figura 2. Parte superior de la página inicial del periódico digital Asturias.com.....	39
Figura 3. Página del servicio de VoD del periódico Asturias.com.	40
Figura 4. Ranking de accesos en el Periodo 1.	42
Figura 5. Valor de theta en el Periodo 1.	43
Figura 6. Ranking de accesos en el Periodo 2.	43
Figura 7. Valor de theta en el Periodo 2.	44
Figura 8. Ranking de accesos en el Periodo 3.	45
Figura 9. Valor de theta en el Periodo 3.	46
Figura 10. Ranking de accesos en el Periodo 4.	47
Figura 11. Ranking de accesos en el Periodo 4.	48
Figura 12. Distribución de la frecuencia de accesos (%) en cada periódico digital.	53
Figura 13. Página categoría/subcategoría "Sociedad" del periódico digital "La Nueva España".....	54
Figura 14. Agrupación de las categorías en secciones en los seis periódicos digitales..	56
Figura 15. Porcentaje de accesos por sección en cada periódico digital.	60
Figura 16. Nº de accesos diarios en los distintos periódicos digitales.	67
Figura 17. Esquema para la obtención de patrones de comportamiento.	74
Figura 18 Porcentaje de días modelados con Zipf-like.	75
Figura 19. Distribución Zipf-like con $\theta = 1,5971$, $R = 0,9498$	76
Figura 20. Test F-Anova para θ entre pares de días de la semana.	77
Figura 21. Test F-Anova para θ entre pares de días de la semana.	79
Figura 22. Test F-Anova para θ entre pares de días de la semana.	80
Figura 23. Ajuste con el modelo general del miércoles y el específico del día.....	84
Figura 24. Ajuste de accesos con dos two Zipf-like.....	86
Figura 25. Ajuste de accesos con dos two Zipf-like.....	87
Figura 26. Test F-Anova para θ entre pares de días de la semana.	88
Figura 27. Ajuste con dos Zipf-like, modelo general y específico.....	89
Figura 28. Ajuste con modelo general, modelo general del martes y específico.	90
Figura 29. Test F-Anova para k entre pares de días de la semana.	92

Figura 30. Test F-Anova para k entre pares de días de la semana.	94
Figura 31. Test F-Anova para θ , entre pares de días de la semana.	96
Figura 32. Test F-Anova para el parámetro a.	97
Figura 33. Resultados del F-Anova test.	98
Figura 34. Test F-Anova del parámetro a.	100
Figura 35. Resultado de F-Anova para el parámetro a.	101
Figura 36. Transformación de Box-Cox con tres modelos.	105
Figura 37. Transformación de Box-Cox con tres modelos en la sección S3.	106
Figura 38. Test F-Anova para el parámetro "a" entre pares de días de la semana.	107
Figura 39. Transformación de Box-Cox con tres modelos.	108
Figura 40. Transformación de Box-Cox con tres modelos.	109
Figura 41. Comparación del número medio de vídeos accedidos entre dos secciones.	110
Figura 42. Transformación de Box-Cox con cuatro modelos.	111
Figura 43. Test F-Anova para el parámetro " λ " entre pares de días de la semana.	112
Figura 44. Distribución de peticiones con distintos intervalos de tiempo.	114
Figura 45. Número de peticiones (%) destinadas a nuevos vídeos (%) en su primer día de vida.	123
Figura 46. Proceso de generación de poblaciones sintéticas.	134
Figura 47. Ajuste con tres distribuciones de tipo Stretched.	136
Figura 48. Ajuste con tres distribuciones de tipo Stretched.	137
Figura 49. Proceso de validación de modelos.	138
Figura 50. Test F-Anova para los parámetros a y b de la distribución de Box-Cox. ...	139
Figura 51. Pares de secuencias que forman parte del cálculo de pR.	148
Figura 52. pR en "La Nueva España"	152
Figura 53. Validación con $\Delta = 0$	152
Figura 54. pR con $\Delta = 0$	154
Figura 55. pR en "La Nueva España"	156
Figura 56. Valores pendiente y origen.	156
Figura 57. pR "La Opinión A Coruña", B=162.	157
Figura 58. pR "Faro de Vigo", B=264.	157
Figura 59. pR "Levante-EMV", B=534.	157
Figura 60. pR "La Provincia", B=1.160.	157
Figura 61. pR "La Nueva España", B=2.244.	157

Figura 62. pR "Superdeporte", B=2.292.....	157
Figura 63. pR "La Opinión A Coruña", B=162.....	158
Figura 64. pR "Faro de Vigo", B=264.....	158
Figura 65. pR "Levante-EMV", B=534.....	158
Figura 66. pR "La Provincia", B=1.160.....	158
Figura 67. pR "La Nueva España", B=2.244.....	159
Figura 68. pR "Superdeporte", B=2.292.....	159

ÍNDICE DE TABLAS

Tabla 1. Caracterización de la popularidad.	16
Tabla 2. Resumen estadístico de los valores del parámetro θ	21
Tabla 3. Valores de los parámetros con k-transformación para las cargas HPC y HPL.	26
Tabla 4. Valores típicos de los parámetros para BIBS y eTeach.	28
Tabla 5. Valores típicos de los parámetros.	29
Tabla 6. Pseudocódigo para el análisis de la redundancia parcial entre pares de secuencias.	35
Tabla 7. Casos de estudio ante la incorporación de un vídeo nuevo.	41
Tabla 8. Asignación de colores por categoría.	57
Tabla 9. N° de vídeos accedidos.	58
Tabla 10. N° de accesos.	59
Tabla 11. Número de vídeos creados por secciones.	61
Tabla 12. Resultados del coeficiente de Pearson (r).	62
Tabla 13. Número medio de vídeos y accesos en "La Opinión A Coruña".	64
Tabla 14. Número medio de vídeos y accesos en el "Faro de Vigo".	64
Tabla 15. Número medio de vídeos y accesos en el "La Provincia".	64
Tabla 16. Número medio de vídeos y accesos en "Levante-EMV".	65
Tabla 17. Número medio de vídeos y accesos en "La Nueva España".	65
Tabla 18. Número medio de vídeos y accesos en "Superdeporte".	65
Tabla 19. Número de vídeos creados por días de la semana de Enero a Junio.	68
Tabla 20. Algoritmo de búsqueda de la mejor partición para las peticiones.	83
Tabla 21. Variación del parámetro θ con periodos de distinta longitud.	115
Tabla 22. Número de grupos modelados con periodos de distinta longitud.	116
Tabla 23. Número de grupos modelados con periodos de distinta longitud.	117
Tabla 24. Número de grupos modelados con periodos de distinta longitud.	118
Tabla 25. Número de grupos modelados con periodos de distinta longitud.	120
Tabla 26. Número de grupos modelados con periodos de distinta longitud.	121
Tabla 27. Porcentaje de accesos sobre el conjunto de vídeos de cada periódico digital.	124
Tabla 28. Datos relativos a los vídeos creados.	127
Tabla 29. Datos relativos a los vídeos creados.	127

Tabla 30. Datos relativos a los vídeos creados.....	128
Tabla 31. Número de días con el test de Chi-cuadrado / Kolmogorov-Smirnov positivo.	140
Tabla 32. Número de días con el test de Kolmogorov-Smirnov y Chi-cuadrado positivo en el modelo general.....	142
Tabla 33. Número de días con el test de Kolmogorov-Smirnov y Chi-cuadrado positivo en el modelo asociado a cada día de la semana.....	143
Tabla 34. Algoritmo para calcular la redundancia parcial.....	149
Tabla 35. Algoritmo para el cálculo de la redundancia parcial final.....	150
Tabla 36. Número de accesos y vídeos en los dos semestres del año 2010.	151
Tabla 37. Tamaño B en los periódicos digitales.....	152
Tabla 38. Resultados de gR.	153
Tabla 39. Parámetros específicos para cada periódico digital.....	155
Tabla 40. Valores de Δ , γ y β en los seis periódicos digitales.....	156

Capítulo 1.

INTRODUCCIÓN

1.1. Introducción

Internet se ha convertido en una de las principales fuentes de información a nivel mundial para la difusión de contenido de distinta índole (texto, imágenes, audio y vídeo). En un principio, el contenido publicado por los distintos sitios web era texto e imágenes, sin embargo la incorporación de contenido multimedia (audio/vídeo) ha tomado un papel predominante, sobre todo con la llegada de redes sociales tan populares como YouTube, Yahoo o Facebook. Así pues, la distribución de contenido multimedia de forma eficaz y/o eficiente se ha convertido en uno de los principales pilares a tener en cuenta en el diseño de distintas arquitecturas de red [1, 2]. El objetivo sería conseguir un mejor reparto de recursos y reducir el impacto que podría causar a la red un incremento de la demanda [3], con alternativas tales como la realización de réplicas (caching o prefetching) de los objetos más populares en determinados servidores [4, 5], multicast, patching o transmisión de broadcast de forma periódica, entre otras. Por este motivo, hay diversos estudios orientados a la construcción de herramientas que intentan generar cargas sintéticas con el fin de poder evaluar distintos diseños de distribución de este tipo de contenido, entre ellos cabría destacar GISMO [6], MediSyn [7] y SURGE [8]. Dichas herramientas modelan la distribución de peticiones sobre los contenidos, y por otro lado el comportamiento del usuario con cada contenido, es decir, el tamaño del contenido solicitado, tipo de codificación, así como la forma de interactuar con el mismo (pausas, retrocesos y/o avances) [9]. Algunos de los objetivos concretos serían la predicción de posibles cuellos de botella en el sistema, determinar el contenido de la caché, mantener y/o actualizar las listas de recomendaciones, y ayudar a los motores de búsqueda. Además, en lo que respecta a la caracterización de los patrones de demanda sobre el contenido, cabe destacar que dependerá en gran medida de la naturaleza del mismo, pues la filosofía de acceso sobre páginas web, vídeos o bien contenido de tipo "*live streaming*" es completamente distinta.

1.2. Entorno y motivación de la tesis

Esta tesis está centrada en un entorno de vídeo bajo demanda (Video-on-Demand, VoD). A veces los términos televisión interactiva y servicios bajo demanda

son considerados sinónimos del mismo, e incluso otros términos como "movies on demand" o "interactive news television" entre otros han sido considerados como subcategorías [10]. En concreto, esta tesis se ha centrado en un servicio de noticias bajo demanda (News-on-Demand, NoD), dado que el estudio ha sido focalizado hacia seis periódicos digitales de ámbito autonómico, que ofrecen vídeos relativos a noticias de diversa índole clasificadas en secciones, tales como el mundo del deporte, ámbito local, nacional e internacional, sociedad y cultura, entre otras. Estos periódicos ya son conocidos a nivel de prensa escrita en las cuatro comunidades autónomas a las que pertenecen, como son Asturias, Islas Canarias, Galicia y Valencia. Como se ha comentado anteriormente, la distribución de contenido eficaz y/o eficiente es compleja y engloba muchas características. Por lo tanto, esta tesis se centrará en una parte específica dentro de esta problemática, como es la caracterización del patrón de demanda sobre los vídeos, o lo que en la literatura es ampliamente conocido como la "popularidad" de los vídeos. La popularidad de un vídeo se podría definir como el número de veces que dicho vídeo es solicitado en un intervalo de tiempo determinado. En general, hay que tener presente que la caracterización de la popularidad podría verse distorsionada debido a factores externos a la naturaleza intrínseca del propio vídeo, entre los cuales se podrían mencionar los siguientes:

1. En muchos casos la aplicación cliente (aplicación que se ejecuta en el equipo del usuario) mantiene una caché con el resultado de algunas de las peticiones efectuadas, siguiendo alguna política de caching [11]. Por lo tanto, si el usuario realizase una petición ya solicitada con anterioridad, cuyo resultado estuviese en la caché, dicha petición sería despachada de forma local en el propio usuario, no quedando constancia de la misma en el servidor correspondiente.
2. La existencia de proxies con caché que recogen las peticiones de los usuarios, origina el mismo efecto que ha sido comentado en el apartado anterior. Dichos proxies atienden las peticiones de los vídeos o páginas web más demandadas.
3. A veces cuando un usuario hace una solicitud sobre un vídeo o una página web, se arranca automáticamente otro vídeo, tal es el caso de los vídeos de

tipo publicitario, sobre el cual no se había realizado solicitud alguna, sin embargo de este vídeo extra quedaría registrada su aparición en el servidor como una solicitud más por parte del usuario. Por lo tanto, existen entradas en el fichero de registros de accesos del servidor que no reflejarían peticiones reales, y que por lo tanto estarían introduciendo ruido en el modelado de la popularidad.

4. La presencia de listas de recomendación o "*pre-filtros*", así como los motores de búsqueda o "*post-filtros*" tienden a favorecer la elección de unos vídeos frente a otros, incluso algunos objetos podrían llegar a no ser conocidos nunca por los usuarios debido a la actuación de los filtros, fenómeno conocido como "*niche content*". Ambos factores, en mayor o menor medida, estarían distorsionando la caracterización de la demanda [12].

En particular, en el estudio que nos concierne se presentan otra serie de dificultades adicionales, tales como:

1. Cuando un vídeo es incorporado al sistema se le pone una fecha de creación, lo cual no significa que esté disponible al usuario. Cuando realmente se publica, se le pondría una fecha de publicación y pasaría a estar visible, y por lo tanto disponible para su acceso. Sin embargo, a veces el gestor del sistema podría cambiar el estado de un vídeo a no visible, y en una fecha posterior volverlo a poner visible, y por lo tanto se modificaría la fecha de publicación del mismo. Sin embargo, un vídeo se tiene en cuenta desde el mismo momento en que recibe un acceso con independencia de su fecha de publicación, y de sus intervalos de inaccesibilidad.
2. En algunas ocasiones, la base de datos presentaba inconsistencias de información, dado que existían vídeos con una fecha de publicación anterior a su fecha de creación.
3. Normalmente, en la parte derecha de la página inicial del periódico digital se presenta uno de los vídeos que son noticia. Por lo tanto, este vídeo tendría un trato preferente respecto al resto de los vídeos, dado que podría atraer la atención de un mayor número de usuarios, dando pie a la captación de un mayor número de accesos. Este hecho no ha podido ser constatado, dado que

no se disponía de información sobre que vídeos aparecen en portada y durante cuanto tiempo.

Tras el estudio de los diferentes artículos se ha observado que la caracterización de la popularidad de los vídeos se ha llevado a cabo con varias distribuciones estadísticas, siendo la más destacada Zipf-like, seguida de Stretched, two Zipf-like y Mandelbrot. En este estudio se han tenido en cuenta todas las distribuciones anteriores, así como un método alternativo conocido como "transformación de Box-Cox". El modelado, así como la medición de la bondad de ajuste y los contrastes de hipótesis para determinar la validez de los modelos, ha sido realizado a través de programas desarrollados en Matlab versión 7.0.1. y el paquete estadístico R versión 2.10.1., mediante el uso de los lenguajes de programación propios que incorporan.

Principalmente, el estudio de la popularidad ha sido caracterizado día a día, con dos enfoques, por un lado teniendo en cuenta el día de la semana concreto en que se han producido (lunes, martes, etc), y por otro lado la clasificación de los vídeos por categorías. Otro aspecto más que ha sido considerado, es la longitud del periodo, dado que también podría influir en la elección de la distribución más adecuada, pues si es largo, el número de vídeos con una popularidad similar aumentaría, sin embargo si el periodo es corto el número de datos disponibles podrían no ser suficientes. Por lo tanto, también se han tenido en cuenta periodos con una longitud de dos, tres y hasta un total de quince días. Por otro lado, el comportamiento de los usuarios ha sido analizado desde el punto de vista de la redundancia en los accesos considerando el periodo de tiempo en estudio como un todo (redundancia global), o bien sobre distintos periodos de tiempo sobre el total (redundancia parcial).

1.3. Objetivos de la tesis

La tesis tiene como objetivo la búsqueda de un modelo estadístico para la caracterización de la popularidad de los vídeos que son distribuidos por un servicio de noticias bajo demanda (NoD). En concreto, los vídeos representan noticias y son publicados en distintos periódicos digitales, los cuales se encuentran asociados a distintas comunidades autónomas de España. Dichos periódicos presentan las siguientes características:

1. Gran diversidad de usuarios.
2. Introducción diaria de nuevos vídeos.
3. Gran variedad de contenido.
4. Clasificación de los vídeos por categorías en función de su temática, con el fin de mejorar la "usabilidad" del sitio web [13, 14].
5. Ordenación de los vídeos de forma cronológica en función de su fecha de publicación, desde el más actual al menos actual, en grupos con un máximo de doce vídeos por página. El acceso a los vídeos restantes pertenecientes a la misma categoría se presentan en páginas sucesivas.

Para abordar el problema se ha establecido una serie de objetivos parciales, los cuales son detallados a continuación:

1. **Recogida y filtrado de datos.** A partir de los datos proporcionados por el operador TeleCable de Asturias S.A. sobre los accesos a los vídeos publicados por los periódicos digitales a los que da soporte, se seleccionarán aquellos que pertenezcan a los periódicos digitales que van a ser objeto de estudio. En concreto han sido seleccionados los seis periódicos digitales con el mayor número de accesos. Aunque cabe destacar que la tesis se inicio con el estudio de un diario digital llamado Asturias.com que presentaba un índice de publicación de un vídeo por semana aproximadamente.
2. **Análisis del periodo.** Dado que la caracterización de la popularidad podría variar en función de la longitud del periodo, según es corroborado en los artículos [15-17], se ha decidido dividir el periodo en slots de tiempo de tamaño uniforme, o lo que es lo mismo estudiar el periodo con distintas escalas de tiempo, pero principalmente el estudio se plantea día a día dado que se trata de periódicos, donde las noticias se insertan diariamente, aunque también se plantea un breve estudio de la demanda en grupos de dos días, y así sucesivamente hasta llegar a un estudio quincenal como máximo. Cualquier otra opción podría ser válida si fuese considerada pertinente para la gestión del contenido de la red.

Por otro lado, el estudio de la redundancia presente en una secuencia de peticiones de los vídeos se realizará sobre distintas particiones de la misma, tal y como será explicado en el Capítulo 7.

3. **Caracterización de la popularidad.** Para cada día se procederá al cálculo de los parámetros correspondientes a cada una de las distintas distribuciones estadísticas, que normalmente han sido aplicadas a lo largo de la literatura, para ajustar la distribución de la popularidad, entre las cuales cabe mencionar la distribución Mandelbrot, Stretched, two Zipf-like y Zipf-like. Por lo tanto, dado que el conjunto de distribuciones es variado, no cabe asumir un ajuste a una distribución específica. En esta tesis se plantea un método estadístico alternativo denominado "transformación de Box-Cox" que mejora los resultados en la mayoría de los casos frente a las distribuciones anteriores.

Además, en cada periódico digital se planteó un estudio para la búsqueda de redundancias en las secuencias de peticiones asociadas a cada uno de ellos. Por un lado, se consideró el periodo en estudio como un todo, con el resultado de una redundancia global. Por otro lado, se planteó una partición sobre la secuencia de peticiones del periodo total y se calculó la redundancia entre parejas de particiones, dando como resultado una redundancia parcial.

4. **Búsqueda de un modelo de comportamiento general.** El objetivo es encontrar el modelo estadístico que mejor ajusta el comportamiento de la demanda diaria en cada periódico digital. Si los ajustes son buenos (testeados a través de tests de contraste de hipótesis), se realizará una comparativa de los ajustes de los distintos modelos entre sí. Lógicamente, el modelo elegido será aquel que proporcione el máximo grado de ajuste.

Desde el punto de vista de la redundancia se plantea un modelo único asociado a cada periódico digital, independiente tanto del usuario como del tiempo.

5. **Validación de los modelos.** Se contrastarán los datos obtenidos por los distintos modelos con los datos reales pertenecientes a otro periodo de tiempo, mediante tests de contraste de hipótesis.

6. **Obtención de conclusiones.** A través de los distintos modelos se tratará de realizar estimaciones sobre la popularidad de los vídeos en un futuro cercano, y particularizada para cada uno de los periódicos digitales.

Finalmente, cabe señalar que la elaboración de esta tesis ha sido posible gracias a los datos presentes en una base de datos, que aportaban información en relación a las características de los vídeos que fueron publicados en los distintos periódicos digitales, entre las cuales cabría mencionar la fecha de creación, la fecha de publicación, el periódico digital en el cual fue insertado, y además en concreto la categoría de publicación. Sin embargo, dado el objetivo de este trabajo, el dato crucial sería el conocimiento sobre cuales de los vídeos publicados han recibido peticiones, así como el instante de tiempo en que fueron efectuadas.

1.4. Estructura de la tesis

Esta tesis se estructura en un total de ocho capítulos, a lo largo de los cuales serán abordados los siguientes aspectos: el establecimiento de la problemática, técnicas que han sido aplicadas para la solución de problemas similares, análisis del ámbito de estudio, aplicación de distintas técnicas para su solución, el análisis de los resultados y su validación, y finalmente las conclusiones y trabajos futuros. A continuación, se ofrece una visión general del contenido de los distintos capítulos.

En el capítulo 1 se ha resaltado algunos puntos generales sobre la problemática en la caracterización de la popularidad en un servicio de vídeo bajo demanda (VoD), los cuales han sido comentados a lo largo de parte de la literatura. Por otro lado, en el ámbito de las noticias bajo demanda (NoD), y en particular en los periódicos digitales de ámbito autonómico que han sido objeto de estudio en esta tesis, también se ha subrayado alguna característica que podría ocasionar distorsiones sobre la popularidad real de los videos. Además, también se han planteado cuales han sido las motivaciones y los objetivos de esta tesis.

En el capítulo 2 se hace un recorrido sobre cual ha sido y es la situación de la investigación en la caracterización de la popularidad en contenido multimedia. Aunque en algunas ocasiones también se hace referencia a la caracterización de la demanda sobre páginas web, dado que fue su precursora.

En el capítulo 3 se presenta el primer estudio realizado en esta tesis sobre la popularidad de los vídeos publicados en uno de los primeros periódicos digitales de noticias en España, Asturias.com, y en particular en lengua asturiana.

En el capítulo 4 se explica por un lado que periódicos digitales fueron seleccionados para su estudio y el por qué, así como la organización de los distintos vídeos que el usuario tiene a su disposición en la página web del periódico correspondiente. Por otro lado, también se plantea la forma en que ha sido abordada la problemática de la caracterización de la popularidad en un servicio de tipo NoD.

En el capítulo 5 se explican los resultados obtenidos en cada periódico digital en el modelado de la popularidad de sus vídeos, a través de las distribuciones estadísticas Zipf-like, two Zipf-like, Mandelbrot y Stretched, y finalmente con la transformación de Box-Cox. Dicha caracterización ha sido realizada día a día con dos enfoques, por un lado se plantea un estudio de la popularidad de los vídeos en su conjunto, y por otro lado, tras una clasificación de los vídeos por temática, se plantea la caracterización de la popularidad dentro de la misma. Los modelos finales alcanzados, junto con el resultado de algunos tests estadísticos, vienen reflejados en el anexo correspondiente a cada una de las distribuciones. Finalmente, dado que la caracterización de la demanda a lo largo de los distintos vídeos está relacionada con el reparto de las peticiones, al final del capítulo se ha planteado una aproximación al análisis sobre la concentración de la demanda, a veces referido en la literatura como localidad temporal.

En el capítulo 6 se presenta los resultados de la validación de los distintos modelos que fueron planteados en el capítulo 5. En el capítulo 7 se realiza una caracterización sobre la redundancia de referencias a vídeos que se encuentra presente en cada una de las secuencias de peticiones asociadas a los distintos periódicos digitales. Finalmente, en el capítulo 8 se exponen las distintas conclusiones que han sido alcanzadas y los posibles trabajos futuros.

Capítulo 2.
ESTADO DEL ARTE

2.1. Introducción

Internet se ha convertido en una de las principales fuentes de información a nivel mundial para la difusión de contenido de distinta índole (texto, imágenes, audio y video). En un principio, el contenido publicado por los distintos sitios web era texto e imágenes, sin embargo la incorporación de contenido multimedia (audio/vídeo) ha tomado un papel predominante, sobre todo con la llegada de redes sociales tan populares como YouTube, Yahoo o Facebook. Así pues, la distribución de contenido multimedia de forma eficaz y/o eficiente se ha convertido en uno de los principales pilares a tener en cuenta en el diseño de distintas arquitecturas de red [1, 2, 18, 19]. Por este motivo, hay diversos estudios orientados a la construcción de herramientas que intentan generar cargas sintéticas con el fin de poder evaluar distintos diseños de distribución de este tipo de contenido, entre ellos cabría destacar GISMO [6], MediSyn [7] y SURGE [8]. Dichas herramientas modelan la distribución de peticiones sobre los contenidos, tanto desde el punto de vista de la popularidad como de su localidad temporal, y por otro lado el comportamiento del usuario con cada contenido, es decir, el tamaño del contenido solicitado, tipo de codificación, así como la forma de interaccionar con el mismo (pausas, retrocesos y/o avances). Algunos de los objetivos concretos serían la predicción de posibles cuellos de botella en el sistema, determinar el contenido de la caché, mantener y/o actualizar las listas de recomendaciones, y ayudar a los motores de búsqueda.

La caracterización del patrón de demanda sobre contenido multimedia o bien sobre páginas web es importante, pero diferente entre sí. Un factor crucial que influye directamente en la demanda y que diferencia notablemente el contenido multimedia de las páginas web es que en el primero de los casos el contenido no varía, mientras que en el segundo está siempre en proceso de cambio. Por lo tanto, es normal que los usuarios accedan a la misma página web muchas veces ("*fetch-repeatedly*"), y que en la mayoría de las ocasiones el usuario solicite el mismo contenido multimedia sólo una vez o como mucho dos, este efecto se ha dado a conocer en la literatura como "*fetch-at-most-once*". Este fenómeno ha sido puesto de manifiesto en [11], donde se estudió los accesos realizados en la universidad de Washington durante un periodo de 200 días, cuyo contenido multimedia se distribuía mediante P2P. En varios estudios ha sido resaltada

esta característica, pero no se comentó a que había sido debido, entre ellos se podría mencionar [20-23]. Otro motivo de la aparición de "*fetch-at-most-once*" podría ser debida a la longitud del periodo en estudio, el cual si es largo, el número de vídeos con una popularidad similar aumentaría, sin embargo si el periodo es corto el número de datos disponibles podría no ser lo suficientemente bueno como para concluir la distribución más adecuada.

2.2. Caracterización de la popularidad

A lo largo de la literatura son muchas las distribuciones estadísticas que han sido aplicadas con el fin de ajustar la demanda de peticiones sobre páginas web o bien sobre distintos objetos de streaming de diferente naturaleza, como pueden ser de tipo educacional, películas, noticias, deportes, etc. El hecho de que todo conjunto de peticiones no siga el mismo tipo de distribución puede ser debido a diversos factores, tales como el ámbito (universitario o empresarial), tipo de contenido (películas o educacional), frecuencia de cambio (página web o un vídeo), escala de tiempo (diaria, semanal, quincenal, etc). Por el contrario, cabe mencionar que no en todos los estudios se ha propuesto una distribución estadística como elemento a tener en cuenta para resolver el problema de la distribución de contenido de la forma más eficiente, de hecho en [24] se plantea una arquitectura de servidores orientada a la distribución de contenido multimedia, de manera que los vídeos que cada servidor almacenaría se decide de forma paulatina y dinámica, en función de las peticiones que se irían efectuando sobre los distintos vídeos, sin necesidad de un conocimiento previo sobre la popularidad de los mismos, y sin el planteamiento de una posible distribución a la cual se pudiese ajustar el patrón de demanda. Sin embargo, esta tesis está centrada en el establecimiento de posibles modelos estadísticos, cuyo objetivo sería el poder establecer predicciones sobre el comportamiento de un patrón de demanda. Por consiguiente, los siguientes apartados ofrecerán una descripción de las distintas distribuciones que han sido analizadas a lo largo de diversas investigaciones, así como los artículos en los cuales se plantea una aplicación de las mismas y que resultados fueron obtenidos en sus estudios.

2.2.1. Distribución Zipf-like

George Kingsley Zipf (1902–1950), lingüista y filólogo americano, estudió la ocurrencia estadística de las palabras a lo largo de distintos textos en diversas lenguas [25]. A partir de estos estudios estableció que si las palabras se ordenaban en orden decreciente en función de su frecuencia de aparición en el texto, se cumplía que la frecuencia de cualquier palabra era inversamente proporcional a su posición en el ranking, es decir, la primera palabra en el ranking sería usada dos veces más que la palabra que ocupa el segundo lugar, esta última sería usada tres veces más que la palabra que iría en tercera posición, y así sucesivamente. La idea que subyace es que muy pocas palabras tendían a ser usadas con mucha frecuencia, en contraposición con el resto donde su frecuencia de uso era escasa. A día de hoy, esta distribución ha sido aplicada en distintos ámbitos, como por ejemplo en la economía, y en el campo que nos concierne como es Internet, con el fin de modelar la demanda de páginas web, objetos multimedia, etc.

La distribución Zipf-like se define según la ecuación (1).

$$y = \frac{C}{x^\theta} \quad 1 \leq x \leq N, \quad \text{con} \quad C = \frac{1}{\sum_{x=1}^N \frac{1}{x^\theta}} \quad (1)$$

C es una constante de normalización y N representa el número de entidades. A través de esta distribución se establece que la frecuencia de acceso sobre la x-ésima entidad más popular es proporcional a $1/x^\theta$. El valor del parámetro θ determina la forma de la distribución, o lo que es lo mismo el nivel de concentración de los accesos sobre las primeras entidades. De esta forma, a medida que aumenta el valor de θ significaría que el número de accesos tiende a concentrarse más sobre menos entidades..

En el entorno de Internet, la distribución Zipf-like fue aplicada en un principio para caracterizar la popularidad de páginas web [26-30]. En [27] se estudió los accesos realizados sobre una caché instalada en DEC Palo Alto (California), siendo de entonces del orden de 4.000 peticiones diarias solicitadas por unos 100 usuarios. Las peticiones de páginas internas eran proporcionadas por servidores locales, mientras que las

externas eran servidas a través de una caché instalada para tal fin, o bien en su defecto solicitadas a un servidor remoto. Si se tenían en cuenta tanto las peticiones sobre páginas locales como externas, se observó que la distribución de los accesos se ajustaba a una distribución Zipf-like con un valor de $\theta=1$, es lo que se da a conocer en la literatura como distribución Zipf exacta o Ley de Zipf. Este hecho también se ha dado en [31] donde el valor de θ fue prácticamente igual a 1 ($\theta=0,986$), con algo más de medio millón de peticiones realizadas sobre un total de 46.830 páginas web, al igual que en [28] donde se estudió los accesos sobre 2.135 páginas web durante dos semanas, siendo el valor de θ muy próximo a 1 (0,85) con un grado de ajuste lineal del 0,99. También en [26] los accesos sobre 2.018 páginas se modelaron con un grado de ajuste muy fuerte ($R^2=0,99$) y un valor de θ próximo a la unidad ($\theta=0,95$). En [29] se estudiaron los accesos sobre páginas web en seis conjuntos de datos distintos, que pertenecían a varios ámbitos académicos o ISP's. Algunas características de los mismos, como son su ámbito, longitud del periodo en estudio, y el número de accesos, son presentados en la Tabla 1. La distribución de los accesos fue ajustada a una distribución Zipf-like, donde el valor del parámetro θ osciló entre 0,64 y 0,83. El ajuste resultó ser bueno, siempre y cuando se excluyesen las 100 primeras páginas más solicitadas. Por otra parte, se puntualiza que el ámbito pudiese tener cierta influencia sobre el valor de θ , siendo mayor en entornos donde el tipo de usuario es el mismo, como es el caso del ámbito académico con θ alrededor de 0,8, mientras que en ámbitos donde el tipo de usuarios es más heterogéneo el valor de θ fue más bajo, oscilando alrededor de 0,7. Además, se comprobó que factores como el tamaño de las páginas, y su frecuencia de cambio no tenían influencia en su popularidad.

Nombre	DEC	UPisa	FuNet	UCB	Questnet	NLANR
Ámbito	ISP	Académico	Académico	Académico	ISP	Académico
Periodo	7 días	3 meses	10 días	18 días	7 días	1 día
Nº Accesos	3.543.968	2.833.624	4.815.551	1.907.762	2.885.285	1.766.409
θ	0,77	0,78	0,83	0,69	0,73	0,64

Tabla 1. Caracterización de la popularidad.

Siguiendo con la demanda sobre páginas web en [30] se estableció una relación entre la popularidad y la localidad temporal (comentada en el apartado 2.3), pues una

página que sea muy popular tenderá a ser accedida con mayor frecuencia y por lo tanto el tiempo entre peticiones será menor (inter-request time). En este artículo se establece el siguiente teorema que relaciona ambos parámetros:

"Si la distribución de la popularidad de un documento en una cadena de peticiones sigue asintóticamente una ley exponencial con parámetro θ , donde $0,5 \leq \theta \leq 1$, entonces la distribución del tiempo entre peticiones en una permutación aleatoria de esta secuencia de peticiones puede ser caracterizada asintóticamente usando una ley exponencial con parámetro $(2-1/\theta)$."

Así pues, en este artículo se usan dos distribuciones exponenciales, por un lado la distribución (parámetro θ) que caracterizará la popularidad de las páginas web, y por otro lado la distribución (parámetro β) para la caracterización del tiempo entre peticiones. Tras el estudio de cuatro cargas durante un tiempo comprendido entre una semana y quince días aproximadamente en cada una de ellas, se obtuvieron los siguientes resultados: $\theta \in [0,66; 0,71; 0,72; 0,77]$ y $\beta \in [0,4; 0,46; 0,47; 0,64]$. Justifican la diferencia de los valores entre las cuatro cargas a dos razones: en primer lugar, tres de las cargas registran peticiones que provienen de clientes de distinta índole, y en segundo lugar la presencia de una caché en el lado del cliente hace que muchas de sus peticiones no sean registradas, dado que la petición es servida por la propia caché. En [32] se estudia el efecto que producen las cachés de páginas web, centrándose en aquel conjunto de peticiones que siguen una distribución Zipf-like. En este artículo se comprueba que ante un mismo valor para el parámetro θ , a medida que el tamaño de la caché aumentaba el número de peticiones registradas en el servidor principal sobre las páginas más populares disminuía. Por lo tanto, para dichas páginas se tiende a registrar un número aproximado de peticiones, originando que la mayor parte del tráfico existente en la red sea ocasionado por las peticiones hacia aquellas páginas web menos populares no disponibles en la caché, es lo que se ha dado a conocer como el efecto "*trickle-down*".

Por otro lado, en [21] las peticiones sobre las películas disponibles en un servidor de vídeo centralizado fueron modeladas con una distribución Zipf-like, con un valor para θ de 0,271, con el fin de analizar distintas políticas de distribución de contenido mediante el tratamiento conjunto de una serie de peticiones (batching) sobre

el mismo, para su posterior difusión a través de multicast. En una de las políticas (FCFS-n, First Come First Served) se estudió la reducción del tiempo de espera de los usuarios, de forma que se reservó parte de la capacidad del servidor para difundir las n películas con mayor probabilidad de ser demandadas, puesto que hacer lo mismo para las menos demandadas evidentemente supondría un desperdicio de recursos. En [22] se estudia una técnica de multicast denominada "*Skyscraper Broadcasting*" para aplicaciones de VoD, donde en términos generales se podría decir que realiza un broadcast de manera periódica de los vídeos más populares, donde la caracterización de la popularidad es realizada igual que en [21], y por lo tanto estaría suponiendo que el 80% de las peticiones se concentrarían sobre los 10 ó 20 objetos más populares. Así mismo, en [33] se estudia un algoritmo de caching de vídeos que tiene en cuenta tanto el tamaño de los mismos como su popularidad, para este último caso se supone que los objetos siguen una distribución Zipf-like cuyo valor de θ es nuevamente 0,271. Por el contrario, en [10] se modeló la demanda sobre un total de 250 películas, con la ley Zipf, donde se prueba a realizar el ajuste sobre dos días, el que menos y el que más valor de hit-rate tenían sobre las 10 películas más demandadas en el periodo de un mes, y en ninguno de los dos casos se consigue un buen ajuste. Por otro lado, los vídeos tienden a ocupar bastante espacio en la caché debido a su tamaño, así pues en [34] se presentan algoritmos que plantean dividir un vídeo en segmentos ("*chunk*"), con el fin de almacenar en la caché uno o varios trozos del vídeo y no en su totalidad. Además tanto el tamaño del segmento, así como el número de los mismos que serán seleccionados para su almacenamiento en la caché se plantea distinto, siendo mayor a medida que aumenta la popularidad del vídeo. Por lo tanto, en este artículo el tamaño de los segmentos se plantea como una métrica variable (Variable Chunk Size, VCS). En definitiva, las políticas de reemplazo de caché utilizadas, "Least Recently Used" (LRU) y "Least Frequently Used" (LFU), así como VCS basan sus mediciones en un modelado de la popularidad de los vídeos de tipo Zipf-like, donde el valor elegido para el parámetro θ fue 0,8. En [35] se analizan distintos escenarios con vídeo bajo demanda interactivo (IVoD) mediante una función de coste que tiene en cuenta tanto el ancho de banda como las necesidades de caché, se simuló el almacenamiento en caché de los 3, 5, 10 y 15 vídeos más populares, los cuales recibían del orden del 24%, 31%, 39% y 44% de las peticiones respectivamente en relación con el total, de acuerdo a la ley de Zipf.

En [36] se estudió los accesos en un servidor mMOD (the multicast Media On Demand) durante seis meses, sobre un total de 139 vídeos clasificados en dos categorías general y educacional. En los primeros, los accesos estaban más distribuidos a lo largo del tiempo, sin embargo en los segundos la demanda se concentraba en periodos más cortos, como por ejemplo la época de exámenes. Por lo tanto, en función de la categoría del vídeo el comportamiento de la demanda resultó ser diferente. De hecho, el patrón de acceso a las películas no seguía una distribución Zipf-like, dado que el 10% de las películas más populares acaparaban aproximadamente el 50% de los accesos, y por lo tanto representa un porcentaje bastante superior al que podría ser modelado con Zipf-like.

En entornos académicos existen estudios realizados por [37]. En [37] la longitud del periodo en estudio fue de una semana, durante el cual fueron accedidos un total de 23.738 objetos multimedia ubicados en 866 servidores distintos de Internet. Tras un estudio de la popularidad de los objetos se concluyó que el 78% fueron accedidos una sola vez, un 1% fueron accedidos 10 o más veces y que los 12 objetos más populares fueron accedidos más de 100 veces. La distribución de la popularidad se ajustó a una distribución Zipf-like con un valor para θ de 0,47. Los autores comparan su trabajo con [29], y concluyen que la concentración de los accesos sobre los objetos multimedia más populares es menor que en los objetos web. Además, los objetos multimedia presentaban una alta localidad temporal, lo cual dio pie a afirmar, que sería útil la distribución de los objetos a través de políticas de multicast, y usar caching sólo en periodos de alta demanda.

En [38] realizaron un modelado del patrón de acceso sobre los vídeos disponibles sobre dos servidores de la corporación Hewlett-Packard con un contenido diferente en cada uno de ellos a distintas escalas de tiempo. En el servidor designado como HP Corporation Media Solutions server se ofrecían vídeos sobre eventos, charlas, reuniones, información sobre productos hardware y software, así como material de tipo educacional. La longitud del periodo disponible fue de dos años y cinco meses, el cual fue estudiado con una escala semestral. La distribución Zipf-like fue aplicada, obteniendo un valor para θ de 1,6, siempre y cuando no se tuviesen en cuenta entre los 15 y los 20 primeros vídeos del ranking. Por otro lado, en el servidor HPLabs Media Server se proporcionaba información sobre los laboratorios HP, seminarios y meetings,

y donde la longitud del periodo disponible fue de un año y nueve meses. En este último caso, el ajuste se realizó con una escala mensual y donde el valor de θ osciló entre 1,4 y 1,6, y siempre y cuando no se tuviesen en cuenta los accesos entre los 10 y 15 primeros ficheros. El hecho del descarte de los primeros elementos del ranking también fue considerado en [29] como ha sido mencionado previamente. Más tarde, las mismas autoras Ludmila Cherkasova y Gupta Minaxi puntualizarían que la longitud del periodo en estudio, o lo que es lo mismo, la escala de tiempo, tiene una influencia notable a la hora de realizar la caracterización de la popularidad. Este hecho ha sido corroborado en el artículo [15], donde se analiza la popularidad con cuatro tipos de periodo (un mes, seis meses, un año y en su totalidad, que oscila alrededor de los dos años), y donde los resultados fueron los mismos que fueron comentados en [38]. Además en este estudio se señala que la no linealidad de los accesos en los objetos más populares es debida a la mínima diferencia que existe entre sus frecuencias.

En [39] se analizaron cuatro cargas, y sólo una de ellas se caracterizó con la distribución Zipf-like. Dicha carga representaba accesos a programas de televisión cuyo contenido se trataba de anuncios (duración: < 2 min.; n° ficheros: 82; n° peticiones: 61), trailers (duración: < 5 min.; n° ficheros: 15; n° peticiones: 33), así como noticias, entrevistas y eventos sociales (duración: < 15 min.; n° ficheros: 3; n° peticiones: 6). El valor del parámetro θ osciló entre 0,66 y 0,82 en el caso de anuncios, entre 1 y 1,2 para los trailers y entre 1 y 1,3 para la tercera categoría. Se comenta que podría ser debido a que cada vez que el usuario solicita un contenido al servidor, además se envía automáticamente un anuncio. Este hecho se plantea como una posible hipótesis, y que por lo tanto necesitaría ser investigado en más detalle.

En [40] se estudió un servicio de vídeo bajo demanda desarrollado por China Telecom en el año 2004, durante un periodo de siete meses, con alrededor de 21 millones de peticiones sobre más de 6.700 vídeos. Un detalle importante que se analiza es que factores externos, como una lista de vídeos recomendados o de los vídeos más populares, tenían una influencia notable sobre la popularidad de los vídeos. Además, en la mayoría de los días, el patrón de acceso siguió una distribución Zipf-like, con la excepción de una larga cola que representaba el acceso a aquellos vídeos que no eran populares. Por otro parte, cabe mencionar que los valores del parámetro θ se ajustaban a

una distribución normal según el test de Kolmogorov-Smirnov ($p_value = 0,940 > 0,05$), y su resumen estadístico se presenta en la Tabla 2.

N	Min.	Max.	Mean	Standard Deviation	Skewness	Standard Error	Kurtosis	Standard Error
209	0	0,348	0,199	0,070	-0,253	0,168	0,048	0,335

Tabla 2. Resumen estadístico de los valores del parámetro θ .

Actualmente, la construcción de arquitecturas de red, destinadas en su mayor parte al servicio de vídeo bajo demanda, implica cuestiones sobre como distribuir el contenido multimedia a lo largo de los servidores que conforman la red e incluso sobre el número de copias óptimo que deberían ser realizadas de cada vídeo. En el primer caso, en [41] se estudiaron una serie de algoritmos de reparto de contenido entre cachés como son RBC ("*Resource Based Caching*") que tiene en cuenta tanto el tamaño del fichero como el ancho de banda para su distribución, el algoritmo LFU, así como otros dos algoritmos que mejorarían cada uno de los dos anteriores. En dicho estudio se consideró que la popularidad seguía una distribución Zipf-like exacta ($\theta=1$), aunque si θ fuese igual a 0,271 como en [21], se comenta que los resultados obtenidos fueron los mismos. En [42] plantea un algoritmo de reparto de vídeos denominado "*Popularity and Partial Replication Load Sharing*" a través de los distintos servidores que constituyen la red con el objetivo de minimizar el tiempo de espera del usuario, de forma que los vídeos se distribuían de acuerdo a su popularidad. En este algoritmo los vídeos "más populares" serían copiados en todos los servidores, los cuales representarían un porcentaje sobre el conjunto de vídeos total, es lo que se ha denominado porcentaje de replicación, y el resto serían repartidos siguiendo algún otro tipo de algoritmo, como podría ser *Round-Robin*. En este estudio se pudo comprobar que si el valor de θ era bajo el porcentaje de réplica tenía un impacto alto sobre el tiempo de espera, por el contrario cuando el valor de θ era elevado apenas presentaba influencia sobre el mismo, los valores testeados para θ fueron 1 y 1,8 con unos porcentajes de réplica de 100%, 75% y 50% para ambos casos. En [2] se plantea un algoritmo de replicación basado en la popularidad de los vídeos, la cual fue caracterizada mediante una distribución Zipf-like, donde el valor del parámetro θ oscilaba entre 0,271 y 1. La idea era clasificar la popularidad de los vídeos en N intervalos, de forma que aquellos vídeos cuya popularidad estuviese comprendida

dentro del mismo intervalo tendrían el mismo número de copias. En [1] se estudia la influencia de la popularidad sobre distintas formas de dar servicio de VoD, con el fin de estimar un coste de instalación para una arquitectura de red descentralizada, se concluye que cuando sólo se permitía un servicio de VoD interactivo (iVoD) la popularidad no influía en el coste, por el contrario si se permitía dar servicio de tipo broadcast cuando el valor de θ aumentaba el coste total de la instalación disminuía, en concreto cuando θ aumentó de 0,7 a 1 el coste total disminuyó en un 10%. En [23] se estudió el reparto de tamaños de cachés a lo largo de una estructura de servidores en forma de árbol con el fin de medir la calidad del servicio (QoS, "*Quality of Service*") desde el punto de vista del tiempo de demora en la recepción del principio del vídeo. A cada vídeo se le asignaba un peso, el cual venía dado aplicando la ley de Zipf, justificando esta asignación a partir de los estudios realizados por [29, 39]. En [43] se analiza la fragmentación de los vídeos con el fin de disminuir el tiempo de espera del usuario. Inicialmente, se almacenaría la parte de comienzo del vídeo en la caché de un determinado tamaño, y posteriormente mediante un algoritmo de reemplazamiento, se sustituiría tanto la parte inicial como aquellos fragmentos del vídeo ya visualizados, por el fragmento o fragmentos que les suceden, en prevención de su posible demanda en un futuro cercano. Uno de los factores que tienen en cuenta es el impacto del patrón de acceso, el cual es modelado a través de una distribución Zipf-like, con θ comprendida entre 0,5 y 0,9, donde dichos valores estarían basados en los estudios realizados en [29]. La conclusión alcanzada en este análisis fue que a medida que el interés del usuario se centraba en cada vez menos objetos, y por lo tanto θ incrementaría su valor, el tiempo de espera disminuiría dado que un menor número de objetos tendrían más probabilidad de ser solicitados, y gozarían de más posibilidades para estar en la caché, con el consiguiente aumento del hit-ratio.

Últimamente también se está estudiando el comportamiento de la demanda sobre los vídeos presentes en redes sociales como YouTube y Yahoo entre otras. En [16] se estudió el servicio de Yahoo durante 46 días. Los accesos que se estudiaron fueron sobre los vídeos de las diez primeras páginas dentro de cada categoría, con un total de 1600 vídeos. La distribución de los accesos sobre una semana siguieron una distribución Zipf-like con un valor para θ de 0,85. La popularidad se analizó con cuatro escalas de tiempo (30 min, 1 hora, 1 día y una semana), y la distribución de la

popularidad resultó ser similar. En [17], se hizo un estudio de YouTube durante tres meses sobre los accesos a más de 600.000 vídeos desde un entorno académico y también externo. Dado que en YouTube tiene los vídeos clasificados en categorías, se comprobó que algunas de ellas son más populares que otras. Las referencias a los vídeos dentro del campus siguió una distribución Zipf-like con un valor de $\theta = 0,56$ ($R^2 = 0,97$). En [44] se estudiaron los accesos realizados desde un campus universitario a YouTube, durante un periodo de tres meses (85 días), y se comprobó que se ajustaban a una distribución Zipf-like con un valor para θ igual a 0,56, siendo el ajuste lineal R^2 de 0,97. En este artículo se justifica que θ tendría un valor bajo debido a dos motivos. En primer lugar, debido a la infraestructura de YouTube, dado que cuando un usuario solicita el mismo vídeo debe realizar una nueva petición y el vídeo volverá a ser descargado, y en segundo lugar se alude al hecho de que existirían una serie de vídeos muy populares entre los usuarios del campus. En [45] se realizó un estudio de YouTube sobre un periodo de tres meses con peticiones sobre más de tres millones de vídeos realizadas en el año 2007, tras el análisis de uno de los días de este periodo, en el cual se realizó el ajuste de la demanda con las distribuciones Zipf-like ($\theta=0,668$), Weibull ($k=0,513$; $\lambda=6.010$) y Gamma ($k=0,372$; $\theta=23.910$), se observó que tanto Weibull como Gamma caracterizarían mejor todo el perfil de la misma, dado que a partir de los 2.000 primeros vídeos más populares la demanda sufre un decrecimiento considerable, y por lo tanto la distribución Zipf-like no conseguiría un buen ajuste a partir de ese punto, justificando que los vídeos no populares serían accedidos en el momento que hubiesen sido subidos al sitio por sus creadores para comprobar su validez, y ya no serían accedidos nunca más.

En [46] se estudia la distribución de VoD en el sitio MSN durante un periodo de nueve meses, con la ayuda de la técnica de peer-to-peer de forma limitada (“*single video approach*”). El usuario que esté actualmente viendo un vídeo podrá distribuirlo a otro usuario que lo solicite, y no otro vídeo que pudiese haber visto con anterioridad, con el fin de paliar la carga sobre el servidor. Se observó que la distribución de la popularidad sobre los tres días con un tráfico mínimo, medio y máximo fue similar y seguía una distribución Zipf-like más aplanada en la cabeza, debido a una alta concentración de los accesos sobre los diez vídeos más populares, los cuales eran relativos a noticias y negocios.

La distribución Zipf-like también se aplica para modelar los accesos sobre contenido de tipo "*live streaming*", pero el enfoque que se plantea es diferente dada su naturaleza, pues se analizaría el número de veces que un cliente accedería a este tipo de contenido durante su transmisión. Así en [47] los clientes son ordenados de forma decreciente en función del número de sesiones que efectuarían sobre un determinado contenido durante su transmisión, y se modeló en escala logarítmica la relación entre el número de sesiones que un cliente realizó (eje Y) frente al ranking del cliente (eje X), según la ordenación comentada, obteniendo un valor para θ de 0,47. Por lo tanto, en VoD es el cliente el que elige un objeto en el momento en que el cliente lo desee ("*user driven*"), sin embargo en "*live streaming*" es el contenido el que determina el intervalo de tiempo durante el cual el cliente tendría la posibilidad de acceder al mismo ("*object driven*").

Desde el punto de vista de NoD, en [48-50] se aplica la distribución Zipf-like para modelar la demanda sobre el periódico digital noruego con más popularidad denominado "*Verdens Gang*". En algunos casos, el periodo considerado para el estudio fue de dos años y con 4,6 millones de accesos sobre un total de 3.500 ficheros de los cuales 1.000 eran de tipo audio. El valor del parámetro θ fue de 1,2, pero sólo se consigue ajustar los accesos de los vídeos más populares. Así mismo, en [51] se modela la popularidad sobre un periódico español de ámbito autonómico denominado "La Nueva España", el cual también ha sido objeto de análisis en esta tesis. En este estudio se resalta que es importante la localización de periodos de estabilidad, es decir, aquellos intervalos de tiempo en los cuales no es introducido nuevo contenido al servicio, y se plantea un algoritmo donde se tendrían en cuenta dichos periodos para realizar el cálculo del valor para el parámetro θ . Sin embargo, cuando un usuario entra en un servicio de NoD no suele ver una sola noticia sino un grupo de ellas. En [52] se plantea un algoritmo basado en lo que se ha denominado Multi-Selection Zipf distribution, en la cual se supone que si un usuario por término medio selecciona k noticias sobre un total de N esto da lugar a un total de M combinaciones, y la popularidad de los vídeos dentro de cada una de ellas seguirá una distribución Zipf-like. Este modelo representa la popularidad a corto plazo sobre un conjunto de noticias en periodos de tiempo equivalentes a una sesión de usuario. A través de este algoritmo se podría determinar cuales son los mejores vídeos candidatos para ser seleccionados en un prefetching.

2.2.2. Distribución Zipf-like generalizada

Si x representa el ranking de un fichero, e y representa la frecuencia de acceso sobre el mismo, y siendo k_x y k_y parámetros de escala, a través de una k -transformación se consigue cambiar los datos originales x e y según la expresión (2).

$$x_k = \frac{x + k_x - 1}{k_x} \quad e \quad y_k = \frac{y + k_y - 1}{k_y} \quad (2)$$

Desde el punto de vista del ranking, la escala es k_x , es decir, se estarían agrupando k_x ficheros en una clase. Por un lado, un fichero con ranking i -ésimo en la escala original pasaría a tener ranking $((i-1)*k_x+1)$ -ésimo con la k -transformación. Por otro lado, los ficheros ubicados en la cola presentan pequeñas diferencias en su número de accesos, y también se podrían agrupar k_y ficheros por clase. A continuación se realizaría un ajuste lineal a los datos x_k e y_k para obtener el valor del parámetro θ . Obviamente, la distribución Zipf-like es un caso particular de la distribución Zipf-like generalizada cuando $k_x = k_y = 1$.

En [53, 54] se estudiaron los accesos sobre un conjunto de ficheros presentes en dos servicios de streaming, durante periodos de 29 y 21 meses. En este estudio se analizaron ambas cargas, denotadas por HPC y HPL, de acuerdo a una serie de propiedades divididas en dos grupos las estáticas y las dinámicas. Como propiedades estáticas se consideró la duración del fichero, la velocidad de codificación, la popularidad y el acceso a la parte inicial del fichero, y como propiedades dinámicas se tomaron en cuenta la evolución de los accesos sobre el contenido multimedia a lo largo del tiempo, así como el orden y el tiempo de llegada de las peticiones. En este estudio la distribución de los accesos no seguía una distribución Zipf-like cuando se tenía en cuenta el periodo completo, tanto en su inicio como en su fin. En otros estudios como en [15, 29] también ocurría el mismo problema, y la forma de solucionarlo fue despreciando los accesos sobre los n primeros ficheros más populares. Sin embargo, sobre los ficheros más populares es donde recae el peso de los accesos, y por lo tanto no deberían de ser despreciados. Además, hay que tener presente, que ambos periodos engloban un elevado número de días, y por consiguiente existirán muchos ficheros con una popularidad similar, los cuales podrían ser considerados en un único grupo (clase de equivalencia). Por lo tanto, es en [53, 54] donde se aplicó la distribución Zipf-like

generalizada, con el objetivo final de que la popularidad se ajustase a una distribución Zipf-like, mediante la agrupación de los ficheros en clases, tanto desde el punto de vista del ranking como del número de accesos. De esta forma se consiguió un ajuste lineal en ambas cargas y para los periodos en su totalidad. Además, tal y como se puede observar en la Tabla 3 cada carga presenta el mismo valor para k_x y k_y .

Carga	θ	R^2	k_x	k_y	Frecuencia	Nº Ficheros
HPC	1,561	0,995	12	12	17.831	1.434
HPL	1,23	0,995	7	7	961	364

Tabla 3. Valores de los parámetros con k-transformación para las cargas HPC y HPL.

Dado el gran tamaño de los vídeos, existen técnicas de caching orientadas al tratamiento de partes de un vídeo [43, 55, 56], donde cada una de dichas partes es lo que se ha denominado "*segmento*". En [56] se propone un algoritmo de caching, basado en la popularidad interna de los distintos segmentos que componen un vídeo, entendiendo como popularidad de un segmento el número de veces que ha sido accedido. El estudio se realizó sobre los vídeos de dos servidores, seleccionando aquellos que recibieron más de 70 peticiones durante un periodo de 210 días, siendo bajo esta condición un total de 121 (2,1%) y 29 (5,6%) los vídeos que fueron estudiados en cada servidor, los cuales recibieron el 22,2% y el 31,3% del total de peticiones respectivamente. Se propone ajustar la popularidad mediante la técnica de Zipf-like k-transformada, de forma que se concluyó que para conseguir un coeficiente de determinación (R^2) superior a 0,9, el valor de k_x debería ser igual a 10 y el valor de k_y superior a 50, pero no más allá de un determinado umbral, pues empeoraba R^2 .

2.2.3. Combinación de dos Zipf-like (two Zipf-like)

En la mayoría de los estudios realizados hasta la fecha, la distribución de los accesos tiene dos partes bien diferenciadas, por un lado la parte correspondiente a los objetos más populares, y que por lo tanto reciben un mayor número de peticiones, que se podría denominar la cabeza (head), y otra parte con los objetos menos populares y que reciben un número de peticiones bastante inferior respecto a los objetos de la cabeza, que se podría denominar la cola (tail). Como consecuencia de este hecho, en

varios estudios se ajusta la distribución de la popularidad mediante la aplicación de dos distribuciones Zipf-like combinadas, una para la cabeza y otra para la cola, es lo que en la literatura se ha dado a conocer como "*two Zipf-like*". Así pues, si denotamos por 'y' el conjunto de peticiones sobre N objetos, 'l' el último índice del ranking que formaría parte de la cabeza, siendo entonces el rango de la cabeza de 1 a 'l', y donde por lo tanto el rango de la cola sería de (l+1) a N. Las distribuciones Zipf-like que modelarían la cabeza y la cola serían las expresadas en las ecuaciones (3) y (4).

$$y1 = \frac{C_1}{x_1^\theta} \quad 1 \leq x_1 \leq l, \quad \text{con} \quad C_1 = \frac{1}{\sum_{x_1=1}^l \frac{1}{x_1^\theta}} \quad (3)$$

$$y2 = \frac{C_2}{x_2^\theta} \quad (l+1) \leq x_2 \leq N, \quad \text{con} \quad C_2 = \frac{1}{\sum_{x_2=(l+1)}^N \frac{1}{x_2^\theta}} \quad (4)$$

El modelo final se formará a partir de la concatenación de las ecuaciones (3) y (4), dando como resultado la expresión (5).

$$f = p * y1 \oplus (1 - p) * y2, \quad \text{donde} \quad p = \left(\frac{\sum_{i=1}^l y_i}{\sum_{i=1}^N y_i} \right) \quad (5)$$

En [57] se hace un estudio sobre los accesos a 5.000 URL's, con contenido de tipo live streaming, durante un periodo de tres meses, en el cual se produjeron más de 70 millones de peticiones. Tras el ajuste, la cabeza representaba las peticiones sobre las 1.000 URL's más populares, oscilando entre las 10.000 peticiones y los 7,3 millones, y donde resultó un valor para θ de 1,01. Por otro lado, la cola representaba los accesos sobre las URL's restantes que acapararon menos de 10.000 peticiones, y con un valor para θ bastante superior a 1. Cabe destacar que solamente el 7% de los objetos eran vídeos que captaron un 1% de accesos, siendo los más demandados objetos de tipo audio, y en concreto de tipo radio ("*non-stop audio programs*").

En [20] se analizaron los accesos realizados en un entorno universitario sobre dos servidores con contenido educacional, e-Teach y BIBS. El primero de ellos contenía las lecciones correspondientes a un curso que no tenía clases presenciales y en el cual estaban matriculados 280 alumnos. El segundo contenía las lecciones de once cursos de

distintas disciplinas, pero en los cuales, los alumnos disponían además de clases presenciales. En este artículo se hace un estudio de los accesos a los vídeos clasificados en cuatro grupos en función de su duración en minutos {0-5, 5-10, 10-15, 50-55} en el cual se observó que el nº de ficheros, la probabilidad total y el valor del parámetro θ de cada parte dependía del tamaño del fichero, tal y como aparece reflejado en la Tabla 4.

Tamaño Fichero (min.)	Servidor	Primera Zipf-like			Segunda Zipf-like		
		Prob.	Nº Fich.	θ	Prob.	Nº Fich.	θ
0-5	eTeach	[0,6-0,85]	{2; 7}	[0,2-2,5]	[0,2-0,45]	[3-17]	[0,8-2]
	BIBS	[0,1-0,2]	[2-4]	[0-0,33]	[0,8-0,9]	[30-85]	[0,35-0,55]
5-10	eTeach	{0,8;0,85}	2	{0,3625;0,5208}	{0,2;0,15}	2	{1,6; 10,047}
10-15	eTeach	{0,89;0,97}	2	{3,6;5,4}	{0,03;0,1}	2	{0;2,4}
50-55	BIBS	{0,5;0,6}	[6-18]	[0,2-0,45]	[0,4-0,5]	[58-130]	[0,45-0,65]

Tabla 4. Valores típicos de los parámetros para BIBS y eTeach.

Basándose en el estudio de [20] en [58] se simularon cuatro escenarios de accesos para cuatro valores de θ diferentes, pero para modelar los objetos más populares, es decir, sólo se estaría considerando la primera distribución Zipf-like. La finalidad de este artículo fue presentar un estudio sobre el mejor reparto para transmitir conjuntamente varios contenidos, y conseguir aprovechar el ancho de banda al máximo posible cuando se transmite usando multicast. La conclusión fue que a medida que aumentaban el número de objetos el incremento de ancho de banda que se necesitaba era cada vez menor.

En [39] fueron estudiadas cuatro cargas, una con contenido educacional, otra con vídeos de corta duración, y las dos restantes de tipo audio. La carga con contenido educacional, así como las de tipo audio fueron modeladas con la distribución two Zipf-like, mientras que la otra carga fue modelada con una distribución Zipf-like. El número de días estudiados en cada una de ellas fue de 411, 24 y 7, sobre un total de ficheros de

230, 70.479 y 42.746 respectivamente. Los resultados obtenidos son presentados en la Tabla 5, donde tanto para la distribución Zipf-like que ajusta la cabeza como para aquella que ajusta la cola, se muestra la probabilidad, el porcentaje de ficheros y el valor del parámetro θ .

Tipo Carga	Tamaño Fichero (min.)	Primera Zipf-like			Segunda Zipf-like		
		Prob.	%Fich.	θ	Prob.	%Fich.	θ
Educativo	0-5	0,44-0,81	8-33	0,19-1,6	0,19-0,56	67-92	0,86-2,6
	5-20	0,56-0,88	9-13	0,93-0,97	0,12-0,44	87-91	1,5-2,2
		0,72-0,95	10-25	2,6-3,8	0,05-0,28	75-90	0,96-2,1
	30-40	0,89-0,94	14-25	2,5-4,4	0,06-0,11	75-86	0,85-2,6
Radio/UOL	0-3	0,36-0,58	3-5	0,54-0,77	0,42-0,64	95-97	1,1-1,2
	3-5	0,50-0,72	7-11	0,6-0,85	0,28-0,5	89-93	1,3-1,4
	5-10	0,48-0,69	7-13	0,58-0,79	0,31-0,52	87-93	1,3-1,4
Audio	0-3	0,95-0,96	34-41	0,83-0,85	0,04-0,05	59-66	1,8-2
	3-5	0,65-0,72	13-18	0,63-0,7	0,28-0,35	82-87	1,7-2
	5-10	0,82-0,88	30-38	0,81-0,86	0,12-0,18	62-70	1,9-2,5

Tabla 5. Valores típicos de los parámetros.

También existen estudios en sistemas de vídeo compartido como son YouTube o Yahoo. En [17] se hace un estudio sobre los accesos a los vídeos de YouTube durante un periodo de 46 días, el cual fue estudiado en periodos de 30 minutos, una hora, un día y una semana. En cada uno de los periodos se realizó un ajuste con las distribuciones exponencial, Zipf-like y two Zipf-like, y de todas ellas la que mejor ajustaba era two Zipf-like. Un detalle a destacar es que la forma de la distribución two Zipf-like prácticamente coincidía en todos los tipos de periodos. En este artículo no se proporciona información sobre los valores de los parámetros obtenidos en cada una de las distribuciones, tan solo se proporciona la bondad de ajuste de los accesos para una semana ($R^2=0,8545$). Sin embargo, el estudio de la popularidad no se realiza por categoría sino globalmente.

2.2.4. Distribución Mandelbrot

La expresión correspondiente a la distribución Mandelbrot se encuentra en la ecuación (6). Además, cabe destacar que cuando k vale 0, la distribución Mandelbrot se convierte en la distribución Zipf-like.

$$y = \frac{C}{(x+k)^\theta} \quad 1 \leq x \leq N \quad y \quad k \geq 0, \quad \text{siendo} \quad C = \frac{1}{\sum_{x=1}^N \frac{1}{(x+k)^\theta}} \quad (6)$$

C es una constante de normalización, N el número de objetos y k una constante. En [59] fue aplicada para modelar la demanda de los vídeos más populares distribuidos mediante P2P, y analizar algunos algoritmos de reemplazamiento de contenido en cachés de distinto tamaño. Una de las conclusiones a las que se llegó, es que los algoritmos LFU y LRU no mejoraban el "hit rate" a través de este modelado, y se ha presentado un algoritmo alternativo propio. Los valores de los parámetros obtenidos a partir del estudio de 16 cargas con distinta demanda fueron que el valor de θ estaba comprendido en [0,4-0,7] y el valor típico de k en [5-60]. Además, en aquellas cargas con mayor demanda el valor de k resultó ser más bajo.

2.2.5. Distribución Stretched

La distribución exponencial Stretched tiene un amplio campo de aplicación, como es el modelado de fenómenos naturales (emisiones de radio y luz de las galaxias, tamaño de las reservas de petróleo, variaciones de la temperatura, etc), en economía (tamaño de las poblaciones de países, tamaño de las aglomeraciones, etc), en Internet (patrones de acceso a contenido de tipo media, contribución de contenido en redes sociales por parte de los usuarios, ...), etc.

Su definición viene dada por la expresión (7).

$$P(x)dx = c \frac{x^{c-1}}{x_0^c} e^{-\left(\frac{x}{x_0}\right)^c} dx \quad (7)$$

Y su función de distribución acumulada es la función Weibull expresada en (8).

$$P_c(x) = e^{-\left(\frac{x}{x_0}\right)^c} \quad (8)$$

Los parámetros c y x_0 son constantes. Si los valores de la variable dependiente Y se ordenan de forma decreciente, y a cada valor se le asigna un ranking x_i ($1 \leq i \leq n$) atendiendo a la posición que ocupa dentro de la ordenación, entonces se cumplirá que $P(X \geq x_i) = i/n$. Atendiendo a esta ordenación, la distribución exponencial se puede transformar en una línea recta si la variable dependiente se eleva a una cierta constante c y la variable independiente es el logaritmo del ranking, dando lugar a la función expresada en (9).

$$y_i^c = -a * \log(i) + b \quad \text{con} \quad 1 \leq i \leq n \quad (9)$$

Donde $a = x_0^c$, $b = y_1^c$. El valor del parámetro c define la forma de la distribución, que suele tener forma lineal seguida de una curvatura (cola), de hecho cuantifica el tamaño de la cola (si c disminuye el tamaño de la cola aumenta), $x_0 = a^{1/c}$ determina la escala, y $b^{1/c}$ es el valor que tiene la variable dependiente de ranking 1. En definitiva, hay un cambio de escala de los datos originales, de forma que el eje de abscisas sería $\log(x)$, y el eje de ordenadas sería y^c , con el objetivo de conseguir una linealización.

En [60, 61] se aplica la distribución Stretched sobre dieciseis conjuntos de datos recogidos en diferentes entornos de distribución de contenido como son Web, VoD, P2P y live streaming entre los años 1998 y 2006, y donde los periodos de análisis eran variados, entre los cuales se podrían mencionar periodos de 5 días, 11 días, 16 semanas, 29 meses, etc. El valor del parámetro 'a' aumentaba a medida que la longitud del periodo en estudio era mayor, y convergía a una constante. Se comprobó que aquellos vídeos que se lanzaban automáticamente cuando un usuario entraba a la página web, por supuesto no eran peticiones reales y provocaban una distorsión en el patrón de acceso. Si el comportamiento de la demanda se modelaba teniendo en cuenta estos accesos junto con los reales, en uno de los casos se concluyó que el patrón se ajustaba a una distribución Zipf-like con θ igual a 0,71; sin embargo si los accesos no reales eran

eliminados, los cuales equivalían al 31% de las peticiones, la demanda se ajustaba mejor a una distribución Stretched.

La aplicación de esta distribución también ha llegado a las redes sociales como en [62], pero en este artículo el enfoque es al revés, es decir, no se modela la demanda sobre el contenido sino la cantidad de contenido genuino incorporado por los usuarios, y por lo tanto los usuarios son ordenados de forma decreciente en función del número de contribuciones que han aportado ("*rank order distribution function*"), con el fin de identificar a aquellos que suben la mayor cantidad de contenido.

2.3. Concentración de la demanda

El estudio de la concentración de la demanda o la "*localidad temporal*", como es conocido en la literatura, se podría definir como "un objeto que ha sido referenciado tiene una alta probabilidad de ser referenciado en un futuro cercano" según [63]. Esta característica sería un dato más que ayudaría a tomar una decisión sobre que tipo de algoritmo de reemplazamiento de contenido podría ser más apropiado para ser aplicado en las cachés. Así por ejemplo, si el porcentaje de accesos tuviese una alta concentración sobre los vídeos más populares, es decir, se cumpliera el Principio de Pareto, el algoritmo LFU ("*Least Frequently Used*") podría ser una buena alternativa, incluso mejor que LRU ("*Least Recently Used*") según [64].

En [65] se estudia la localidad temporal cuando las peticiones son servidas a través de una serie de cachés unidas entre sí de forma jerárquica. Aquí se plantea una métrica para su medición como es la "*entropía*", donde su definición viene dada a través de la ecuación (10).

$$H(x) = - \sum_{i=1}^n p_i \log_2 p_i \quad (10)$$

La variable 'n' representa el número de peticiones y p_i la probabilidad de ocurrencia de la petición que ocurrió en i-ésimo lugar. El máximo valor de $H(X)$ se alcanza cuando todas las peticiones tienen la misma probabilidad de ocurrencia ($p_i=1/n$, para todo $i \in [1, n]$), y el mínimo valor cuando sólo un objeto recibe todas las

peticiones, y por lo tanto $H(X)=0$. Esta métrica también ha sido utilizada en [63] para analizar la distribución de las referencias a objetos Web.

Según estudios realizados en [29] entre el 25% y el 40% de las páginas web acapararon el 70% de los accesos. En entornos de tipo empresarial la localidad temporal podría ser diferente. De hecho, según estudios realizados por [38] en la empresa HP, entre el 14% y el 30% de los vídeos disponibles capturaron el 90% de los accesos.

En [64] se analizó la concentración de los accesos sobre páginas web en seis conjuntos de datos, donde no todos provenían del mismo ámbito, de hecho tres de ellos pertenecían a un ámbito académico, dos a un entorno científico y el último se trataba de un proveedor comercial de Internet. En los seis conjuntos de datos se comprobó que el 10% de las páginas más solicitadas acaparaban entre el 80% y el 95% de las peticiones.

En [1] se observó la concentración de los accesos sobre los ficheros más populares para distintos valores del parámetro θ de la distribución Zipf-like, así por ejemplo cuando θ fue igual a 0,7 el 50% de los accesos se concentró sobre el 33% de los ficheros más populares.

En [40] se estudia un sistema de VoD desarrollado por China Telecom denominado PowerInfo, donde los accesos sobre el 10% de los vídeos más populares acapararon el 60% de los accesos, mientras que el 23% de los objetos capturaron el 80% de los accesos. Por lo tanto, presenta unos resultados más moderados que el Principio de Pareto en sí.

En redes sociales como YouTube, y de acuerdo a los estudios realizados en [44] entre el 10% y el 20% de los vídeos más populares acaparaban el 39,7% y el 52,4% de las peticiones respectivamente, y por lo tanto no seguía el Principio de Pareto. Además el 68,1% de los vídeos fueron accedidos una sola vez, y capturaron el 35,3% de las peticiones. Justifican este hecho a la gran variedad y cantidad de contenido disponible que hace que la demanda esté más diversificada, y por lo tanto se traduce en un menor número de peticiones por vídeo. Por otro lado, el 10% de los vídeos vistos un día, fueron vistos al día siguiente, salvo en periodos de bajo tráfico o durante los fines de semana, con una reducción de un 5%. Sin embargo, según los estudios de [17] la popularidad seguía el Principio de Pareto (regla 80-20), pues el 20% de los vídeos más populares acapararon el 80,1% de los accesos. En el momento de estudio del artículo

(17 Julio al 31 de Agosto de 2007) los vídeos de YouTube estaban clasificados en 16 categorías, y los accesos en cada una de ellas oscilaron entre el 14,76% y el 2,2% para la más y la menos demandada respectivamente. Según el estudio realizado en [16] para la red social Yahoo, donde se hizo un análisis de 46 días en cuatro escalas de tiempo (30 min., 1 hora, 1 día y una semana) de las 16 categorías, que de entonces presentaba este servicio para clasificar sus vídeos, se observó que la concentración de la popularidad seguía el Principio de Pareto, al igual que en el estudio realizado en [17] para YouTube.

2.4. Redundancia global y parcial

El estudio de la redundancia aparece en [66, 67] con el objetivo de encontrar patrones de regularidad en el comportamiento de los usuarios en su búsqueda de información a través de la red. El análisis se centra en las palabras más usadas por los usuarios, así como las URL's más visitadas. Ambos estudios se apoyan en los datos recogidos de los proxies-cachés de France Telecom R&D durante un periodo de 17 meses correspondientes a 331 usuarios, con un total de 1.510.358 peticiones realizadas sobre 392.853 objetos. Para tal fin, se apoya en dos métodos, la redundancia global (gR) y parcial (pR). La redundancia global la define como la proporción del número de items diferentes que son referenciados a lo largo de una serie de peticiones entre el número total de items presentes en dicha serie, tal y como se muestra en la ecuación (11).

$$gR = 100 - 100 * \left(\frac{\# \text{ items } \text{ úni cos}}{\# \text{ total } \text{ items}} \right) \quad (11)$$

Por otra parte, si la serie de peticiones S se divide en secuencias de peticiones Q_x , no solapadas y del mismo tamaño T (% items), se calcularía el grado de redundancia que existe entre un par de secuencias entre sí, Q_i y Q_j siendo $i < j$. Así pues, considerando parejas de secuencias de peticiones con el mismo tamaño T no solapadas entre sí, la redundancia parcial la define como la proporción del número de items diferentes de una secuencia de peticiones Q_i que se repiten en otra secuencia de peticiones posterior Q_j , tal y como se refleja en la ecuación (12).

$$pR(Q_i, Q_j) = 100 * \left(\frac{\# \text{ items } Q_i \in Q_j}{\# \text{ items } Q_i} \right) \quad (12)$$

La redundancia parcial fue calculada con secuencias de un mismo tamaño T (% items) y distinto número de secuencias de separación (Δ) entre las mismas, tal y como se muestra en la Fig. 1.

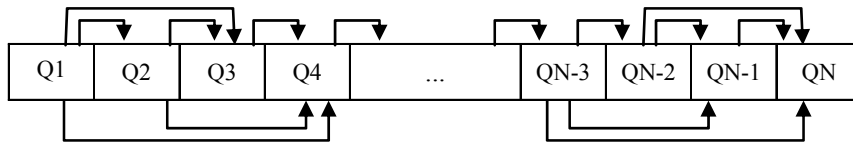


Figura 1. Pares de secuencias involucradas en el cálculo de pR.

En primer lugar, el proceso consistía en establecer el tamaño de las secuencias a un valor T (% items). En segundo lugar, la serie total de items S se dividía en N secuencias (S/T). En tercer lugar, pR se calculaba entre pares de secuencias, de forma que pR se calcularía entre la primera secuencia denotada por Q_0 y la segunda secuencia Q_1 ($pR(Q_0, Q_1), \Delta=0$), después Q_0 con Q_2 ($pR(Q_0, Q_2), \Delta=1$), y así sucesivamente. Este proceso sería reiterado hasta que todos los pares de secuencias fueron cubiertos (Q_i se compararía con $Q_{(i+N/2)}$). El algoritmo final aplicado en este artículo para el cálculo de pR se encuentra representado en la Tabla 6.

```

Para cada usuario

  Para cada secuencia  $T$  desde 1% hasta 20% sobre el total de items de  $S$ 

     $N = \text{n}^\circ \text{ items de } S / \text{n}^\circ \text{ items } T$ 

    Para cada secuencia  $i$  desde 0 hasta  $N/2$ 

      Para cada secuencia  $j$  desde  $(i+1)$  hasta  $(i+N/2)$ 

        Calcular  $pR(Q_i, Q_j)$ 

      Fin Para  $j$ 

    Fin Para  $i$ 

  Fin Para  $T$ 

Fin Para usuario
  
```

Tabla 6. Pseudocódigo para el análisis de la redundancia parcial entre pares de secuencias.

Se comprobó que cuando se comparaban parejas de secuencias cercanas en el tiempo, el valor de pR se incrementaba, lo cual es lógico dado que es más probable que el número de items en común sea mayor. Por lo tanto, el valor máximo para pR se consigue cuando la distancia entre las dos secuencias que se comparan es mínima ($\Delta=0$). A partir de los valores de los parámetros T y Δ se concluyó la relación expresada en (13), donde pR_{max} representa el valor máximo de pR y k una constante.

$$\log(pR) = -kT \log(\Delta) + \log(pR_{max}) \quad (13)$$

En consecuencia, $pR=f(T, \Delta)$ resultó ser independiente tanto del usuario como del tiempo, dado que el cálculo de pR está ligado al tamaño de la secuencia de peticiones T y a la distancia entre secuencias Δ . El modelo asegura que a medida que Δ aumenta, pR disminuye.

Capítulo 3.

ANÁLISIS DEL PERIÓDICO DIGITAL

ASTURIES.COM

3.1. Introducción

El estudio de la popularidad de los vídeos se inició en uno de los primeros diarios digitales de noticias en España, Asturias.com, cuyos resultados dieron pie a una publicación en un congreso de ámbito nacional denominado Jitel [68]. De entonces, se trataba del tercer diario en Internet más visitado de Asturias, tras La Nueva España y El Comercio Digital, y del más popular en términos absolutos entre los servicios de este tipo que ofrecían información en lengua asturiana. El servicio de vídeo bajo demanda Asturias.com se incorporó al diario en el año 2004, ganando una notable popularidad. En él se ofrecía información de actualidad sobre Asturias, así como sobre la cultura asturiana (música, tradiciones, lengua, cultura, ecología, etc.). Además, también disponía de múltiples canales de radio temáticos que permitía a los músicos asturianos promocionar sus trabajos. El acceso al servicio se realizaba a través de un enlace situado en la parte izquierda de la página inicial del periódico, tal y como se muestra en la Fig. 2, y si se hiciese un desplazamiento a lo largo de la página también se encontraría un marco incrustado con información sobre el último vídeo publicado.



Figura 2. Parte superior de la página inicial del periódico digital Asturias.com.

La página del servicio presentaba dos zonas diferenciadas, como se observa en la Fig. 3. La parte superior mostraba información sobre el último vídeo publicado, además de permitir su reproducción mediante un plug-in de Flash o RealNetworks (en función

de su formato). Este vídeo o “vídeo en portada”, se reproducía de forma automática cuando se entraba en el servicio. Por otro lado, la parte inferior permitía acceder a vídeos que habían sido publicados con anterioridad.



Figura 3. Página del servicio de VoD del periódico Asturias.com.

La frecuencia de incorporación de nuevos vídeos aproximadamente era de uno por semana, característica que lo hacía singular, pues el hecho de introducir contenidos de forma tan espaciada en el tiempo hacía que vídeos con una gran demanda siguiesen siendo solicitados pese a la introducción de nuevos contenidos, de hecho a lo largo de algo más de los dos años sólo fueron incorporados un total de 42 vídeos. Esta situación no se daría en otros periódicos digitales, en los que la información es introducida diariamente, haciendo que vídeos anteriores pasen a ser desapercibidos con rapidez.

3.2. Análisis de la popularidad

El estudio de la popularidad ha sido realizado sobre un intervalo de tiempo con una duración de dos años y ocho meses, durante el cual han sido introducidos un total de 42 vídeos. Además es importante destacar, que todo vídeo nuevo que sea introducido siempre se ubicará en portada. De este modo, este vídeo siempre tendrá mayor protagonismo en el servicio frente a los posibles clientes, y reemplazará al vídeo anterior, que pasará a convertirse en un vídeo más, junto con todos los vídeos ya ofertados en Asturias.com. Dada la extensa amplitud del mismo, el impacto que produce

la llegada de un vídeo nuevo sobre la popularidad del servicio (influencia de la popularidad de un vídeo nuevo frente a la popularidad del vídeo anterior), se ha decidido focalizar el estudio sobre cuatro periodos en base a los criterios establecidos en la Tabla 7.

CRITERIOS	
Periodo 1	Baja popularidad en el vídeo anterior y en el nuevo.
Periodo 2	Baja popularidad en el vídeo anterior y alta en el nuevo.
Periodo 3	Alta popularidad en el vídeo anterior y baja en el nuevo.
Periodo 4	Alta popularidad en el vídeo anterior y alta en el nuevo.

Tabla 7. Casos de estudio ante la incorporación de un vídeo nuevo.

En cada uno de los cuatro criterios se realizará un análisis tanto del ranking, principalmente de los dos primeros vídeos de cada día, así como del valor de θ que modelará la popularidad del servicio diaria según la distribución Zipf-like. Ambos análisis irán acompañados de las gráficas correspondientes.

Dada la variabilidad de la longitud del periodo entre la introducción de dos vídeos consecutivos, con el fin de unificar el estudio para cada criterio se ha decidido estudiar sólo los siete primeros días tras la introducción de un nuevo vídeo. Además como se puede observar en la Tabla 7, para cada uno de los periodos elegidos se tiene en cuenta la popularidad del vídeo publicado en el periodo anterior, pues se necesita analizar el efecto que producirá la introducción del nuevo vídeo sobre él. Así pues, este dato también es relevante. Para tal fin, se ha decidido incorporar el valor de θ del último día del periodo anterior, así como su situación en el ranking para ese mismo día, el cual vendrá representado en la gráfica a través del valor 0 en el eje de abscisas. En los apartados siguientes se estudiará en detalle cada uno de los cuatro periodos.

3.2.1. Baja popularidad en el vídeo anterior y en el nuevo

El vídeo nuevo introducido en el servicio es el 1362, siendo el anterior el 1320. Como se observa en la Fig. 4, el vídeo 1362 a partir de su segundo día siempre ocupa la primera posición en el ranking, sin embargo el primer día el vídeo más demandado sigue siendo el vídeo introducido en el periodo anterior, es decir, el 1320. La causa

principal de la baja demanda en su primer día seguramente sea debido a la hora de publicación en portada, pues lógicamente si ha sido a última hora de la tarde la demanda será menor que si es publicado por la mañana.

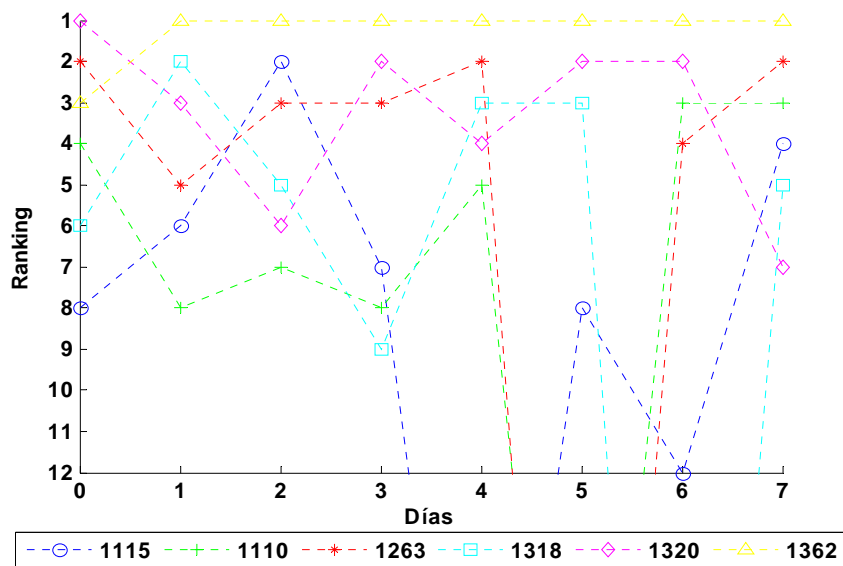


Figura 4. Ranking de accesos en el Periodo 1.

Por otro lado, el porcentaje de accesos al vídeo nuevo 1362 oscila desde el 15% en su primer día de publicación, presentando una variación para el resto de los días entorno al 50%, excepto en el sexto día donde alcanzaría su máximo apogeo llegando a alcanzar el 67% de accesos. Por el contrario, el 1320 presenta un 45% de accesos en su primer día, para ir decreciendo paulatinamente hasta finalmente alcanzar un valor inferior al 1% de accesos. Además como se puede observar en la Fig. 4 no mantiene su permanencia en segunda posición, pues dado que es un vídeo con baja demanda siempre estará compitiendo con el resto de los vídeos presentes en el servicio.

Dado que aparece un nuevo vídeo en el servicio, este hecho se ve reflejado con un incremento en el valor de θ en el primer día de su publicación con respecto al día anterior, que como se puede observar en la Fig. 5 pasa de 1,21 a 1,37. Aunque este incremento no es elevado debido a su bajo porcentaje de accesos. Este valor de θ va incrementando a medida que aumenta el porcentaje de accesos diaria sobre el vídeo nuevo, presentando un pequeño descenso en los días 4 y 5. Sin embargo, como ya se ha comentado en el sexto día, el vídeo nuevo alcanzaba el 67% accesos, hecho que se ve corroborado con un valor elevado para θ de 2,05, para luego descender al 50% dando lugar a un descenso de θ a un valor de 1,52.

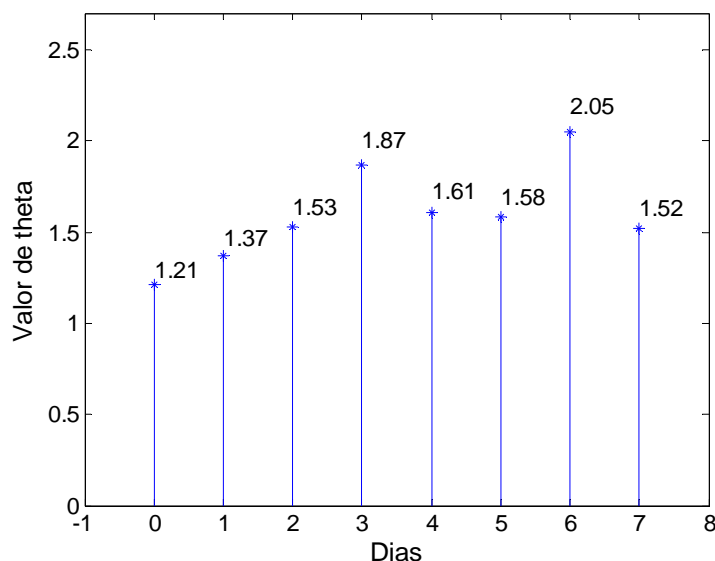


Figura 5. Valor de theta en el Periodo 1.

3.2.2. Baja popularidad en el vídeo anterior y alta en el nuevo

El vídeo nuevo introducido en el servicio es el 1370, siendo el anterior el 1362. Como se observa en la Fig. 6, el vídeo 1370 desde su primer día siempre ocupa la primera posición en el ranking, pues se trata de un vídeo que siempre va a tener una alta popularidad. Sin embargo, el vídeo 1362 permanece en segunda y tercera posición hasta el cuarto día, y a partir del quinto día su ranking decae paulatinamente hasta llegar finalmente a la novena posición.

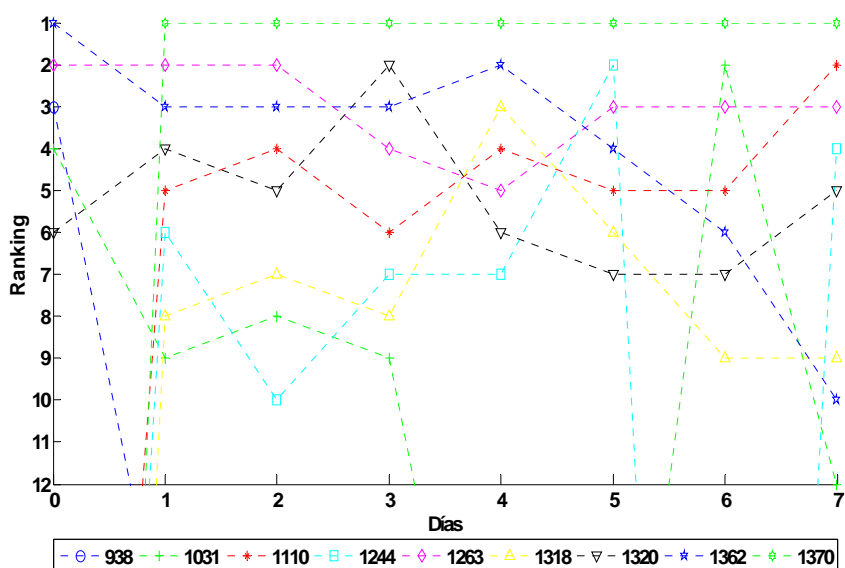


Figura 6. Ranking de accesos en el Periodo 2.

El porcentaje de accesos al vídeo nuevo 1370 es del 62% en su primer día, y va decayendo ligeramente hasta alcanzar el 48% en su sexto día. En el séptimo día vuelven a subir los accesos al 58%. Es un vídeo que presenta un porcentaje de accesos elevado, prácticamente siempre superior al 50%, con respecto a los vídeos que ocupan la segunda y tercera posición del ranking cuyo porcentaje no supera el 9% de accesos durante todo el periodo.

En la Fig. 7 se puede apreciar que el primer día que se introduce el vídeo el valor de θ es de 2,01, bastante elevado con respecto al último día del periodo anterior, antes de ser introducido en el servicio, cuyo valor es de 1,38, corroborando el hecho de que el vídeo introducido presenta un alto porcentaje de accesos con respecto a los demás, ya desde el primer momento. Para los sucesivos días el valor de θ decae a medida que disminuye el porcentaje de accesos, pero suavemente y siempre manteniéndose en valores elevados comprendidos entre 1,88 y 1,53. Finalmente, dado que el séptimo día el porcentaje de accesos se vuelve a incrementar en un 10%, esto se refleja con un incremento en el valor de θ tomando el valor de 1,87.

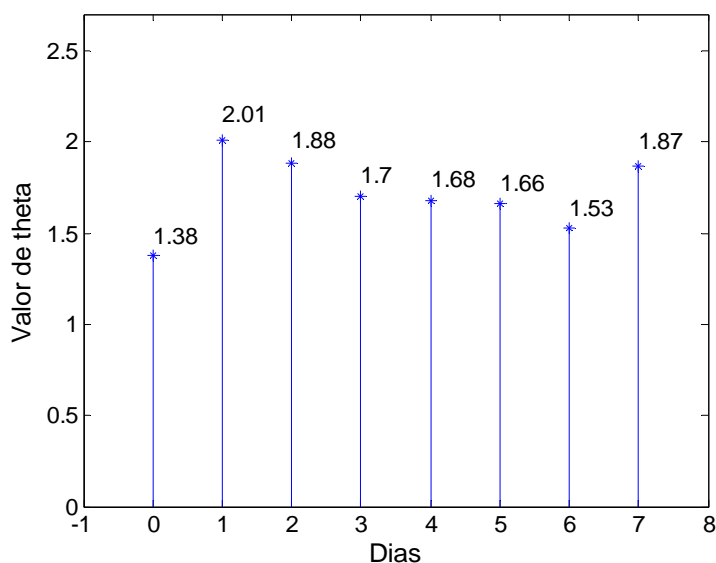


Figura 7. Valor de theta en el Periodo 2.

3.2.3. Alta popularidad en el vídeo anterior y baja en el nuevo

El vídeo nuevo introducido en el servicio es el 1311, siendo el anterior el 1263. Como se observa en la Fig. 8, el vídeo 1311 está continuamente compitiendo por la primera posición con el vídeo anterior 1263.

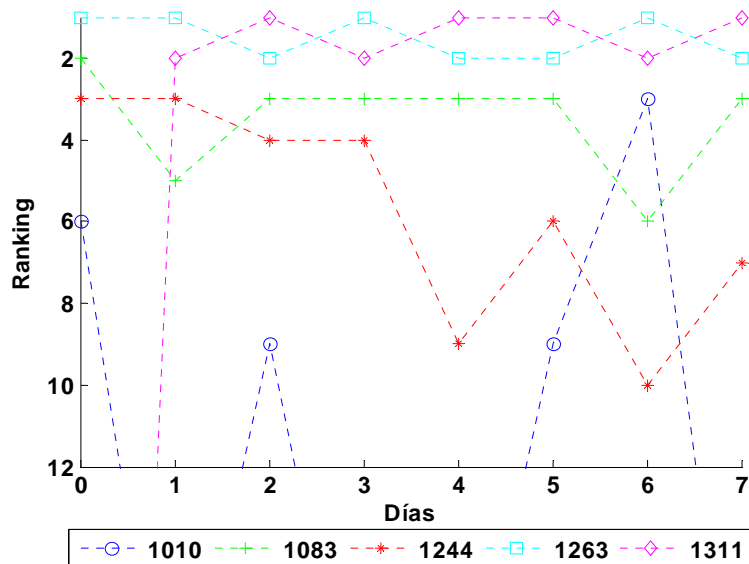


Figura 8. Ranking de accesos en el Periodo 3.

Dado que el vídeo nuevo tiene una baja demanda, el porcentaje de accesos sobre el vídeo anterior siempre debería ser mayor, pues es el que sigue siendo solicitado con mayor asiduidad. Pero dado que el nuevo vídeo siempre es reproducido automáticamente cuando un cliente entra al servicio, el porcentaje de accesos sobre el vídeo anterior deberá ser igual o inferior al vídeo nuevo. Pero, según se observa en la Fig. 8 el primer día el vídeo anterior se sitúa el primero en el ranking con un porcentaje de accesos muy elevado del 61% respecto al nuevo vídeo que sólo alcanza el 8%, esto puede ser debido a que el vídeo anterior ha permanecido en portada la mayor parte del día. Una vez que el vídeo de portada ha sido reemplazado por el nuevo, el resto de los días ambos vídeos presentan una situación muy pareja respecto a su porcentaje de accesos, con una diferencia aproximada del 6% entre ambos, estando siempre el vídeo nuevo en primera posición en el ranking, excepto el tercer y el sexto día. Esta situación puede ser debida al camino elegido para acceder al vídeo anterior, pues no sólo es posible acceder al mismo a través de la página de inicio del servicio Asturias.com, sino a través de enlaces externos presentes en otros servicios, evitando de esta forma la

reproducción automática del nuevo vídeo. Por lo tanto, si el vídeo anterior tiene mucha demanda seguirá compitiendo bastante fuerte por su primera posición en el ranking, consiguiéndolo tal y como ocurre en los días tercero y sexto, con un 1% más de accesos sobre el nuevo.

El último día antes de ser incorporado el nuevo vídeo θ presenta el valor de 2,17 como se puede observar en la Fig. 9, lo cual implica que el vídeo anterior presenta una alta popularidad, pues acapara el 66% de los accesos. Tras ser incorporado el nuevo vídeo su valor decae muy levemente pasando a ser de 1,96, pero sigue presentando un porcentaje del 61% de los accesos, debido a que aún no ha sido retirado de portada. Una vez que el nuevo vídeo realmente aparece en el servicio, a partir del segundo día la popularidad del anterior vídeo es compartida con el nuevo vídeo incorporado, por lo que el valor de θ desciende a 1,16, manteniéndose con un valor bajo, salvo que el vídeo anterior acapare más porcentaje de accesos que el primero debido a que sea demandado a través de otros enlaces. Esta situación aparece reflejada también en el valor de la θ sobre todo en el sexto día con un ascenso en su valor a 1,34.

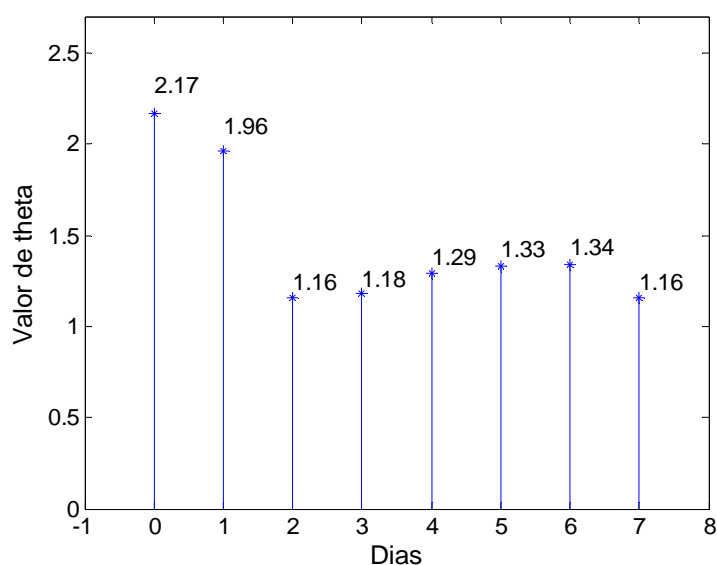


Figura 9. Valor de theta en el Periodo 3.

3.2.4. Alta popularidad en el vídeo anterior y alta en el nuevo

El vídeo nuevo introducido en el servicio es el 1381, siendo el anterior el 1370. Una vez que entra en el sistema siempre permanece en primera posición, y el vídeo anterior en segunda posición, tal y como se observa en la Fig. 10. El porcentaje de accesos al vídeo nuevo 1381 oscila entre el 50% y el 39%, a diferencia del vídeo anterior 1370 que oscila entre el 17% y el 30%.

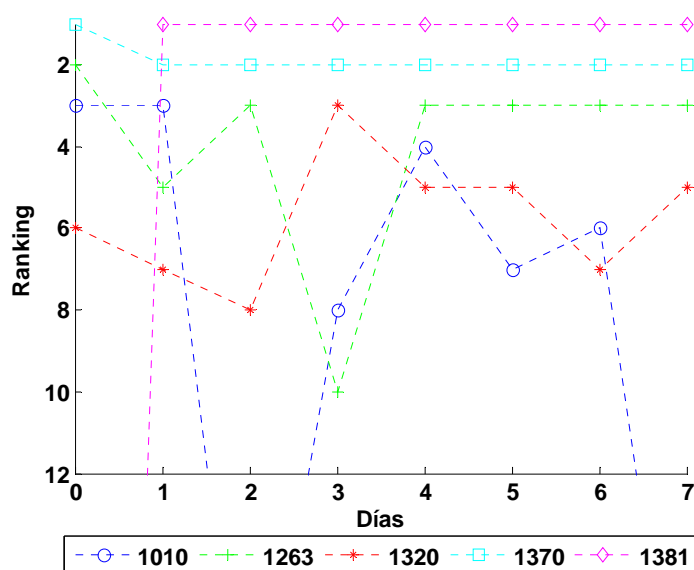


Figura 10. Ranking de accesos en el Periodo 4.

El hecho de que el vídeo anterior sea muy popular se ve reflejado en la Fig. 11, donde el valor de θ en su último día es de 2,37, hecho también ratificado en el criterio 3. En este caso dado que se introduce también un vídeo muy popular, la demanda será repartirá principalmente entre ambos vídeos. Este reparto ya aparece reflejado el primer día con un descenso en el valor de θ a 1,56, hecho que implicaría que el nuevo vídeo ha sido puesto en portada por la mañana, pues ya se registra un porcentaje de accesos de un 46% frente a un 26% del vídeo anterior. A partir de este momento, dado que como se puede observar en la Fig. 10 el nuevo vídeo permanece el primero en el ranking durante todo el periodo, según aumente o disminuya su porcentaje de accesos con respecto al vídeo anterior el valor de θ aumentará o disminuirá. Por ejemplo, en el segundo día de su publicación el valor de θ ha aumentado de 1,56 a 1,62, dado que el porcentaje de accesos sobre el primer vídeo ha aumentado en un 13%, y siendo los que acaparan el mayor porcentaje de accesos sobre el servicio con un total del 67%; sin embargo, en el

tercer día ha disminuido de 1,62 a 1,29, esto reflejaría que no sólo se ha acortado la diferencia entre el porcentaje de accesos en ambos, siendo de un 10%, sino que el porcentaje de accesos es más repartido entre el resto de los vídeos.

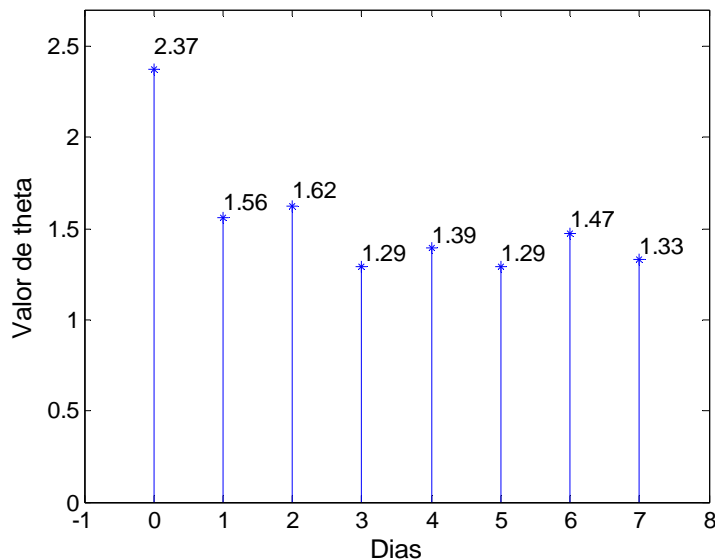


Figura 11. Ranking de accesos en el Periodo 4.

3.3. Conclusiones

Los cuatro periodos analizados permiten concluir que el valor del parámetro θ de la distribución Zipf-like es un buen estimador de la popularidad del servicio en periodos de estabilidad del contenido.

Si se introduce un nuevo vídeo con alta popularidad el valor de θ sufre un incremento inicial considerable superior a 2, manteniéndose el resto de los días con valores altos, superiores a 1,5 (Periodo 2). Sin embargo, si el vídeo anterior presenta también una alta popularidad, el valor de θ sufre un incremento inicial para luego descender y mantenerse en valores por debajo de 1,5, dado que el nuevo vídeo tiene que compartir su popularidad con el vídeo anterior (Periodo 4). En este último caso, dado que los dos últimos vídeos incorporados al sistema son muy populares acaparan el mayor porcentaje de accesos.

Por otro lado, si el vídeo nuevo tiene una baja popularidad, el valor de θ se incrementa paulatinamente para llegar a estabilizarse en valores relativamente altos

comprendidos entre 1,5 y 1,6 (Periodo 1). Dado que el nuevo vídeo es de baja demanda al competir con un vídeo anterior también de baja demanda, el nuevo siempre predominará en su periodo. Sin embargo, si el vídeo anterior tiene una popularidad alta, el valor de θ se mantiene bajo con valores comprendidos entre 1,15 y 1,35 (Periodo 3). Esto es debido a que el vídeo que realmente desean ver los clientes es el vídeo anterior, pero el cliente cuando entra en el sistema no tiene opción a elegir, pues automáticamente es lanzado el vídeo actual, creando el efecto de que el usuario lo está solicitando. Por lo tanto, el vídeo actual de baja demanda tendrá prácticamente el mismo porcentaje de accesos que el vídeo anterior, es decir, se está produciendo un reparto de accesos entre ambos vídeos de prácticamente el 60%.

Capítulo 4.
CARACTERÍSTICAS DE LOS
PERIÓDICOS DIGITALES BAJO
ANÁLISIS

4.1. INTRODUCCIÓN

El grupo Editorial Prensa Ibérica (EPI) da soporte a más de 16 periódicos digitales de ámbito autonómico, de los cuales sólo seis han sido elegidos para su estudio. El motivo de esta selección radicó en el número de accesos que presentaban, siendo superior a 50.000, ya que con pocos accesos el modelado de la popularidad se mostraba muy irregular, y por lo tanto casi imposible de encontrar un patrón de comportamiento. Los periódicos seleccionados han sido los siguientes: “La Opinión A Coruña” (www.laopinioncoruna.es) de A Coruña, “La Nueva España” (www.lne.es) de Asturias, “La Provincia” (www.laprovincia.es) de Islas Canarias, “Faro de Vigo” (www.farodevigo.es) de Vigo, y los dos últimos “Levante-EMV” (www.levante-emv.com) y “Superdeporte” (www.superdeporte.es) de Valencia. Dichos periódicos han sido estudiados durante un periodo de nueve meses, desde el 1 de Enero al 30 de Septiembre de 2009. En la Fig. 12 se puede observar el número de accesos, en porcentaje, que fueron efectuados sobre cada uno de los periódicos digitales seleccionados durante dicho periodo. El periódico "La Nueva España" es claramente el más solicitado con un 35% (586.398), en segunda posición se encuentra "Levante-EMV" con un 23% (383.031), "Superdeporte" en tercera posición con un 15% (238.114), "La Provincia" en cuarta posición con un 13% (217.285), y finalmente "Faro de Vigo" y "La Opinión A Coruña" con un 9% (153.199) y un 5% (85.758) respectivamente.

Porcentaje de accesos de Enero a Septiembre

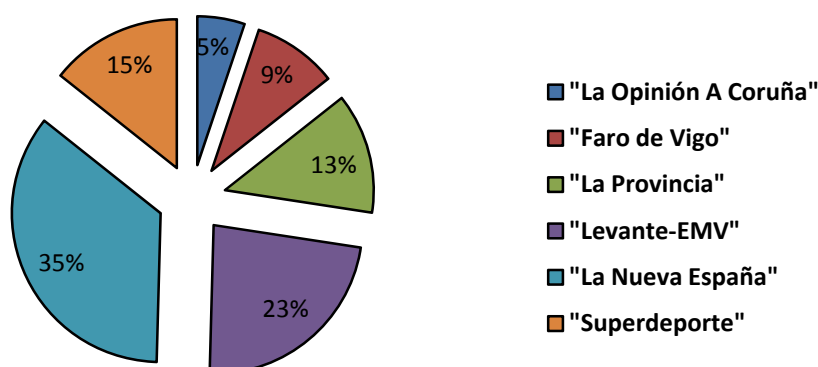


Figura 12. Distribución de la frecuencia de accesos (%) en cada periódico digital.

Estos periódicos digitales muestran los vídeos a los usuarios clasificados por categorías, siguiendo el mismo formato de presentación, y nada mejor que un ejemplo visual para aclararlo. En la Fig. 13 se puede observar la página web del servicio "La Nueva España" correspondiente a la categoría "Sociedad", la cual fue capturada el día 23 de Diciembre de 2010. En dicha imagen se puede observar que a la izquierda aparece la lista de todas las categorías que contienen vídeos con noticias que el usuario de la página podría seleccionar, y si se pinchase sobre cualquiera de ellas aparecería un submenú con las subcategorías asociadas a la categoría elegida, donde de nuevo el usuario debería seleccionar una. A continuación, el usuario vería los últimos doce vídeos que hubiesen sido incorporados al servicio, pertenecientes a la subcategoría seleccionada. Si el usuario deseara ver más vídeos con una fecha anterior debería de elegir la opción "Sigüientes", situada en la esquina inferior derecha. En definitiva, los vídeos son presentados al usuario en orden cronológico en función de su fecha de publicación en el servicio, del más al menos actual y en grupos de doce.

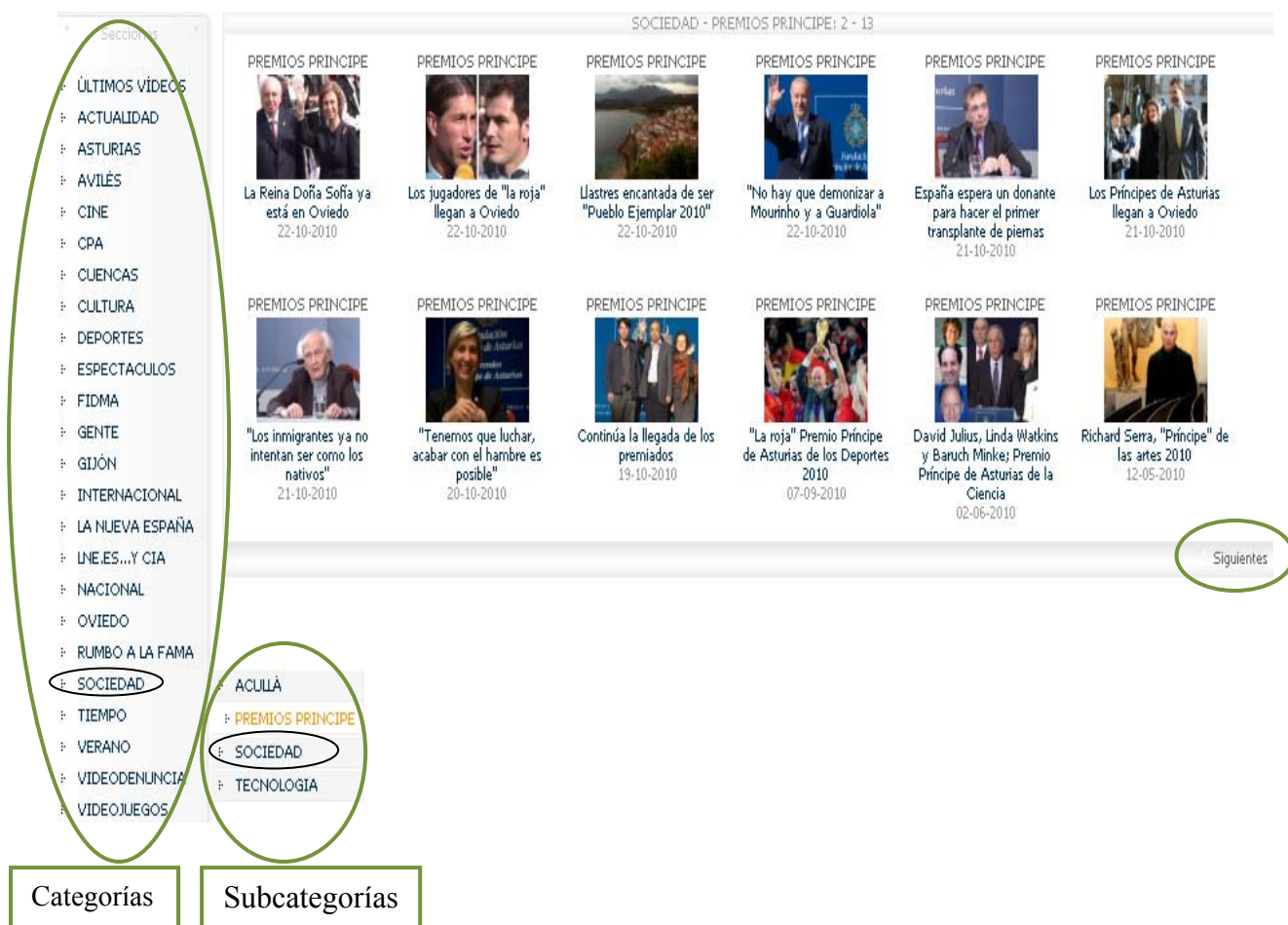


Figura 13. Página categoría/subcategoría "Sociedad" del periódico digital "La Nueva España".

Otros periódicos españoles, muy conocidos y con un alcance nacional, tales como “El país” (www.elpais.es), “El mundo” (www.elmundo.es) y “ABC” (www.abc.es), presentan un enlace en su página de inicio a una sección de vídeos denominado "Multimedia" o “Vídeos”. Además, cada periódico presenta sus vídeos a los usuarios de acuerdo a su propia política de publicación. Sin embargo, enlaces como "el más visto" y "los últimos vídeos" están presentes en cada uno de ellos. Algunos ejemplos de periódicos internacionales serían los siguientes: en U.K. "The Metro" (www.metro.co.uk) presenta sus vídeos clasificados en nueve categorías, en Francia "Le Monde" (www.lemonde.fr) clasifica sus vídeos en orden descendente de acuerdo a su fecha de publicación, y en U.S.A. el "New York Times" (www.nytimes.com) divide sus vídeos en 24 categorías, donde una de ellas está especialmente dedicada a noticias de ámbito local denominada "N.Y./Region".

4.2. ÁMBITO DE APLICACIÓN DE LA POPULARIDAD

La caracterización de la popularidad ha sido realizada en cada periódico digital a nivel independiente, y considerando en cada uno de ellos los vídeos como un conjunto único. Sin embargo, dado que las noticias presentadas a través de vídeos son mostradas al usuario de acuerdo a una posible clasificación de las mismas en función de su temática, también se ha llevado a cabo una caracterización a nivel de categoría. Dentro de este último aspecto, y con el objetivo de conseguir un enriquecimiento del estudio, en esta tesis se ha planteado un análisis comparativo del comportamiento de la popularidad entre los distintos periódicos digitales, y con este fin se han buscado las categorías comunes entre ellos, y en base a este criterio se ha decidido realizar un total de diez secciones en función de la temática de la categoría, las cuales han sido las siguientes: deportes, economía, autonómico, nacional, internacional, sociedad, cultura, sucesos, tiempo y otros. A cada sección se le ha asignado un color, tal y como aparece reflejado en la Tabla 8, con el fin de discernir de forma visual que categorías han sido agrupadas dentro de cada sección. La clasificación de las categorías por sección para cada periódico digital aparece reflejada en la Fig. 14.

"Faro de Vigo"	"La Nueva España"	"La Opinión A Coruña"	"La Provincia"	"Levante-EMV"	"Super-deporte"
Actualidad	Actualidad	Coruña/Metro	Canarias	Gente	Alpinismo
A pie de calle	Asturias	Ciencia	Carnaval	Ciencia	Atletismo
Curiosidades	Avilés	Cultura	Ciencia	Cine	Automovilismo
Deportes	Cine	Deportes	Concursos	Club diario Levante	Baloncesto
Economía	CPA	Economía	Cultura	Comunitat Valenciana	Fórmula I
Gran Vigo	Cuencas	El Tiempo	Deportes	Cultura	Fútbol
Internacional	Deportes	Encuesta	Economía	Deportes	Futbol Base
Política	Especiales	España	Encuentros Digitales	Economía	Juegos Olímpicos
Sociedad y cultura	Espectáculos	Galicia	España	El Tiempo	Juegos Paralímpicos
Sucesos	FIDMA	Mar	Fiestas	Internet	Motociclismo
	Gijón	Mundo	Fuerteventura	Internacional	Natación
	Internacional	Música	Gran Canaria	España	Polideportivo
	La Nueva España	Sociedad	Hoy en la provincia/DLP	Levante TV	Promoción
	LNE.ES...Y CIA	Sucesos	Lanzarote	Motor	Tenis
	Nacional	Televisión	Las Palmas GC	Música	Valencia
	Ocio		Mundo	Sociedad	Villareal
	Oviedo		Radio Canarias	Sucesos	
	Rumbo a la fama		Sociedad	Tecnología	
	Sociedad		Sucesos	TV	
	Tecnología		Tenerife		
	Tiempo		Tiempo		
	Turismo				
	Verano				
	VideoDenuncia				
	VideoJuegos				

Figura 14. Agrupación de las categorías en secciones en los seis periódicos digitales.

Deportes (S1)	Sociedad (S6)
Economía (S2)	Cultura (S7)
Autonómico (S3)	Sucesos (S8)
Nacional (S4)	Tiempo (S9)
Internacional (S5)	Otros (S10)

Tabla 8. Asignación de colores por categoría.

4.3. CARACTERÍSTICAS DE LOS PERIÓDICOS DIGITALES

4.3.1. Desde el punto de vista de las secciones

Para cada una de las secciones analizadas en los cinco periódicos digitales que las contienen, en la Tabla 9 se resume el número de vídeos solicitados y en la Tabla 10 el número de accesos, junto con el porcentaje que representa dicho número respecto al resto de las secciones dentro del mismo periódico, en el periodo comprendido entre el 1 de Enero y el 1 de Julio de 2009. A raíz de los datos presentes en dichas tablas, se puede observar que no todas las secciones son igual de populares, y por consiguiente podría ser interesante una comparación entre aquellas secciones que ocupasen por ejemplo los tres primeros puestos en cada uno de los periódicos.

Desde el punto de vista del número de vídeos, según la Tabla 9, resaltaría la sección S3 con un porcentaje que oscila entre el 74% (474) y el 27% (884), a excepción del periódico "Levante-EMV" donde la sección "Nacional" (S4) con un 18% (266), sería la que ocupa la primera posición. En segundo lugar, la sección "Sociedad" (S6) en el "Faro de Vigo" con un 17% (59), la sección "Sucesos" (S8) en "La Provincia" con un 20% (98), y la sección "Nacional" (S4) en "La Nueva España" con un 17% (548), tienen el segundo puesto al igual que sus homólogas, desde el punto de vista del número de accesos (Tabla 10). Sin embargo, la segunda posición en "La Opinión A Coruña" sería para la sección "Cultura" (S7) con un 6% (40), y en "Levante-EMV" la sección "Deportes" (S1) con un 14% (199). Respecto a la tercera posición, tampoco existe una

un comportamiento homogéneo entre los periódicos, con un porcentaje que oscila entre 6% (39) y 15% (50).

Secciones	Periódicos digitales				
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"
S1	39 (6,09%)	17 (5%)	32 (6,69%)	199 (14%)	258 (8,13%)
S2	5 (0,78 %)	20 (6%)	11 (2,31%)	98 (7%)	0
S3	474 (74,06%)	132 (39%)	216 (45,19%)	168 (11%)	884 (27,87%)
S4	11 (1,72%)	0	9 (1,88%)	266 (18%)	548 (17,28%)
S5	15 (2,35%)	20 (6%)	8 (1,67%)	92 (6%)	291 (9,17%)
S6	8 (1,25%)	59 (17%)	32 (6,69%)	106 (7%)	148 (4,67%)
S7	40 (6,25%)	0	26 (5,44%)	54 (4%)	0
S8	8 (1,25%)	50 (15%)	98 (20,50%)	169 (11%)	0
S9	19 (2,97%)	0	5 (1,05%)	130 (9%)	162 (5,11%)
S10	21 (3,28%)	40 (12%)	41 (8,58%)	192 (13%)	881 (27,77%)

Tabla 9. N° de vídeos accedidos.

Desde el punto de vista del número de accesos, según la Tabla 10, la sección que más destacaría (resaltada en color verde) es la correspondiente a noticias de ámbito autonómico (S3), hecho que remarcaría, aún más si cabe, la naturaleza de los periódicos, donde el número de accesos oscilaría entre los 30.216 (35%) de "Faro de Vigo" y los 95.110 (22%) de "La Nueva España". A continuación, la sección que ocuparía un segundo lugar (resaltada en color naranja) no coincide en todos los periódicos por igual, siendo en los periódicos de "Levante-EMV" y "La Nueva España" la sección "Nacional" (S4) con 13% (34.904) y un 20% (83.702) respectivamente, en

los periódicos "La Opinión A Coruña" y "La Provincia" la sección de "Sucesos" (S8) con un 6% (4.252) y un 20% (34.500) respectivamente, y finalmente en el periódico "Faro de Vigo" sería la sección "Sociedad" (S6) con un 18% (15.331). Por último, la sección que iría en tercera posición no coincidiría en ninguno de ellos (resaltada en color amarillo) con un porcentaje que oscilaría entre el 3% (2.307) y el 11% (47.063).

Secciones	Periódicos digitales				
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"
S1	2.307 (3,31%)	5.631 (6,66%)	3.157 (1,90%)	22.777 (8,76%)	47.063 (11,30%)
S2	121 (0,17%)	7.628 (9,03%)	2.737 (1,65%)	8.139 (3,13%)	0
S3	56.524 (80,98%)	30.216 (35,75%)	85.021 (51,16%)	64.472 (24,8%)	95.110 (22,84%)
S4	1.291 (1,85%)	0	303 (0,18%)	34.904 (13,43%)	83.702 (20,10%)
S5	352 (0,51%)	5357 (6,34%)	322 (0,19%)	12.870 (4,95%)	43.520 (10,45%)
S6	142 (0,2%)	15.331 (18,14%)	17.912 (10,78%)	14.911 (5,73%)	36.450 (8,76%)
S7	2.240 (3,21%)	0	4.991 (3%)	4.571 (1,76%)	0
S8	4.252 (6,09%)	8.067 (9,54%)	34.500 (20,76%)	29.817 (11,47%)	0
S9	1.724 (2,47%)	0	8.193 (4,93%)	27.636 (10,63%)	27.753 (6,66%)
S10	843 (1,21%)	12.291 (14,54%)	9.061 (5,45%)	39.873 (15,34%)	82.849 (19,89%)

Tabla 10. N° de accesos.

En la comparación de ambos casos no se ha tenido en cuenta la sección S10 (resaltada en gris), dado que es un compendio de secciones con distinta temática, con el fin de agrupar las secciones restantes.

En la Fig. 15 se representa una comparativa de los accesos por categoría entre los periódicos digitales, y se puede observar con claridad la superioridad de la categoría relativa a noticias de ámbito autonómico (S3) frente al resto de categorías.

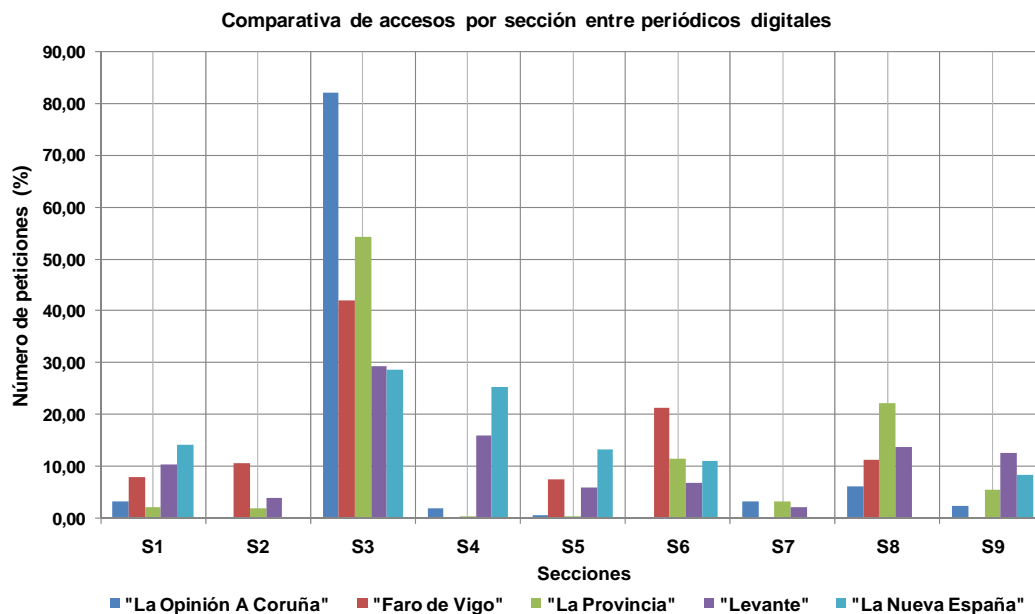


Figura 15. Porcentaje de accesos por sección en cada periódico digital.

Por otro lado, el periódico "Superdeporte", dedicado en su totalidad a los deportes tal y como su nombre indica, presentó un total de 162.467 accesos sobre 424 vídeos. Si tanto el número de vídeos accedidos, así como el número de accesos, se compara con los valores obtenidos en la sección de "Deportes" (S1), presente en los otros periódicos digitales, destaca la superioridad de este periódico en ambos aspectos. En concreto, si se compara con el periódico digital que presentó el mayor número en ambos factores, el cual sería "La Nueva España" con un total de 47.063 accesos sobre 258 vídeos, se puede deducir que el periódico "Superdeporte" obtuvo un número de accesos 3,45 veces superior y con un número de vídeos accedidos 1,6 veces superior.

Otro factor importante a tener en cuenta es el número de vídeos creados por sección. En la Tabla 11 se representa el número de vídeos creados desde el 1 de Enero de 2009 al 1 de Julio de 2009.

Secciones	Periódicos digitales				
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"
S1	12 (3,35%)	7 (7,14%)	7 (3%)	133 (12,37%)	152 (9,52%)
S2	0	6 (6,12%)	3 (1%)	79 (7,35%)	0
S3	297 (82,96%)	17 (17,35%)	165 (63%)	115 (10,7%)	82 (5,13%)
S4	2 (0,56%)	0	3 (1%)	254 (23,63%)	600 (37,57%)
S5	2 (0,56%)	9 (9,18%)	1 (0%)	84 (7,81%)	293 (18,35%)
S6	1 (0,28%)	23 (23,47%)	21 (8%)	77 (7,16%)	92 (5,76%)
S7	22 (6,14%)	0	11 (4%)	43 (4%)	0
S8	3 (0,84%)	12 (12,25%)	31 (12%)	89 (8,28%)	0
S9	13 (3,63%)	0	3 (1%)	127 (11,82%)	112 (7,01%)
S10	6 (1,68%)	24 (24,49%)	18 (7%)	74 (6,88%)	266 (16,66%)

Tabla 11. Número de vídeos creados por secciones.

Como se puede observar, la sección de ámbito autonómico (S3) y la sección "Nacional" (S4) son las que más vídeos nuevos han recibido en los periódicos "La Opinión A Coruña" y "La Provincia" por un lado, y "Levante" y "La Nueva España" por otro respectivamente, con un número de vídeos que osciló entre los 297 y los 600. Sin embargo, en el "Faro de Vigo" la sección que más vídeos incorporó fue la sección "Sociedad" (S6) con un total de 23. La segunda sección que más vídeos incorporó fue totalmente diferente para cada periódico, las cuales si se enumeran de mayor a menor serían las siguientes: la sección "Nacional" (S5) en "La Nueva España" con un total de 293 vídeos, la sección "Deportes" (S1) en "Levante-EMV" con 133, la sección "Sucesos" (S8) en "La Provincia" con 31, la sección "Cultura" (S7) en "La Opinión A

Courña" con 22, y por último la sección "Autonómico" (S3) en "Faro de Vigo" con 17. Finalmente, la tercera sección que más vídeos ha incorporado vuelve a ser dispar en casi todos los periódicos, oscilando entre los 11 y los 152 nuevos vídeos, a excepción de los periódicos "La Opinión A Coruña" y "Levante-EMV" que incorporarían 13 y 127 vídeos respectivamente en la sección "Tiempo" (S9). Adicionalmente, siguiendo con los comentarios sobre el contenido de la Tabla 10, existen secciones en las cuales aparece reflejado cero inserciones, como es el caso de las secciones "Nacional" (S4), "Cultura" (S7) y "Tiempo" (S9) en el "Faro de Vigo", o bien las secciones "Economía" (S2), "Cultura" (S7) y "Sucesos" (S8) en "La Nueva España", dado que dichas secciones no existen en los periódicos digitales mencionados, tal y como aparecía reflejado en la Fig. 14 del apartado 4.2 de este capítulo.

Además, a través del coeficiente de correlación de Pearson se ha comprobado la existencia de una relación lineal entre el número de vídeos creados con el número de vídeos accedidos y el número de accesos, donde los resultados se reflejan en la Tabla 12. En "La Opinión A Coruña" y "La Provincia" es donde la relación es más fuerte, con un coeficiente superior al 0,9 en ambos casos, así como en "Faro de Vigo" con un valor de 0,8. Por otro lado, en "Levante-EMV" existe una relación muy alta entre el número de vídeos accedidos y el número de vídeos creados ($r=0,9$), pero no con el número de accesos ($r<0,5$). Finalmente, en "La Nueva España" el número de vídeos creados presentó una relación baja con ambas variables, pues el coeficiente fue bajo con un valor inferior a 0,65.

Periódicos digitales					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"
Vídeos accedidos	0,9992	0,79	0,9659	0,9015	0,4781
Accesos	0,9966	0,8040	0,9722	0,4625	0,6435

Tabla 12. Resultados del coeficiente de Pearson (r).

En "La Nueva España" el número de vídeos publicados fue de 1.597, es un número bastante alto y por lo tanto es normal que muchos de dichos vídeos no reciban peticiones o bien sean escasas ($r=0,4781$). Desde el punto de vista de los accesos,

muchos de los vídeos que son novedad reciban más peticiones que los demás, sobre todo los que representen noticias de interés, sin embargo aunque capturen gran parte de las peticiones otros vídeos anteriores siguen captando la atención del usuario, y por lo tanto parecen no captar la mayoría de las peticiones ($r=0,6435$). En "Levante-EMV" la mayoría de los vídeos publicados reciben accesos ($r=0,9015$), con un total de 1.075, sin embargo, no parece que capturen la mayoría de las peticiones ($r=0,4625$). Estos dos periódicos ofrecen una mayor diversidad de contenido al usuario, y por lo tanto la elección del usuario tiende a ser más dispersa que en los anteriores. En el caso de "La Nueva España" el número de vídeos publicados es 20 veces superior al "Faro de Vigo", 6 veces superior a "La Provincia", y 4,5 veces superior a "La Opinión A Coruña", donde el número de vídeos publicados en cada una de ellos fue de 81, 263 y 358 respectivamente. En "Levante-EMV" el número de vídeos publicados es 20 veces superior al "Faro de Vigo", 4 veces superior a "La Provincia" y 3 veces superior a "La Opinión A Coruña".

4.3.2. Desde el punto de vista de los días de la semana

Dado que la incorporación de nuevos vídeos a los diferentes periódicos digitales es diaria, en este apartado se pretende ofrecer una visión desde el punto de vista de los días de la semana, aunque también se acompaña considerando el periodo completo. En las Tablas 13, 14, 15, 16, 17 y 18 se presenta para cada periódico digital, tanto el número medio de vídeos que han sido accedidos, como el número de accesos, donde en este último se representa a través de cinco datos, los cuales son el número total de accesos, el número mínimo, el número máximo, la media y la desviación típica. Además, si sólo se considera el número de accesos, el día de la semana que ocupa el primer lugar es resaltado en verde, el segundo en naranja, y el tercero en amarillo.

Intervalo Tiempo	Servicio "La Opinión A Coruña"					
	Nº medio vídeos	Nº accesos	Mínimo	Máximo	Media	σ
Completo	30	69.796	92	1.915	383	274
Lunes	30	9.792	152	1.719	377	317
Martes	31	11.035	178	1.347	424	263
Miércoles	31	11.883	176	1.407	457	307
Jueves	29	10.599	148	941	408	213
Viernes	32	9.745	113	866	375	211
Sábado	27	9.097	92	1.173	350	239
Domingo	29	7.645	96	1.915	294	340

Tabla 13. Número medio de vídeos y accesos en "La Opinión A Coruña".

Intervalo Tiempo	Servicio "Faro de Vigo"					
	Nº medio vídeos	Nº Accesos	Mínimo	Máximo	Media	σ
Completo	23	84.521	52	5.611	464	713
Lunes	23	10.096	102	1.603	388	408
Martes	24	10.756	81	1.940	414	475
Miércoles	26	18.199	63	4.111	677	877
Jueves	24	19.115	69	5.611	735	1.181
Viernes	21	15.651	52	3.287	602	843
Sábado	23	4.867	72	409	187	97
Domingo	22	5.837	82	1.574	224	291

Tabla 14. Número medio de vídeos y accesos en el "Faro de Vigo".

Intervalo Tiempo	Servicio "La Provincia"					
	Nº medio vídeos	Nº Accesos	Mínimo	Máximo	Media	σ
Completo	40	166.197	60	6.444	950	1.001
Lunes	40	24.195	64	3.412	968	1.005
Martes	41	19.431	116	3.666	777	816
Miércoles	40	24.572	85	3.367	983	890
Jueves	37	24.169	60	4.209	967	981
Viernes	42	24.701	202	3.442	988	931
Sábado	38	24.404	120	6.444	976	1.333
Domingo	43	24.725	167	5.418	989	1.131

Tabla 15. Número medio de vídeos y accesos en el "La Provincia".

Intervalo Tiempo	Servicio "Levante-EMV"					
	Nº medio vídeos	Nº Accesos	Mínimo	Máximo	Media	σ
Completo	55	259.960	29	12.077	1.494	1.134
Lunes	56	36.921	29	2.652	1.477	655
Martes	56	50.100	633	12.077	2.004	2.306
Miércoles	56	35.993	331	4.451	1.440	953
Jueves	52	38.808	587	4.126	1.617	780
Viernes	54	35.801	756	2.487	1.432	495
Sábado	53	27.984	1.119	363	2.793	638
Domingo	59	34.353	96	1.915	294	340

Tabla 16. Número medio de vídeos y accesos en "Levante-EMV".

Intervalo Tiempo	Servicio "La Nueva España"					
	Nº medio vídeos	Nº Accesos	Mínimo	Máximo	Media	σ
Completo	103	416.447	539	11.166	2.288	1.559
Lunes	105	60.593	539	7.763	2.330	1.674
Martes	100	63.603	690	6.205	2.446	1.308
Miércoles	97	70.210	1.043	11.166	2.700	2.344
Jueves	105	64.758	580	6.370	2.491	1.430
Viernes	99	66.863	878	7.725	2.572	1.517
Sábado	105	46.741	765	6.408	1.798	1.149
Domingo	108	43.679	541	5.107	1.680	974

Tabla 17. Número medio de vídeos y accesos en "La Nueva España".

Intervalo Tiempo	Servicio "Superdeporte"					
	Nº medio vídeos	Nº Accesos	Mínimo	Máximo	Media	σ
Completo	25	160.382	43	6.942	978	1.175
Lunes	25	19.548	67	4.104	814	961
Martes	26	22.553	50	4.635	902	1.078
Miércoles	24	22.188	43	4.757	887	1.023
Jueves	25	27.146	51	3.081	1.131	1.007
Viernes	25	27.601	59	6.942	1.104	1.532
Sábado	24	23.820	40	5.559	953	1.206
Domingo	25	19.603	51	6.055	784	1.274

Tabla 18. Número medio de vídeos y accesos en "Superdeporte".

A la vista de los resultados, aunque todos los días de la semana aparecen en algún momento en una de las tres primeras posiciones en alguno de los periódicos, los accesos parecen ser predominantes el martes, miércoles, jueves y viernes. La diferencia en el número de accesos efectuados en los diferentes periódicos entre los días de la semana que ocupan los tres primeros puestos es muy baja en prácticamente todos los casos. De hecho, la diferencia entre el primer y el segundo puesto es mínima en el periódico "La Provincia" y máxima en el periódico "Levante-EMV" siendo un 29% superior, en el resto de los periódicos la diferencia oscila entre un 2% en el caso de "Superdeporte" y un 8% en "La Opinión A Coruña". La diferencia entre el segundo y el tercer puesto es un poco más alta, y oscila como mínimo entre un 1% en el periódico "La Provincia" y alcanzando su valor máximo en el periódico "Faro de Vigo" con un 16%.

En la Fig. 16 se puede observar una comparativa de como varía el porcentaje de accesos entre los distintos periódicos digitales para cada día de la semana. El lunes presenta pequeñas variaciones entre ellos, donde el mínimo sería un 12% en los periódicos "Faro de Vigo" y "Superdeporte", y siendo en el resto de un 14% aproximadamente. En el martes se destacaría claramente "Levante-EMV" con un 19%, en segundo lugar estaría "La Opinión A Coruña" con casi un 16%, y en tercer y cuarto lugar "La Nueva España" y "Superdeporte" con un 15% y un 14% respectivamente, el resto estaría entorno al 12%. El miércoles, jueves y viernes destaca el periódico "Faro de Vigo", donde en los dos primeros días destacaría con notable superioridad con un 21% y un 22% respectivamente, siendo en el caso del viernes de un 18%, para el resto de los periódicos el porcentaje de accesos estaría comprendido entre un 13% y un 17%. En contraposición, el periódico "Faro de Vigo" es el que menos porcentaje de accesos tiene el sábado y el domingo, con un 6% y un 7% respectivamente, donde en los demás periódicos oscilaría entre un 10% y un 14%.

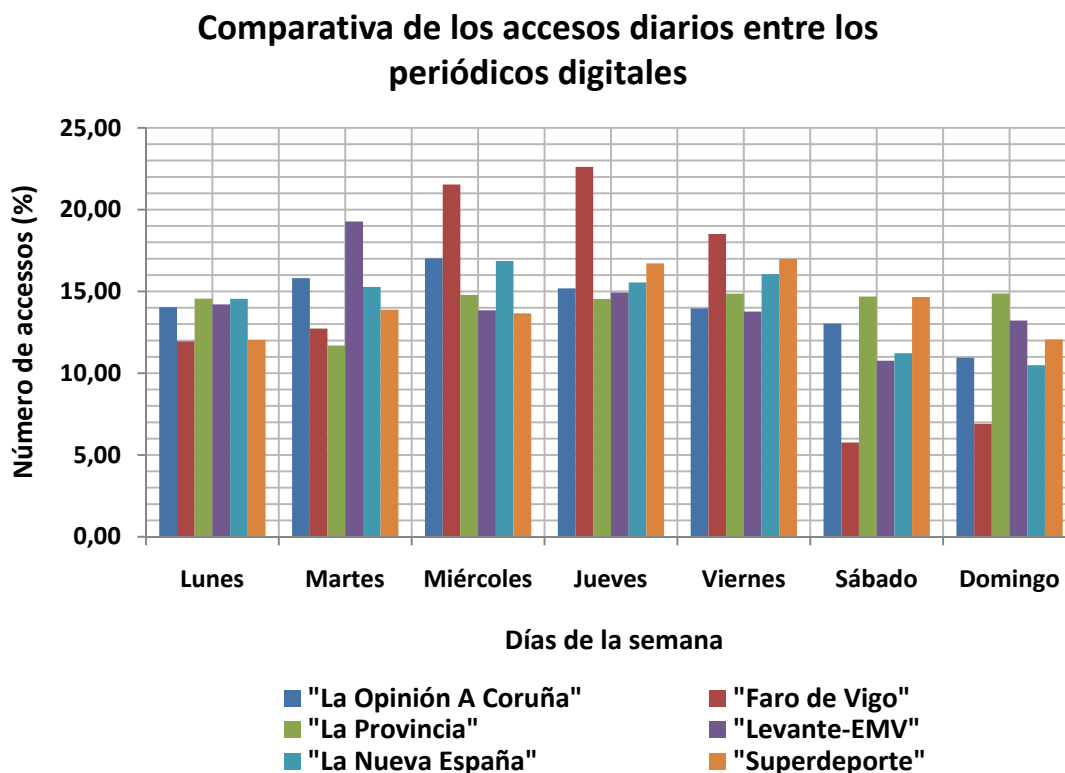


Figura 16. N° de accesos diarios en los distintos periódicos digitales.

En la Tabla 19 se presenta el número de vídeos creados cada día de la semana y en cada periódico. Además, para cada periódico se resalta aquellas casillas que ocupan los tres primeros puestos, donde el primer lugar es resaltado en color verde, el segundo en naranja y el tercero en amarillo, siguiendo la misma filosofía que en tablas anteriores. Si se establece una relación entre los resultados presentados desde la Tabla 13 a la Tabla 18 y la Tabla 19, y se considera para su comparación los tres primeros puestos, en general se puede concluir que los días en los cuales más vídeos han sido publicados coincide con alguno de los días en los cuales se han producido un mayor número de accesos. Otro aspecto a destacar es el hecho de que las inserciones de vídeos que se producen un día pudiesen tener una repercusión en el número de accesos del día siguiente. Así por ejemplo, en el periódico "Levante-EMV" el lunes es el tercer día de la semana con más accesos, sin embargo el número de vídeos incorporados dicho día sólo representa el 6% (65 vídeos nuevos), esto podría ser debido a las inserciones realizadas el día anterior, de hecho el domingo se produjo un gran número de inserciones, siendo el día más destacado de la semana con casi un 20% (211 vídeos nuevos). Este mismo efecto parece ocurrir en el resto de los periódicos, es decir, los efectos de haber

introducido un elevado número de vídeos un día aparecen visibles al día siguiente, hecho que no ha sido constatado en esta tesis, y que se plantea como una ampliación futura.

Días de la semana	Número de vídeos creados					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"	"Superdeporte"
Lunes	4 (1,12%)	0	25 (10%)	65 (6,04%)	150 (8,25%)	20 (8%)
Martes	61 (17,04%)	22 (22,45%)	51 (19%)	180 (16,73%)	296 (16,27%)	48 (21%)
Miércoles	3 (0,84%)	2 (2,04%)	42 (16%)	59 (5,48%)	132 (7,26%)	13 (5%)
Jueves	56 (15,64%)	21 (21,43%)	45 (17%)	181 (16,82%)	375 (20,61%)	48 (21%)
Viernes	59 (16,48%)	18 (18,37%)	29 (11%)	214 (19,89%)	287 (15,78%)	32 (14%)
Sábado	113 (31,56%)	13 (13,26%)	39 (15%)	166 (15,43%)	290 (15,94%)	37 (16%)
Domingo	62 (17,32%)	22 (22,45%)	32 (12%)	211 (19,61%)	289 (15,89%)	34 (15%)
Total	358	98	263	1.076	1.819	232

Tabla 19. Número de vídeos creados por días de la semana de Enero a Junio.

4.4. Conclusiones

El estudio de la tesis se centrará en los seis periódicos digitales con el mayor número de accesos. En todos los periódicos, los vídeos relativos a noticias se presentaban al usuario clasificados por categorías en función de su temática. Por lo tanto, con el objetivo de establecer un marco común de referencia en el análisis de la popularidad de los vídeos, se buscaron las categorías comunes entre los seis periódicos, y se establecieron diez secciones (nacional, internacional, autonómica, etc).

Tanto desde el punto de vista del número de vídeos presentes por sección, como desde el punto de vista del número de accesos, la sección correspondiente al ámbito autonómico resultó ser la más destacada, hecho que resalta más el carácter regional de

este tipo de periódicos. En el caso específico de los deportes, el periódico "Superdeporte" sobresale con bastante superioridad respecto a la sección "Deportes" del resto de los periódicos. Por otro lado, en función del periódico unas secciones publican más vídeos nuevos que otras, siendo la sección "Nacional", "Autonómica" y "Sociedad" las más destacadas. La popularidad de unas categorías frente a otras fue corroborada en otros servicios, como es el caso de YouTube [44, 45].

Dado que la incorporación de nuevos vídeos a los diferentes periódicos digitales es diaria, también se ofrece una visión desde el punto de vista de los días de la semana, tanto del número de vídeos publicados y accedidos, como del número de accesos. En el análisis se comprobó que los días en los que más vídeos eran publicados coincidía con los días en los que había un mayor número de accesos, o bien con el día siguiente de su publicación.

Capítulo 5.
RESULTADOS DEL AJUSTE
MEDIANTE DISTINTAS
DISTRIBUCIONES
ESTADÍSTICAS

5.1. Introducción

El objetivo de este capítulo es la caracterización de la popularidad de los vídeos presentes en los seis periódicos digitales de ámbito autonómico con el mayor número de accesos, con el fin de valorar posibles ampliaciones de recursos en el servicio de NoD. El estudio del modelado ha sido realizado en el periodo comprendido entre el 1 de Enero de 2009 y el 1 de Julio de 2009, ambos inclusive, siendo un total de 182 días, en base a dos criterios. En primer lugar, el ajuste se ha realizado teniendo en cuenta los accesos sobre el conjunto de vídeos vistos en su totalidad, es decir, sin tener en cuenta la sección a la que pertenecen. En segundo lugar, dado que los vídeos son presentados al usuario clasificados por categorías, también ha sido realizado un ajuste a nivel de sección, tal y como fue explicado en el capítulo 4.

El estudio del comportamiento sobre el patrón que sigue la demanda en los accesos a los distintos videos presentes en cada uno de los periódicos, ha sido realizado día a día, dado que la inserción de contenido es diaria. Por lo tanto, se trataría de analizar la popularidad en periodos de estabilidad del contenido [51]. En consecuencia, los accesos diarios serán considerados, en lo que a términos estadísticos se refiere, como una muestra. Para cada una de estas muestras se buscará el mejor ajuste posible en cada una de las distribuciones estadísticas descritas con anterioridad, como son Zipf-like con y sin k-transformación, two Zipf-like, Mandelbrot y Stretched. A estas distribuciones se añade un nuevo método estadístico denominado transformación de Box-Cox que pretende ser una nueva aportación a este campo. Por otra parte, se ha tratado de analizar diferencias de comportamiento en el acceso entre los distintos días de la semana. Por este motivo, los días se han agrupado en siete clases, donde cada una de ellas representaría un día diferente de la semana. De esta forma, todos los días que se correspondiesen con un lunes serian agrupados en una clase, los correspondientes a los martes en otra, y así sucesivamente.

El proceso seguido ha sido reflejado en la Fig. 17, y sería aplicado para cada uno de los periódicos digitales, tal y como se describe a continuación.

- ✓ El estudio del comportamiento sobre el patrón que sigue la demanda en los accesos a los distintos vídeos presentes en el servicio ha sido realizado día a día.

Por consiguiente, los accesos diarios serán considerados, en lo que a términos estadísticos se refiere, como una muestra.

- ✓ Para cada muestra se buscará el mejor ajuste posible en cada una de las distribuciones estadísticas descritas con anterioridad, como son Zipf-like con y sin k-transformación, two Zipf-like, Mandelbrot, Stretched y Box-Cox.
- ✓ Es necesario un proceso de filtrado, con el fin de eliminar aquellos modelos que no podrían ser considerados como un buen ajuste de los datos, a través de técnicas estadísticas destinadas a medir la bondad del ajuste, como serían el contraste de hipótesis mediante Kolmogorov-Smirnov o Chi-cuadrado.
- ✓ Estudio diario de los accesos, los cuales se agruparán en siete clases en función del día de la semana correspondiente.
- ✓ Una vez hecha la clasificación se trata de buscar un modelo general que describa el comportamiento de la popularidad en su totalidad y de cada clase en particular.

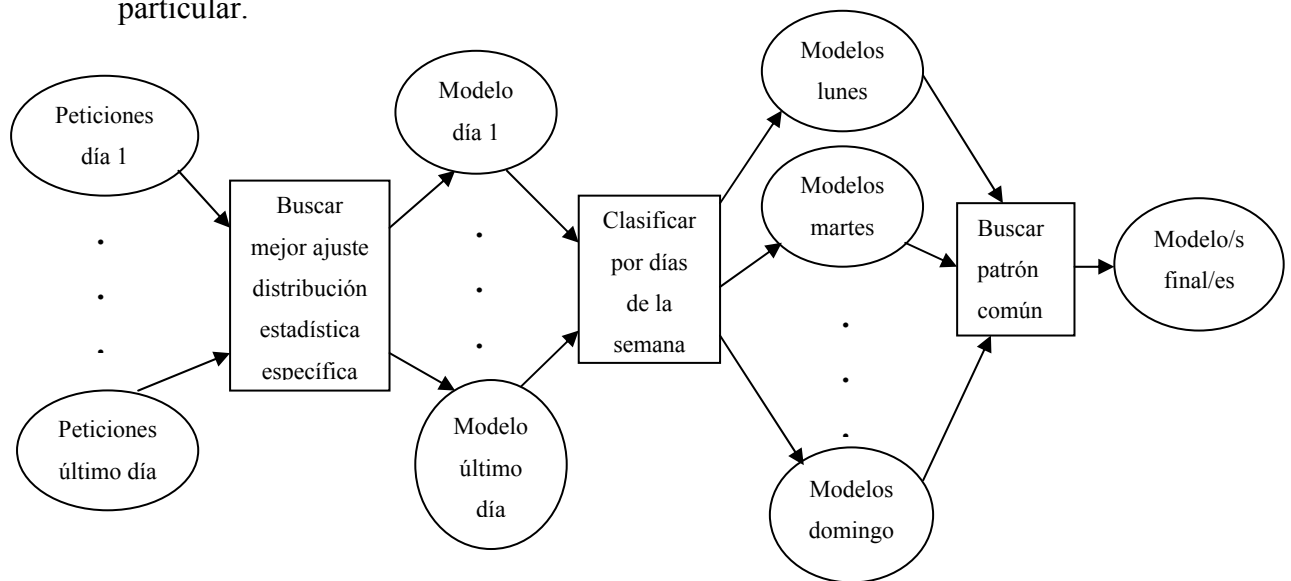


Figura 17. Esquema para la obtención de patrones de comportamiento.

Además, con el fin de analizar la validez del conjunto de valores obtenidos a lo largo de las distintas muestras, y comprobar si existen diferencias de comportamiento entre las siete clases, en cada uno de los parámetros correspondientes a las diferentes distribuciones, se han aplicado una serie de tests estadísticos, los cuales han sido los siguientes:

- ✓ Test de Shapiro-Wilk: se usa para comprobar si el conjunto de valores del parámetro que provienen de las muestras sigue una distribución normal.
- ✓ Test de Bartlett: en el caso de que el test de Shapiro-Wilk resulte positivo ($p_value > 0,05$) para un parámetro, se comprobará si el conjunto de valores del mismo presenta la misma varianza entre las distintas clases.
- ✓ Test de Levene: tiene la misma finalidad que el test de Bartlett, pero este test se usaría si el test de Shapiro-Wilk fuese negativo ($p_value < 0,05$).
- ✓ Test F-Anova: se usa para comprobar si el conjunto de valores del parámetro presenta la misma media entre las distintas clases.

5.2. Distribución Zipf-like

En la Fig. 18 se presenta el porcentaje de días desde el 1 de Enero de 2009 a 1 Julio de 2009, en los cuales ha sido posible aplicar la distribución Zipf-like con y sin k-transformación. En "La Nueva España" y en "Levante-EMV" ha sido posible aplicar la distribución sobre prácticamente el 100% de los días, seguido de "La Opinión A Coruña" con el 86%, "Faro de Vigo" y "La Provincia" con el 68% y 67% respectivamente, y en último lugar "Superdeporte" con un 37% de cobertura. Desde el punto de vista de la técnica de k-transformación no aporta una mejoría o bien no es significativa, tal es el caso de los periódicos "La Nueva España" y "La Provincia" donde la mejora es nula, y en el resto oscilaría entre el 1% y el 5%.

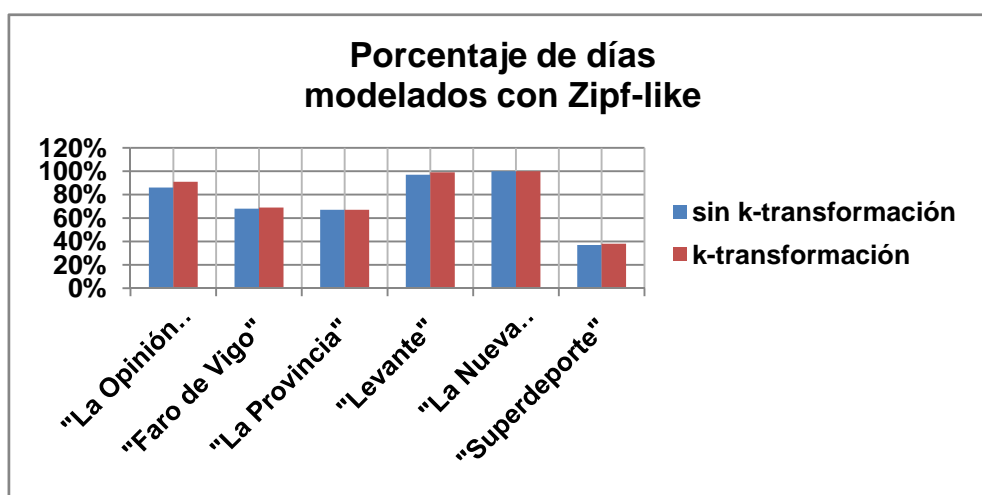


Figura 18. Porcentaje de días modelados con Zipf-like.

Todos los resultados conseguidos con esta distribución se encuentran en el Anexo I en más detalle.

5.2.1. Periódico "La Opinión A Coruña"

Desde el punto de vista del estudio diario de la demanda durante todo el periodo, el número de días modelados fue de 158 (87%). El conjunto de valores del parámetro θ siguió una distribución normal según el test de Shapiro-Wilk ($p_value = 0,969 > 0,05$), siendo el valor medio de θ de 1,5971 ($\sigma: 0,2242$), y con un grado medio de ajuste lineal R de 0,9457 ($\sigma: 0,02$). Con la técnica de la k-transformación se consiguió modelar un 5% más, originando un ligero incremento en el valor medio de θ , alcanzando el valor de 1,7875 ($\sigma: 0,6535$), y con un grado medio de ajuste lineal R de 0,9415 ($\sigma: 0,0332$). La Fig. 19 presenta un ejemplo de ajuste en los accesos realizados el día 10 de Enero de 2009.

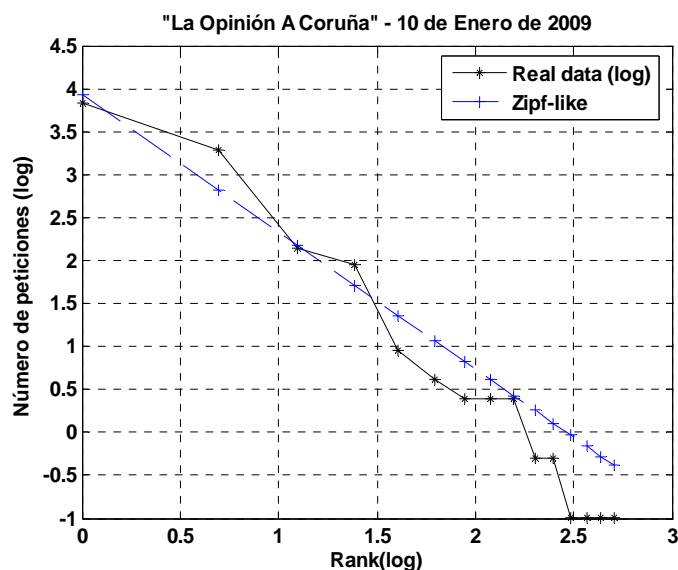


Figura 19. Distribución Zipf-like con $\theta = 1,5971$, $R = 0,9498$

Desde el punto de vista de las secciones, solamente se podría plantear un modelo para la sección S3, con un total de 94 días (52%) modelados. El valor medio del parámetro θ sería de 1,7334 ($\sigma: 0,2656$) con un ajuste lineal medio R de 0,952 ($\sigma: 0,0198$). Si se aplicase una k-transformación se conseguiría modelar un 7% más. En el resto de las secciones, el porcentaje alcanzó un máximo de un 5% (10 días) e incluso en algunas secciones el porcentaje fue de un 0%, quedando descartado el planteamiento de

cualquier modelo específico para cada una de ellas, y por lo tanto en su defecto sólo daría pie a la aplicación del modelo general.

5.2.2. Periódico "Faro de Vigo"

El número de días modelados ha sido inferior al servicio "La Opinión A Coruña" con un total de 123 días (68%). Desde el punto de vista del estudio diario sobre el periodo completo, el parámetro θ según el test para la media F-Anova señalaba que no presentaba el mismo valor medio en los siete días de la semana. En la Fig. 20, se representa una comparación de la media del parámetro θ , donde se puede observar que el valor medio del parámetro θ del domingo y del sábado es diferente del viernes, y donde el par (sábado, jueves) estaría en el límite de ser considerado aceptable.

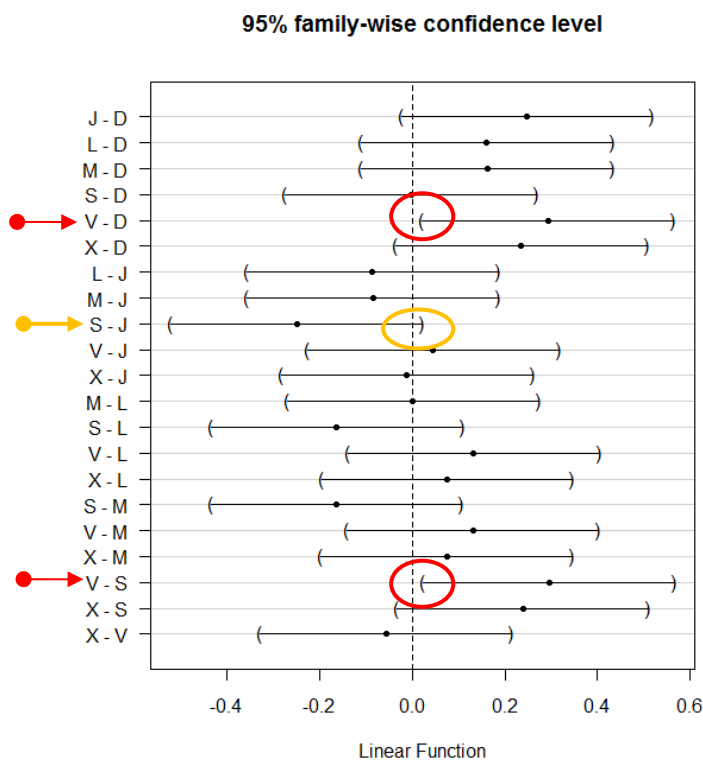


Figura 20. Test F-Anova para θ entre pares de días de la semana.

Por lo tanto, de acuerdo a los resultados del test se han planteado dos modelos de comportamiento, el modelo que englobaría los accesos realizados de lunes a viernes, y el modelo que describiría la demanda durante los fines de semana. En el primer caso, el valor medio del parámetro θ ha sido de 1,6273 (σ : 0,3659) con un ajuste lineal medio R de 0,9532 (σ : 0,0175), y en el segundo caso de 1,3687 (σ : 0,2023) siendo el valor medio

de R de 0,9518 (σ : 0,0162). Con la técnica de k -transformación se consiguió un incremento insignificante de un 1% en el número de días modelados, en concreto fueron 125 días (69%).

Desde el punto de vista del estudio por secciones, en todas ellas se alcanzó un grado de cobertura igual o inferior a 9 días ($\leq 5\%$), y por lo tanto no se calcula un modelo concreto para ninguna de ellas. En su defecto se aplicaría el modelo general, al igual que ocurrió en el servicio “La Opinión A Coruña”. Con el uso de la k -transformación se consiguió un incremento en el grado de cobertura, aunque sigue siendo bastante bajo, en concreto para la sección S3 se consiguió el modelado de 21 días (11%), pasando de un grado de cobertura del 5% al 11%.

5.2.3. Periódico "La Provincia"

El número de días modelados fueron 117 (67%). Se podría plantear un modelo general donde el valor medio de θ fuese de 1,6084 (σ : 0,3521), y con un grado medio de ajuste lineal R de 0,9338 (σ : 0,0307). Desde el punto de vista de las secciones, todas ellas plantean un grado de cobertura igual o inferior a 5 días ($\leq 3\%$), siendo la sección S3 la que ofrece un grado de cobertura superior que llegaría a los 62 días (34%). Por lo tanto, estos resultados no conllevarían a plantear ningún modelo específico para ninguna de las secciones, excepto la sección S3, donde el valor medio del parámetro θ sería de 1,686 (σ : 0,5401) con un ajuste lineal medio R de 0,9399 (σ : 0,0222).

Por otro lado, desde el punto de vista de la k -transformación no ofrece ninguna mejoría o bien una mejora sustancial. En el modelo general sin secciones se ha obtenido el mismo grado de cobertura del 67%, y en el caso de las secciones sigue siendo la sección S3 la que ofrece un grado de cobertura superior, con un total de 70 días (38%), siendo para el resto de las secciones igual o inferior a 11 días ($\leq 6\%$). En definitiva, desde el punto de vista de las secciones el grado de cobertura es inferior al 50%.

5.2.4. Periódico "Levante-EMV"

En este periódico se han conseguido modelar 170 días (97%) de los 182 posibles. Sin embargo, según el test F-Anova no sería correcto plantear un modelo único para todos los días de la semana al igual que ocurre en el servicio "Faro de Vigo", dado que la media del parámetro θ para los siete días de la semana no se puede considerar igual. Este hecho se puede apreciar en la Fig. 21, donde se representa una comparación de las medias del parámetro θ entre pares de días de la semana. Claramente, se puede observar que el domingo tiene una media distinta respecto al jueves, y habría dos casos que estarían en el límite de ser aceptados como válidos, el par (viernes, domingo), y el par (sábado, jueves).

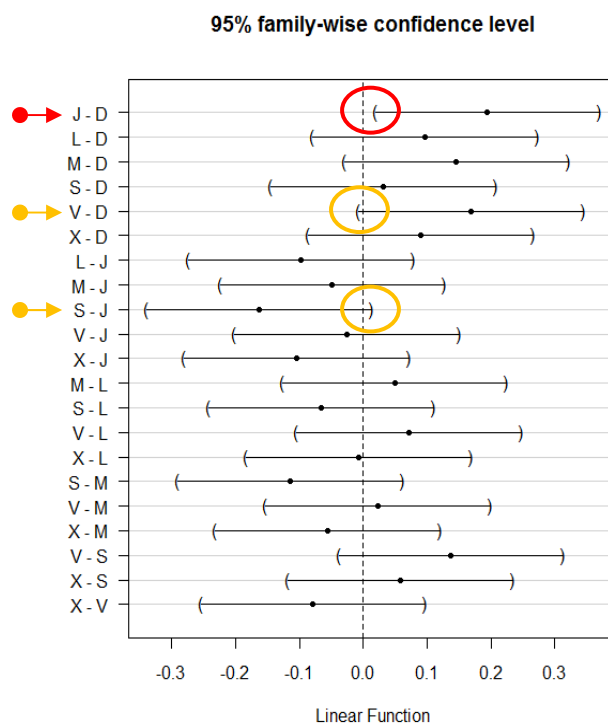


Figura 21. Test F-Anova para θ entre pares de días de la semana.

En consecuencia, a la vista de los resultados se plantea un modelo general que englobaría los días de lunes a viernes, y otro modelo que englobaría el sábado y el domingo. Este mismo planteamiento fue realizado en el servicio "Faro de Vigo". En el primer caso, el valor medio del parámetro θ ha sido de 1,8162 (σ : 0,1907) con un valor medio para el ajuste lineal R de 0,9527 (σ : 0,0187), y en el segundo caso el valor medio del parámetro θ ha sido de 1,6896 (σ : 0,2364) con un valor medio para R de 0,9550 (σ : 0,0197). Desde el punto de vista de las secciones, en todas ellas se plantea un grado de

cobertura igual o inferior a 20 días (11%), y por lo tanto para ninguna se puede plantear un modelo específico, al igual que ocurrió en el servicio “Faro de Vigo”.

Con la técnica de la k-transformación, el servicio en su globalidad tampoco aportó una mejora sustancial, alcanzando a modelar 174 días (99%), un 2% más, al igual que desde el punto de vista de las secciones, donde se obtuvo un grado de cobertura igual o inferior a los 44 días (24%).

5.2.5. Periódico "La Nueva España"

Este periódico presenta un grado de cobertura superior a cualquiera de los anteriormente analizados, pues se ha llegado a modelar el conjunto completo de los días, siendo un total de 182 días (100%). Sin embargo, según el test F-Anova aplicado sobre el parámetro θ englobando los siete días de la semana, no se puede considerar que la media sea la misma en todos. Tal y como se puede apreciar en la Fig. 22, el valor medio de θ el domingo tendría un comportamiento distinto respecto al viernes, y al igual que en el periódico “Levante-EMV” habría dos casos que estarían al límite de ser considerados válidos aunque no se trate de los mismos casos, el par (martes, domingo) y el par (miércoles, domingo).

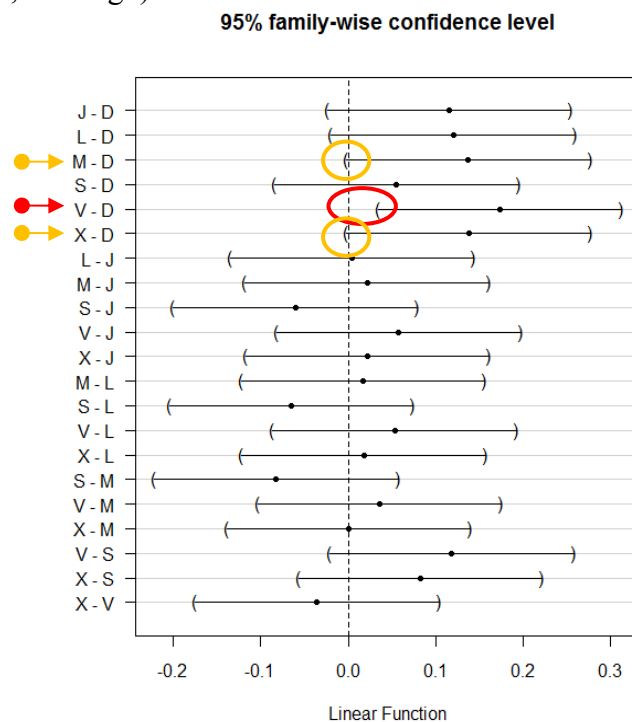


Figura 22. Test F-Anova para θ entre pares de días de la semana.

Por consiguiente, se plantea un modelo general que englobaría los días de lunes a sábado, y otro que representaría solamente el domingo. En el primer caso el valor medio del parámetro θ ha sido de 1,5570 (σ : 0,1797) con un valor medio para el ajuste lineal R de 0,9661 (σ : 0,0168), y en el segundo caso fue de 1,4429 (σ : 0,1293) con un valor medio de R de 0,9638 (σ : 0,0161). Desde el punto de vista de las secciones, al igual que ocurrió en el periódico “La Opinión A Coruña”, cabe destacar la sección S3 con un grado de cobertura de 136 días (75%), a la cual habría que añadir la sección S10 con un grado de cobertura de 165 días (91%), siendo en el resto de las secciones igual o inferior a 14 días (8%). Por lo tanto, el planteamiento de un modelo para la sección S3 sería factible al igual que para la sección S10, pero también serían dos los modelos a plantear, uno de lunes a sábado y otro para el domingo para la sección S3, y para la sección S10, uno que englobaría de lunes a domingo menos el viernes y otro para el viernes. El estudio de la sección S3 ha llevado a los siguientes resultados, de lunes a sábado el valor medio del parámetro θ sería de 1,4592 (σ : 0,279) con un valor medio de R de 0,9193 (σ : 0,0463), y en el caso del domingo el valor medio del parámetro θ sería de 1,2695 (σ : 0,2349) y con un valor medio de R de 0,9179 (σ : 0,0448). En el caso de la sección S10 el valor medio de θ de lunes a domingo menos el viernes sería de 1,2036 (σ : 0,2992) con un valor medio de R de 0,9439 (0,0238), y para el viernes θ tendría un valor de 1,4685 (σ : 0,1963) con un valor medio para R de 0,9552 (σ : 0,021).

Al igual que en los periódicos anteriores la k-transformación sigue sin aportar mejora para el caso general, y en el caso de las secciones para la sección S3 supondría el modelado sobre 144 días (79%), suponiendo un incremento del 4%, la sección S10 no tendría variación, y en el resto de las secciones es igual o inferior a 37 días (20%), lo cual supondría un incremento del 12%.

5.2.6. Periódico "Superdeporte"

Este periódico se podría considerar un caso especial dado que está dedicado en su totalidad al mundo de los deportes. En él se consigue modelar 66 días (36%). El valor medio del parámetro θ obtenido fue de 1,7668 (σ : 0,4219) con un valor medio en el ajuste lineal R de 0,9348 (σ : 0,0218). Si se tiene en cuenta la k-transformación el porcentaje de cobertura aumentaría en un 2%.

5.3. Combinación de dos Zipf-like (two Zipf-like)

Se pretende encontrar una distribución Zipf-like que ajuste las peticiones de los vídeos más demandados (cabeza), y otra para los menos demandados (cola), pero ¿dónde está la frontera entre ambas zonas?, para este fin se prueba con distintas particiones para el conjunto inicial de peticiones. En esta tesis se plantea un algoritmo, representado en la Tabla 20, en el cual se prueba a dividir el conjunto de peticiones inicial en dos partes, que representarían las dos zonas mencionadas anteriormente, y a continuación para cada zona se calcularía el mejor modelo Zipf-like que ajuste el conjunto de peticiones asociado. Si el modelo final, que estaría formado por los dos modelos Zipf-like, caracteriza mejor al conjunto total de peticiones, será el modelo elegido, siendo en caso contrario rechazado.

```
n=Número de vídeos accedidos;
P = {p1, p2, ..., pn | peticiones ordenadas de forma decreciente, siendo pi el número de
    peticiones del vídeo que ocupa el i-ésimo lugar,
    según dicha ordenación}
error_minimo = ∞;

para punto_corte = 2 hasta (n-2) hacer

    Peticiones_1 = P{p1, ..., ppunto_corte};
    Peticiones_2 = P{ppunto_corte + 1, ..., pn};

    theta_1 = calcularZipfLike(Peticiones_1);
    theta_2 = calcularZipfLike(Peticiones_2);
    modelo_1 = calcularModeloZipfLike(theta_1);
    modelo_2 = calcularModeloZipfLike(theta_2);
    modelo_final = concatenar(P, punto_corte, modelo_1, modelo_2);

    error_modelo = calcularError(P, modelo_final);
    norma_modelo = calcularNorma(error_modelo);
```

```
si (norma_modelo < error_minimo)
    entonces error_minimo = norma_modelo;
        theta_1_minimo = theta_1;
        theta_2_minimo = theta_2;
        modelo_minimo = modelo_final;
        punto_corte_optimo = punto_corte;
finsi
finpara
```

Tabla 20. Algoritmo de búsqueda de la mejor partición para las peticiones.

Los resultados obtenidos para cada uno de los periódicos digitales son presentados de forma más precisa, desde el punto de vista numérico en el Anexo II, y de forma más general en los siguientes apartados. Pero dado que en esta tesis los vídeos más populares son los más importantes, aunque la demanda sobre los menos populares podría ser importante en ciertos contextos, los diferentes comentarios irán dirigidos hacia ellos principalmente.

5.3.1. Periódico "La Opinión A Coruña"

El modelado para los vídeos más populares (cabeza), ha sido posible sobre un total de 49 días (27%), aunque es un porcentaje bajo para establecer un modelo general. El punto de corte medio osciló entorno al vídeo 14 (σ : 5,79) sobre un total de 32 vídeos (σ : 12,32). El valor medio de θ obtenido fue de 1,5184 (σ : 0,4127) con un ajuste lineal R^2 de 0,9762 (σ : 0,0112). El número medio de accesos fue de 328, pero no es muy significativo dado que presenta una desviación típica muy alta, con un valor de 228. El número total de accesos fue de 69.796, donde el día con menos accesos fue el sábado 2 de Mayo de 2009 con un total de 92, y el día con más accesos el domingo 7 de Junio de 2009 con un total de 1.915.

Desde el punto de vista de las secciones, ninguna de ellas pudo ser modelada, a excepción de la sección S3, donde se consigue modelar todos los días del periodo, es decir, 182 días. Sin embargo, sólo se consiguen modelos válidos para 44 días (24%).

Dado que el porcentaje es bajo, se ha descartado el planteamiento de cualquier modelo asociado a dicha sección, y con más motivo a cualquiera de las restantes.

5.3.2. Periódico "Faro de Vigo"

La demanda sobre los vídeos más populares ha sido ajustada sobre un total de 77 días (42%), donde el número medio de vídeos populares ha resultado ser de 9 (σ : 6,2), sobre un total de 24 (σ : 7,27) vídeos por término medio, y con un número medio de accesos de 462 (σ : 696,92). El grado medio de ajuste lineal R^2 alcanzado ha sido de 0,9824 (σ : 0,0167), con un valor medio para el parámetro θ de 2,3864 (σ : 1,3156).

Si se tiene en cuenta la cola, sólo en 9 días de los 77 fue posible realizar un ajuste válido en ambas zonas simultáneamente. Un ejemplo de esta situación viene representada en la Fig. 23 que refleja el ajuste de las peticiones efectuadas el miércoles día 24 de Junio de 2009, en la cual se representa el modelo obtenido específicamente para dicho día, así como el modelo de los miércoles (Anexo II), con un ajuste lineal R^2 de 0,9909 y 0,9780 respectivamente. Lógicamente, el ajuste es mejor si se aplica el modelo específico que el general, aunque en este ejemplo sólo es ligeramente superior.

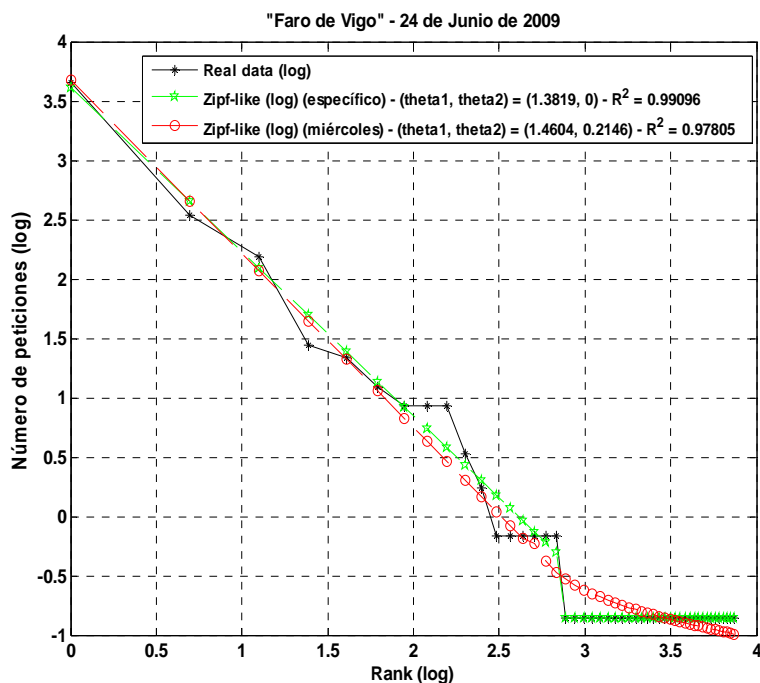


Figura 23. Ajuste con el modelo general del miércoles y el específico del día.

Desde el punto de vista de las secciones sólo la sección S3 consigue un mayor número de días modelados aunque sigue bastante bajo. El número de días modelados fue de 154, pero en sólo 39 días se consigue un ajuste válido (25%). El resto de las secciones el número de días para los que se obtuvo un modelo fue inferior a la sección S3, siendo igual o inferior a los 24 días. Por lo tanto, al igual que en el periódico "La Opinión A Coruña", no se plantea un modelo para ninguna sección.

5.3.3. Periódico "La Provincia"

El número de días modelados han sido 63 (36%) sobre un total de 175 días. El punto medio de corte osciló entorno al vídeo número 11 (σ : 10,72) sobre un número medio total de 40 vídeos (σ : 14,13). El valor medio de θ en la zona de la cabeza ha sido de 2,3916 (σ : 1,3134), con un ajuste medio lineal R^2 de 0,9852 (σ : 0,0148). El conjunto de los vídeos más populares recibió un valor medio de peticiones de 863 (σ : 959,73) respecto al número medio total de peticiones recibidas que osciló entorno a las 950 (σ : 1.009,56).

Desde el punto de vista de las secciones todas presentaron un bajo porcentaje de cobertura, siendo la sección S3 la que estaría un poco por encima de las demás con un porcentaje de un 34% (59 frente a un total de 174 días), frente al resto de secciones cuyo porcentaje es inferior al 19% (\leq 33 días). Por lo tanto, no se plantea ningún modelo en especial.

5.3.4. Periódico "Levante-EMV"

Desde el punto de vista de la cabeza, el número de días que se ha conseguido modelar ha sido de 59 (34%) sobre un total de 174, por lo que es un grado de cobertura bastante bajo para establecer un modelo genérico. Además, cabe destacar que no en todos los días ha sido posible que la *cabeza* y la *cola* ajustasen simultáneamente, sólo en 38 (22%) días de los 59. El punto de corte osciló entorno al vídeo número 19 (σ : 9,31) sobre un total de 58 vídeos (σ : 15,33). El valor medio de θ obtenido fue de 1,3951 (σ : 0,3365) con un ajuste lineal R^2 de 0,9714 (σ : 0,0156).

En la Fig. 24 hay un ejemplo de ajuste para el domingo 11 de Enero de 2009, donde se consigue un ajuste de tanto de la cabeza como de la cola para el modelo

específico del día ($\theta_1 = 0,1412$; $\theta_2 = 1,1843$), y donde el punto de corte está en el vídeo que ocupa la posición tercera sobre un total de 72 vídeos, con un ajuste lineal R^2 de 0,9819. Sin embargo, con la distribución Zipf-like correspondiente al modelo general del domingo, donde el punto de corte se encuentra en el décimo vídeo, se consigue un ajuste lineal muy bajo 0,8312.

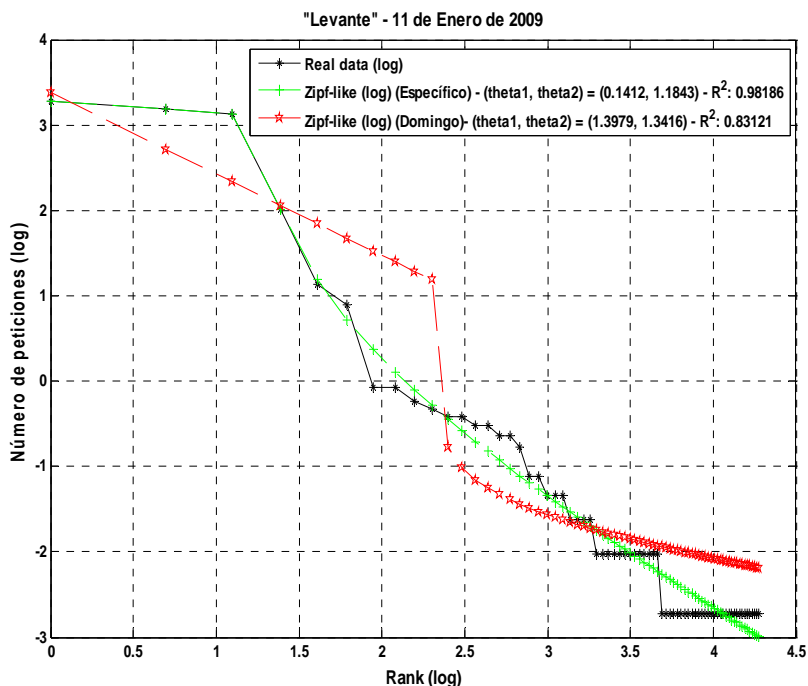


Figura 24. Ajuste de accesos con dos two Zipf-like.

En la Fig. 25 se muestra otro ejemplo de ajuste sobre los accesos realizados el domingo 18 de Enero de 2009, pero este día representaría un contraejemplo del anterior, pues el modelo general del domingo consigue un buen ajuste lineal R^2 de 0,9539. El punto de corte está en el vídeo de la novena posición sobre un total de 78, donde en el modelo específico θ tomó el valor de 0,9491 para la cabeza, y de 0,9766 para la cola, con un grado de ajuste lineal R^2 de 0,9846.

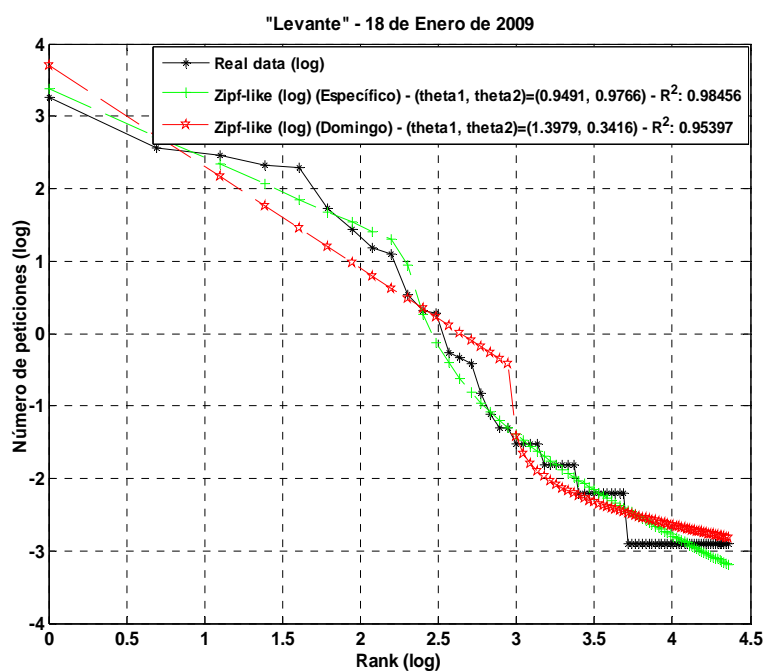


Figura 25. Ajuste de accesos con dos two Zipf-like.

Desde el punto de vista de las secciones, al igual que ocurre en el estudio de los servicios anteriores, tampoco da pie al planteamiento de modelo alguno. De hecho, la sección S8 que cuenta con el mayor número de días para los cuales se ha encontrado un ajuste válido, respecto al resto de las secciones, se han ajustado 38 días (26%) sobre un total de 143 días. Por lo tanto, para la mayor parte de los días no fue posible encontrar un buen ajuste, representando entonces un 26%. El resto de las secciones presentaron una cobertura con un número igual o inferior a los 34 días.

5.3.5. Periódico "La Nueva España"

En este periódico se ha conseguido un ajuste en 144 días (79%) sobre un total de 182. El número medio de vídeos más populares osciló entorno a los 41 (σ : 17,88) sobre un total de vídeos que como término medio fue de 102 (σ : 26,44). El número medio de accesos en la zona de la cabeza estuvo alrededor de los 2.104 (σ : 1.463,72) sobre un valor medio total de 2.288 (σ : 1.558,95), es decir, estaría acaparando casi el 92% de los accesos, aunque este dato debería ser considerado con cautela pues presenta una alta desviación típica, aunque según el test F-Anova podría ser considerado como válido. El grado de ajuste lineal R^2 obtenido es bastante bueno con un valor medio de 0,9733 (σ :

0,0130). Sin embargo, el valor del parámetro θ correspondiente a la zona de la cabeza, como se puede observar en la Fig. 26, presenta un valor medio diferente en lo que respecta el sábado y el domingo frente a los demás días de la semana, pues el test F-Anova ha resultado negativo ($p_valor < 0,05$).

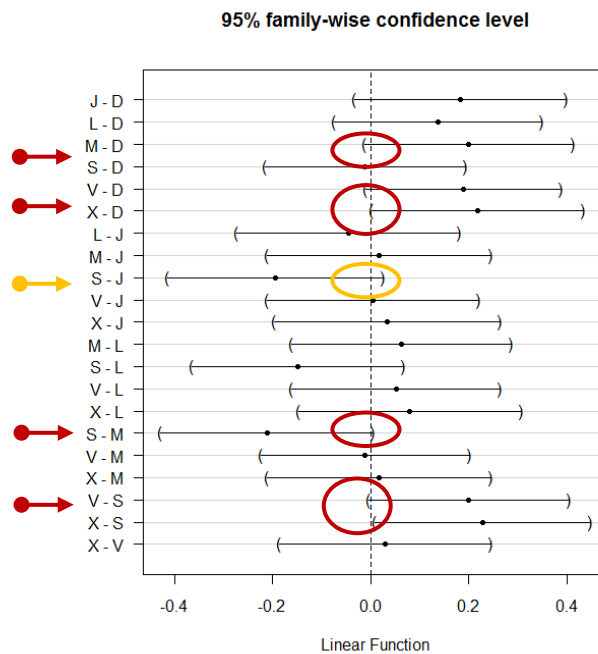


Figura 26. Test F-Anova para θ entre pares de días de la semana.

En la Fig. 27 se puede observar un ejemplo de ajuste para el sábado día 7 de Marzo de 2009, donde el modelo general del sábado, con $(\theta_1, \theta_2) = (1,3022; 0,3121)$, ajusta mejor que el modelo general propiamente del sábado, con $(\theta_1, \theta_2) = (1,2963; 0,4004)$, con un ajuste lineal R^2 de 0,9696 y 0,9448 respectivamente, pero el modelo propio del día es el que mejor ajusta lógicamente con $(\theta_1, \theta_2) = (1,4612; 0)$ y un R^2 de 0,9741. El punto de corte del modelo específico es de 63, en el modelo general de 41 y en el modelo del sábado 39.

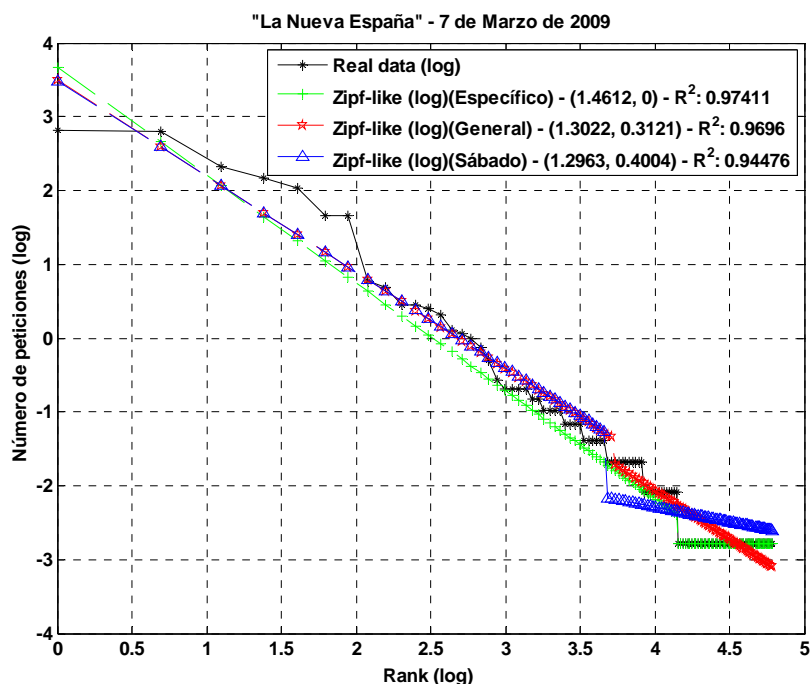


Figura 27. Ajuste con dos Zipf-like, modelo general y específico.

Desde el punto de vista de las secciones, al igual que en casos anteriores la sección S3 es la que más días cubre, con un total de 53 frente a 182 días (29%). Este porcentaje es muy similar al obtenido en los periódicos "La Opinión A Coruña", "Faro de Vigo" y "Levante-EMV". Para el resto de las secciones el número de días cubierto fue inferior o igual a los 36. Además, en "La Nueva España" la sección S10 fue la que más días cubrió con una considerable diferencia si se compara con sus homólogas en el resto de los periódicos, con un total de 101 frente a 182 días (55%), siendo 3 veces superior a "Levante-EMV", 9 veces superior a "La Provincia", 7 veces superior a "Faro de Vigo" y finalmente 101 veces superior a "La Opinión A Coruña".

5.3.6. Periódico "Superdeporte"

El número total de días modelados para la demanda ha sido de 50 sobre un total de 173, representando un porcentaje bastante bajo de un 25%. El número medio de vídeos que forman parte de la cabeza es un total de 4 (σ : 4,26) sobre un total de vídeos de 25 (σ : 8,43). Esta zona recibe un número medio de accesos de 849 (σ : 960) sobre un número medio total de accesos de 696 (σ : 714,75). El grado de ajuste lineal medio alcanzado ha sido de 0,9943 (σ : 0,0087). Finalmente, el valor medio para el parámetro θ ha sido de 3,2918 (σ : 1,38).

En la Fig. 28 se observa un ejemplo de ajuste sobre las peticiones realizadas el martes día 16 de Junio de 2009, donde se aplica el modelo general (punto corte= 6; $\theta_1=1,9115$; $\theta_2 = 0,5260$), el modelo general del martes (punto corte = 7; $\theta_1 = 1,9240$; $\theta_2=0,6089$), así como el específico (punto corte = 9; $\theta_1 = 1,7721$; $\theta_2 = 0,7939$). Como se puede apreciar la diferencia entre ambos modelos generales desde el punto de vista del grado de ajuste lineal es mínima, con un valor alrededor de 0,96.

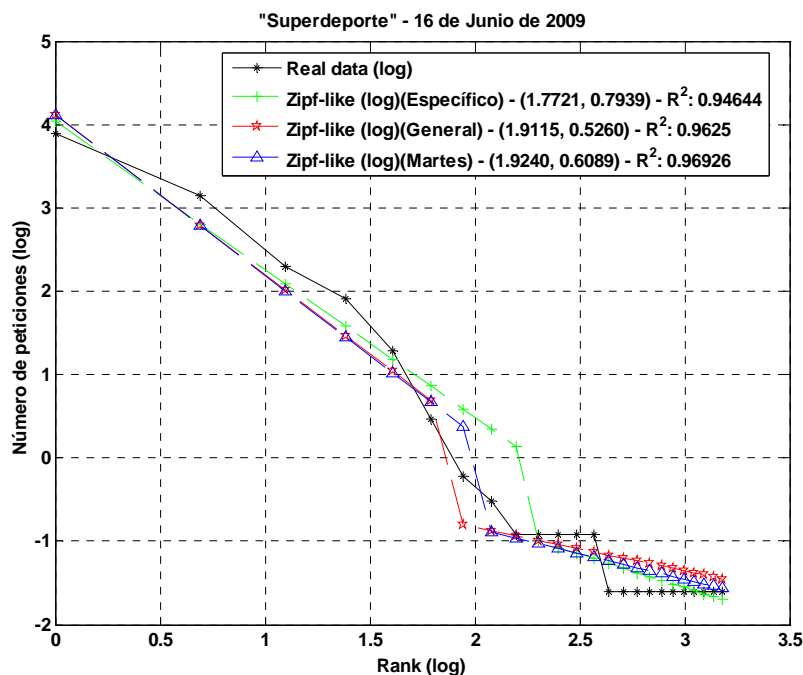


Figura 28. Ajuste con modelo general, modelo general del martes y específico.

5.4. Distribución Mandelbrot

En todos los periódicos fue posible modelar un elevado número de días. El servicio "La Opinión A Coruña" fue el que presentó un mayor grado de cobertura con un 98% (178 días), seguido por "Levante-EMV" y "La Nueva España" con un 96% (174 días aproximadamente), "La Provincia" y "Faro de Vigo" con un 89% (162 días) y un 80 % (145 días) respectivamente, y en último lugar "Superdeporte" con un 73% (133 días).

El grado de ajuste de los datos a esta distribución fue medido a través del test de Kolmogorov-Smirnov, de forma que si el p_valor fuese inferior al nivel de significación

0,05 el modelo obtenido sería rechazado, dado que no representaría un buen ajuste de los datos a modelar.

5.4.1. Periódico "La Opinión A Coruña"

El modelo general tiene un valor medio para el parámetro k de 4,4033 (σ : 11,6168), y un valor medio para θ de 2,7659 (σ : 2,1426). Desde el punto de vista de las secciones, sólo se ha logrado ajustar la sección S3, con un grado de cobertura del 99% (181 días), siendo el valor medio del parámetro k de 3,3404 (σ : 8,4625) y el valor medio del parámetro θ de 2,8896 (σ : 2,2268). Cabe destacar que una gran parte de secciones, como son las secciones S2, S4, S5, S6, S9 y S10 no tuvieron accesos o bien fueron escasos como en el caso de las secciones S1 y S7, donde en caso afirmativo se concentraron sobre tres vídeos como máximo. El resto de las secciones presentaron un grado de cobertura comprendido entre el 4% (9 días) y el 8% (14 días).

5.4.2. Periódico "Faro de Vigo"

El modelo general tiene un valor medio para el parámetro k de 16,6942 (σ : 31,3946), y un valor medio para θ de 4,9714 (σ : 7,0705). Por otro lado, en este periódico las secciones S4, S7 y S9 no se encuentran presentes, y sólo se ha logrado ajustar la sección S3, con un grado de cobertura del 69% (125 días), siendo el valor medio obtenido para el parámetro k de 4,2064 (σ : 11,4768) y el valor medio del parámetro θ de 4,7224 (σ : 9,5971). El resto de las secciones presentaron un grado de cobertura inferior al 36% (66 días).

5.4.3. Periódico "La Provincia"

El modelo general tiene un valor medio para el parámetro k de 2,7468 (σ : 8,28), y un valor medio para θ de 2,2536 (σ : 1,2076). Por otro lado, las secciones S2, S4 y S5 no presentaron accesos, las secciones S1, S7 y S9 presentaron un grado de cobertura inferior al 8% (14 días), y por último las secciones S6 y S10 con un grado de cobertura del 23% (42 días) y del 18% (33 días) respectivamente. Las únicas secciones con

posibilidad de ser modeladas serían la sección S8 con un 51% (93 días) y la sección S3 con un 84% (153 días).

La sección S3 presentaría un valor medio para el parámetro k de 11,2839 (σ : 40,8972), y un valor medio para θ de 4,2002 (σ : 10,0250). Además, la sección S8 presentaría un valor medio para el parámetro k de 5,9579 (σ : 23,7047), y un valor medio para θ de 4,6467 (σ : 13,3083).

5.4.4. Periódico "Levante-EMV"

El planteamiento de un modelo general que caracterizaría la demanda sobre todos los días de la semana, conlleva a la aplicación del test F-Anova con el fin de comprobar si la media de los parámetros podría considerarse la misma. La Fig. 29 muestra el resultado del test para el parámetro k , y como se puede observar, al igual que ocurrió para la distribución Zipf-like, tampoco sería posible el planteamiento de un modelo general que abarcase todos los días, pues tanto el domingo como el sábado tienen un valor medio de k diferente al jueves y al viernes.

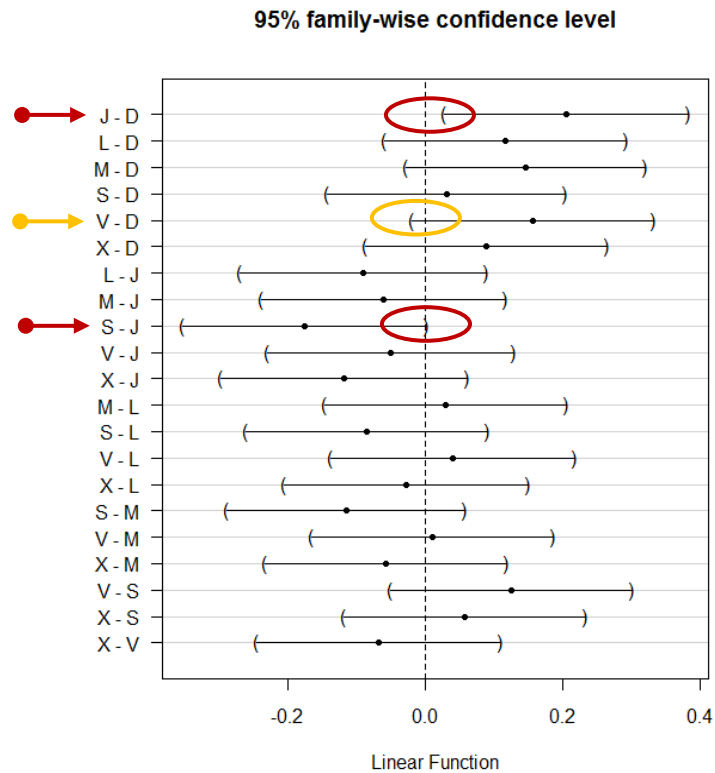


Figura 29. Test F-Anova para k entre pares de días de la semana.

Se plantea dos modelos generales, uno con el fin de caracterizar la demanda de lunes a viernes, y el otro caracterizaría la demanda del sábado y el domingo conjuntamente. Así pues, el primer modelo presentaría un valor medio para el parámetro k de 5,7733 (σ : 5,2116) y un valor para el parámetro θ de 3,0356 (σ : 1,0124), y en el segundo caso el valor medio de k sería de 3,4642 (σ : 2,1455) y un valor medio de θ de 2,5341 (σ : 0,5207).

Desde el punto de vista de las secciones cabría plantear un modelo para las secciones S3, S8 y S10, las cuales presentarían un grado de cobertura del 79% (144 días), 66% (120 días) y 75% (136 días) respectivamente, el resto de las secciones presentan un grado de cobertura igual o inferior al 46% (≤ 83 días). Así pues, tras el análisis de los datos, los modelos que caracterizarían la sección S3, S8 y S10 tendrían un valor medio para el parámetro k de 3,0149 (σ : 9,9744), 5,1916 (σ : 15,1856), y 8,2895 (σ : 18,8347) respectivamente, y en relación al valor medio del parámetro θ presentaría un valor de 3,9416 (σ : 5,2044), 5,6789 (σ : 12,3387) y 2,9902 (σ : 3,0131) respectivamente.

5.4.5. Periódico "La Nueva España"

La caracterización de la demanda de este periódico a través de la distribución Mandelbrot infiere el planteamiento de dos modelos según el test F-Anova, dado que tras la comparación del valor medio para el parámetro k entre los distintos días de la semana ha dado un p_valor inferior 0,05. Este hecho es representado a través de la Fig. 30 que refleja el resultado del test F-Anova para el parámetro k , y donde se observa que la media del valor para dicho parámetro en el sábado estaría en la frontera con el miércoles, y rozando la permisibilidad con el martes y el jueves. Sin embargo, el test F-Anova para el parámetro θ fue positivo para todos los pares de días comparados entre sí.

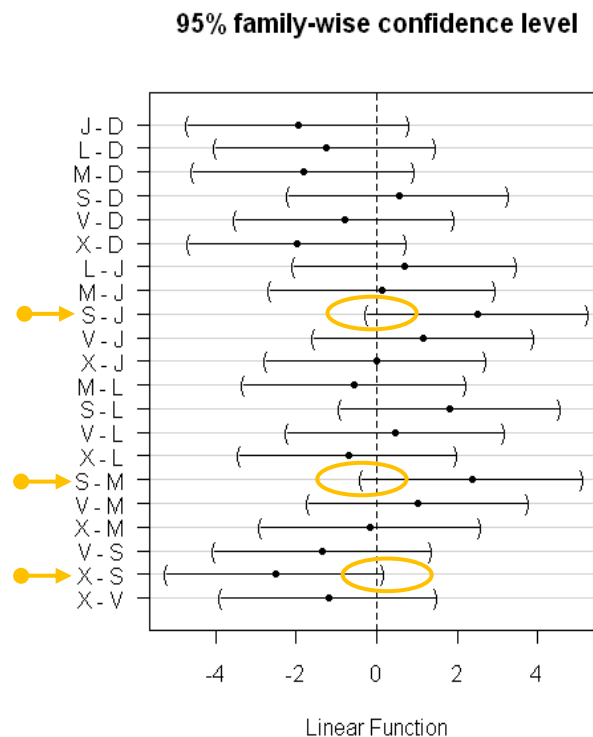


Figura 30. Test F-Anova para k entre pares de días de la semana.

En definitiva, a la vista de los resultados se plantearía un modelo que caracterizaría la demanda del lunes al viernes y la del domingo conjuntamente, donde el valor medio del parámetro k sería de 4,7341 (σ : 3,1228) y el valor medio para el parámetro θ de 2,3610 (σ : 0,2986). Por otro lado, el modelo para el sábado tendría un valor medio para el parámetro k de 6,5764 (σ : 3,9299) y un valor medio para el parámetro θ de 2,4603 (σ : 0,3733).

Desde el punto de vista de las secciones con la distribución Mandelbrot se podría establecer un modelo para las secciones S1, S3, S4, S5, S6 y S10 con un grado de cobertura de 75% (137 días), 91% (166 días), 63% (115 días), 44% (81 días), 57% (104 días) y del 97% (176 días) respectivamente, cuyos resultados han sido los siguientes:

- 1) En la sección S1 el test F-Anova resultó ser negativo tanto para θ como para k , y donde la demanda media del sábado y del domingo presentaba un comportamiento distinto respecto al resto de los días de la semana. Así pues, se plantea un modelo que abarcaría del lunes al viernes donde θ tendría un valor de 5,2550 (σ : 5,3831) y donde k tomaría un valor de 5,1647 (σ : 9,9798). Por otro lado, se plantearía un modelo conjunto para el sábado y el domingo, donde el

valor de θ sería de 12,8247 (σ : 19,0326) y el valor de k tendría el valor de 18,9944 (σ : 31,4975).

- 2) En la sección S3 el test F-Anova fue positivo para ambos parámetros y por lo tanto se plantearía un modelo único que englobaría todos los días, donde θ tomaría el valor de 3,0599 (σ : 3,8271) y k sería de 9,0965 (σ : 30,6962).
- 3) En la sección S4 dado que el test F-Anova resultó ser negativo para todos los días como un conjunto, dado que la demanda media diaria del sábado y del domingo presentaba un comportamiento distinto frente al resto, incluso el sábado fue diferente del domingo, por lo tanto la caracterización se establece en base a los siguientes tres modelos:
 - ✓ Lunes a viernes: el valor de θ sería de 9,4545 (σ : 10,497), con un valor para k de 8,717 (σ : 14,1398).
 - ✓ Sábado: con un valor para θ de 21,0956 (σ : 23,7783) y un valor para k de 18,2762 (σ : 25,4784).
 - ✓ Domingo: el valor de θ sería de 6,4509 (σ : 7,5525) y un valor para k de 3,9384 (σ : 9,1313).
- 4) En la sección S5 el test F-Anova resultó ser negativo para el parámetro θ , y positivo para el parámetro k . De acuerdo a la Fig. 31, el comportamiento de la demanda media diaria del domingo es distinta a la del lunes, miércoles, jueves y viernes. Por lo tanto se plantean dos modelos, uno que caracterizaría la demanda media diaria que englobaría de lunes a sábado y otro modelo que caracterizaría la demanda del domingo.

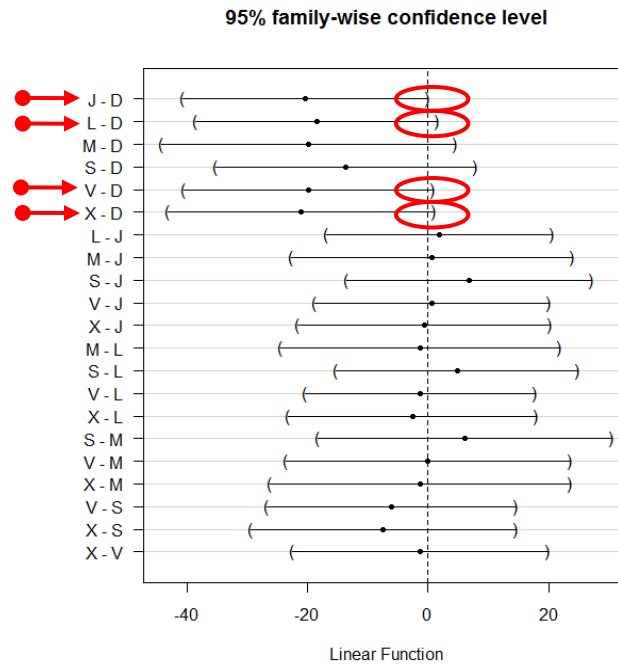


Figura 31. Test F-Anova para θ , entre pares de días de la semana.

Para el grupo de días de lunes a sábado el valor de θ sería de 5,6593 (σ : 8,6088) y un valor para el parámetro k de 4,8424 (σ : 10,5325). Respecto al domingo el valor del parámetro θ propuesto sería de 25,5612 (σ : 39,8096), y con un valor para el parámetro k de 17,2864 (σ : 32,2982).

- 5) En la sección S6 el test F-Anova resultó ser positivo para ambos parámetros, dando pie a una caracterización de la demanda media que englobaría todos los días. El valor de θ sería de 5,7251 (σ : 8,0951) y un valor de k de 4,6132 (σ : 14,4048).
- 6) Al igual que en las secciones S3 y S6, en la sección S10 también se podría plantear un modelo único que englobaría todas los días. El valor de θ sería de 3,0848 (σ : 3,7818) y el valor para k de 14,9026 (σ : 56,7886).

5.4.6. Periódico "Superdeporte"

El resultado del test F-Anova resultó ser positivo tanto para el parámetro θ como el parámetro k , y por lo tanto la demanda media diaria se caracterizará con un modelo cuyo valor para el parámetro θ sería de 2,3028 (σ : 1,2625) y un valor para el parámetro k de 2,0357 (σ : 4,9789).

5.5. Distribución Stretched

5.5.1. Periódico "La Opinión A Coruña"

Desde el punto de vista del periódico en su totalidad, de Enero a Junio, y por clases cabe plantear un modelo único, que viene dado a través de la ecuación (14).

$$y^{0,0622 (\sigma: 0,0966)} = -0,1338 (\sigma: 0,2392) * \log(x) + 1,3434 (\sigma: 0,6152) \quad (14)$$

Cuando el periódico es modelado desde el punto de vista de las secciones, sólo la sección S3 tiene posibilidades con un grado de cobertura del 59% (107 días). Sin embargo, no ha sido posible plantear un modelo único para las siete clases, pues los parámetros a, b y c incumplen el test de la media, F-Anova ($p_valor < 0,05$). En la Fig. 32 aparece representado el resultado del test F-Anova para el parámetro a y se puede observar que la clase del sábado no sigue la misma media que la clase del lunes, martes, miércoles, jueves (señalados mediante un círculo rojo), y estaría rozando el límite de la permisibilidad para las clases del viernes y del domingo (señalados mediante un círculo amarillo).

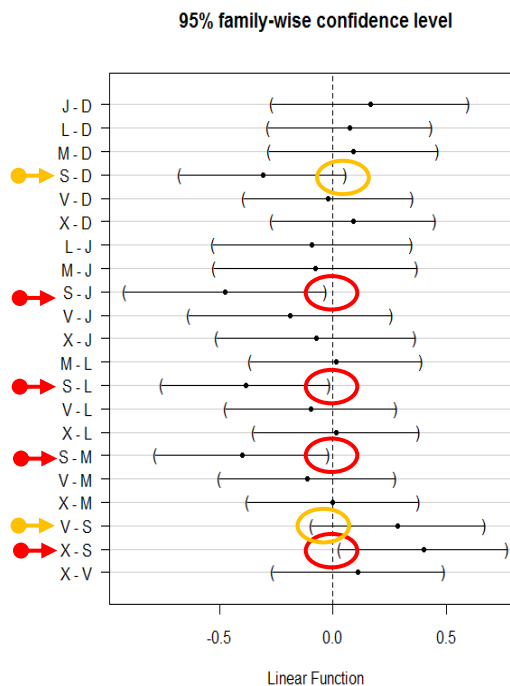


Figura 32. Test F-Anova para el parámetro a.

Por lo tanto, se plantearían dos modelos, uno que determinaría el patrón de la demanda para el sábado, y el otro modelaría el resto de las clases conjuntamente. El

modelo de comportamiento para todas las clases menos el sábado viene expresado mediante la ecuación (15).

$$y^{0,109 (\sigma: 0,1539)} = -0,5001 (\sigma: 1,0776) * x + 1,9677 (\sigma: 1,9006) \quad (15)$$

Mientras que el modelo para el sábado vendría dado a través de la ecuación (16).

$$y^{0,167 (\sigma: 0,1719)} = -0,5216 (\sigma: 0,6974) * x + 2,2287 (\sigma: 1,6346) \quad (16)$$

5.5.2. Periódico "Faro de Vigo"

El conjunto de valores de cada uno de los parámetros que caracterizan la distribución no siguen una distribución normal según el test de Shapiro-Wilk, tampoco el test de Levene para la comprobación de la varianza entre las siete clases, así como también incumple el test de F-Anova para la media, por lo tanto en los tres casos el p_valor ha sido inferior a 0,05. En la Fig. 33 se puede apreciar que la media de la demanda del viernes es diferente de la media de la demanda del lunes (indicado mediante un círculo rojo), y aunque el miércoles, jueves y sábado se puede considerar que la media de la demanda es la misma, estarían en el límite de no ser así (señalados mediante un círculo amarillo).

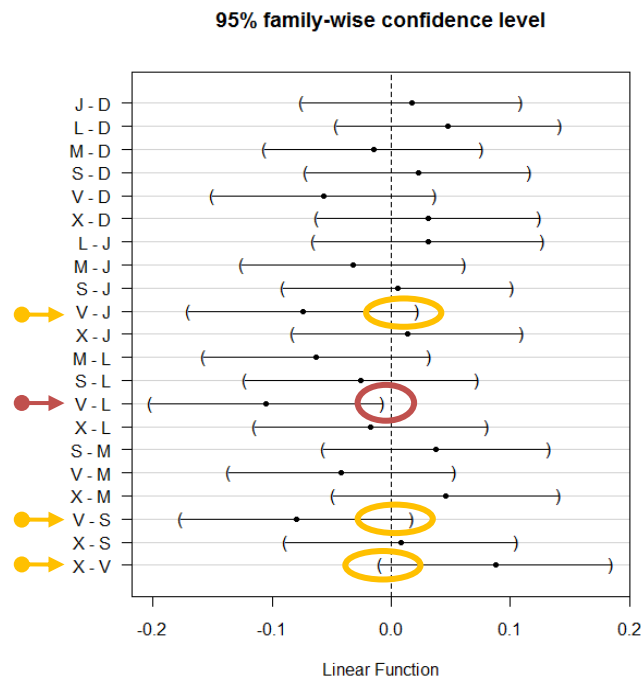


Figura 33. Resultados del F-Anova test.

Así pues, a la vista de los resultados del test F-Anova se plantearía un modelo de popularidad que englobaría todas las clases excepto la clase del viernes, y otro modelo independiente para el modelado de la demanda de esta última. El primer modelo sería el expresado en la ecuación (17).

$$y^{0,0188 (\sigma: 0,0356)} = -0,0354 (\sigma: 0,0721) * x + 1,0862 (\sigma: 0,1725) \quad (17)$$

Mientras que el modelo para la clase del viernes viene expresado a través de la ecuación (18).

$$y^{0,0558 (\sigma: 0,0707)} = -0,1094 (\sigma: 0,1487) * x + 1,2716 (\sigma: 0,3794) \quad (18)$$

Por otro lado, desde el punto de vista de las secciones no ha sido posible plantear modelo alguno, dado que el grado de cobertura máximo ha sido de un 10% (19 días) por parte de la sección S3, siendo menor para el resto.

5.5.3. Periódico "La Provincia"

La demanda media diaria ha podido ser caracterizada a través de un único modelo dado que los resultados del test F-Anova ha resultado ser positivo en los tres parámetros a, b y c. Así pues, el modelo planteado estaría expresado con la ecuación (19) presentada a continuación.

$$y^{0,0207 (\sigma: 0,055)} = -0,0385 (\sigma: 0,1043) * x + 1,1012 (\sigma: 0,2795) \quad (19)$$

Desde el punto de vista de las secciones cabe plantear un modelo para la sección S3, donde el grado de cobertura es del 38% (69 días), y cuya expresión viene representada a través de la ecuación (20). Para el resto de las secciones el grado de cobertura fue igual o inferior al 5% (10 días).

$$y^{0,1109 (\sigma: 0,1495)} = -0,2109 (\sigma: 0,2849) * x + 1,5866 (\sigma: 0,8159) \quad (20)$$

5.5.4. Periódico "Levante-EMV"

El resultado del test de Shapiro-Wilk resultó ser negativo para todos los parámetros a, b y c ($p_valor < 0,05$), al igual que el test de la comparación de la media entre clases F-Anova. Además, el test de la varianza de Levene fue positivo en los parámetros a y b con un p_valor en cada uno de ellos de 0,1161 y 0,09282 respectivamente, pero negativo en el parámetro c ($p_valor < 0,05$). Como se puede observar en la Fig. 34, la demanda media de los domingos presenta un comportamiento distinto al del miércoles, y un poco en la frontera respecto al martes y el jueves.

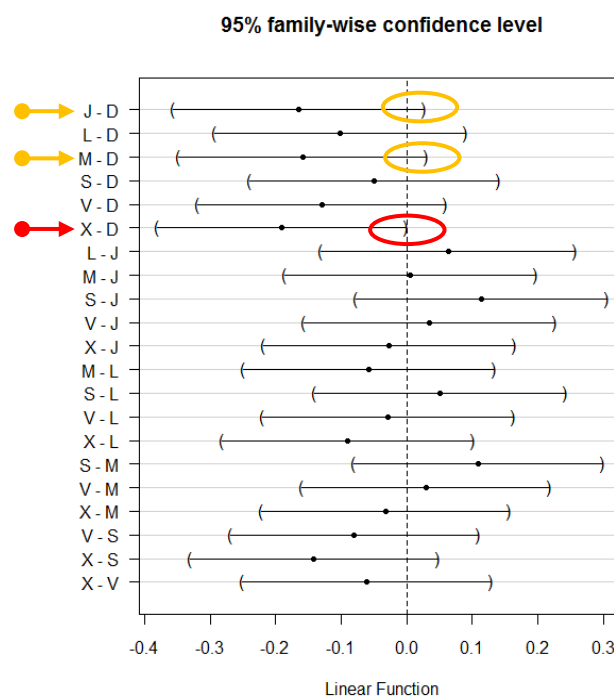


Figura 34. Test F-Anova del parámetro a.

Así pues, a la vista de los resultados del test F-Anova se plantearía un modelo de popularidad que englobaría todas las clases excepto la clase del domingo, y otro modelo independiente para el modelado de la demanda de esta última. El primer modelo sería el expresado en la ecuación (21).

$$y^{0,0982 (\sigma: 0,1041)} = -0,1954 (\sigma: 0,2359) * x + 1,5127 (\sigma: 0,6492) \quad (21)$$

Mientras que el modelo para la clase del domingo viene expresado a través de la ecuación (22).

$$y^{0,0346 (\sigma: 0,0489)} = -0,0631 (\sigma: 0,0914) * x + 1,1569 (\sigma: 0,2272) \quad (22)$$

Respecto a las secciones, el grado de cobertura es igual o inferior al 19% (34 días) en cada una de ellas, por lo tanto no se justifica el planteamiento de ningún modelo.

5.5.5. Periódico "La Nueva España"

Como se puede apreciar en la Fig. 35 el sábado parece presentar un comportamiento medio de la demanda rozando la permisibilidad con las clases del lunes y martes, y en estado crítico para las clases del miércoles y jueves.

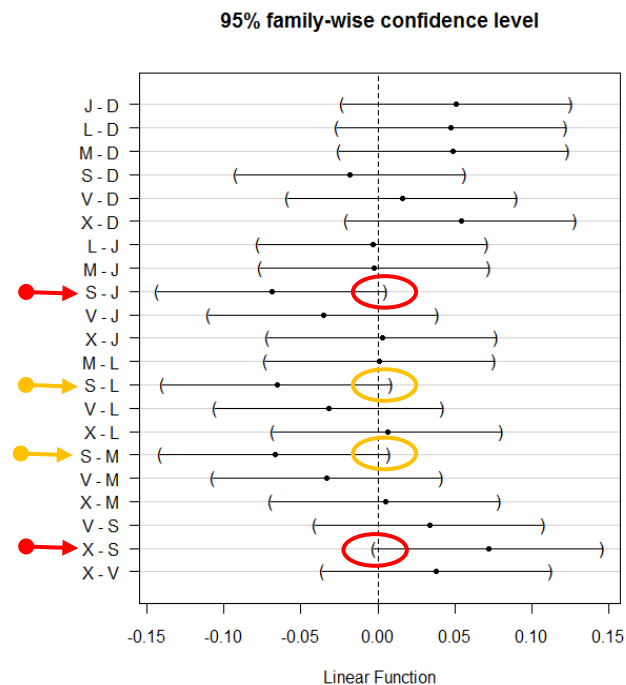


Figura 35. Resultado de F-Anova para el parámetro a.

Dados los resultados del test F-Anova se plantearía un modelo de popularidad que englobaría todas las clases excepto la clase del sábado, y otro modelo independiente para el modelado de la demanda de esta última. El primer modelo sería el expresado en la ecuación (23).

$$y^{0,039 (\sigma: 0,0546)} = -0,0583 (\sigma: 0,0811) * x + 1,1604 (\sigma: 0,2327) \quad (23)$$

Mientras que el modelo para la clase del sábado viene expresado a través de la ecuación (24).

$$y^{0,0782 (\sigma: 0,0779)} = -0,1124 (\sigma: 0,1112) * x + 1,3188 (\sigma: 0,3265) \quad (24)$$

Desde el punto de vista de las secciones sólo la sección S3, así como la última sección S10, da pie al establecimiento de un modelo, dado que es en ellas donde se consigue alcanzar un porcentaje de modelado del 79% y el 91% respectivamente. En el resto de las secciones se consigue un grado de cobertura igual o inferior al 18%.

El modelo correspondiente a la sección S3 viene representado a través de la siguiente ecuación (25).

$$y^{0,0663 (\sigma: 0,1209)} = -0,1231 (\sigma: 0,2614) * x + 0,0662 (\sigma: 0,1209) \quad (25)$$

Y el modelo correspondiente a la sección S10 sería expresado en la ecuación (26).

$$y^{0,0681 (\sigma: 0,1084)} = -0,0998 (\sigma: 0,1763) * x + 1,3048 (\sigma: 0,5444) \quad (26)$$

5.5.6. Periódico "Superdeporte"

La caracterización de la demanda media diaria en este periódico digital solo se consiguió en el 37% de los casos, es decir, sólo en 67 de los 182 días fue posible la realización de un ajuste de los datos a esta distribución. Al igual que en los periódicos digitales anteriores se podría establecer un modelo general de comportamiento, cuya expresión sería la expresada en (27).

$$y^{0,0078 (\sigma: 0,0211)} = -0,0179 (\sigma: 0,051) * x + 1,0382 (\sigma: 0,0107) \quad (27)$$

5.6. Transformación de Box-Cox

El método de Box-Cox es una transformación sobre la variable dependiente. Su diseño está pensado para valores estrictamente positivos y elige la transformación que mejor se ajuste a los datos. El método transforma la variable dependiente y en $y(\lambda)$, donde la familia de transformaciones indexada por λ es la indicada en la ecuación (28).

$$t_{\lambda}(y) = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \text{si } \lambda \neq 0 \\ \log y & \text{si } \lambda = 0 \end{cases} \quad (28)$$

El valor de λ es elegido por el método de máxima verosimilitud, de forma que asumiendo la normalidad de los errores su valor vendrá dado a través de la siguiente ecuación (29).

$$L(\lambda) = -\frac{n}{2} \log\left(\frac{RSS}{n}\right) + (\lambda - 1) \sum_{i=1}^n y_i \quad (29)$$

Se podría calcular λ' para que maximice exactamente $L(\lambda)$, pero normalmente $L(\lambda)$ se maximiza sobre un conjunto de valores tales como $\{-2, -1, -1/2, 0, 1/2, 1, 2\}$, con el fin de asegurar que λ tenga una interpretación más sencilla.

La transformación de la variable dependiente puede llevar a un modelo difícil de interpretar, por lo tanto, no se aconseja realizar un cambio de variable salvo que sea estrictamente necesario. Por este motivo, se construirá un intervalo de confianza para λ de $100(1-\alpha)\%$, cuyo valor es calculado a través de la siguiente expresión (30).

$$\left\{ \lambda : L(\lambda) > L(\lambda') - \frac{1}{2} \chi_1^2 (1 - \alpha) \right\} \quad (30)$$

De forma que, si $L(\lambda)$ estuviese comprendido dentro del intervalo no sería necesario realizar una transformación del modelo; por el contrario si $L(\lambda) \neq 1$ entonces se debería elegir aquel valor comprendido dentro del intervalo que minimizase la dificultad de interpretación del modelo, por ejemplo si $1/2$ y $1/3$ estuviesen en el intervalo la mejor opción sería elegir un $1/2$.

Principalmente, las ventajas ofrecidas por esta transformación son dos [69]. En primer lugar, la distribución estadística a la cual podría ajustarse los datos es desconocida, y por lo tanto sería deseable disponer de un método que evitase el problema de la elección de un modelo específico. En segundo lugar, esta técnica puede ofrecer una predicción a través de un modelo lineal, según la expresión (31), de la demanda media futura basándose en el conocimiento de como fueron las demandas medias diarias en el pasado.

$$t_{\lambda}(y) = -a * \log(x) + b \quad \text{con } 1 \leq x \leq n \quad (31)$$

En los siguientes apartados se comenta los resultados tras su aplicación sobre los diferentes periódicos, siendo en el Anexo V donde se presenta información más detallada sobre el valor obtenido de algunos tests estadísticos, así como otros valores relativos a los parámetros, y por lo tanto en general no serán comentados en los apartados correspondientes.

5.6.1. Periódico "La Opinión A Coruña"

Tras el análisis de los diferentes días, la transformación Box-Cox ha resultado ser válida sobre 106 días (58%) sobre un total de 182 días. Los modelos obtenidos se presentan en la Tabla 14. El grado de ajuste lineal R^2 fue elevado en todos los casos con una media que osciló entre 0,94 y 0,96. El test F-Anova resultó ser positivo para el periodo en su totalidad en los tres parámetros a, b y λ , y por lo tanto se podría plantear un modelo general de comportamiento.

	Modelos
General	$t_{-0,1}(y) = -1,5154 (\sigma: 0,0669) * \log(x) + 3.5112 (\sigma: 0,1673)$
Lunes	$t_{-0,1}(y) = -1,5464 (\sigma: 0,0585) * \log(x) + 3.7051 (\sigma: 0,1499)$
Martes	$t_{-0,1}(y) = -1,5877 (\sigma: 0,0702) * \log(x) + 3.3788 (\sigma: 0,1710)$
Miércoles	$t_{-0,1}(y) = -1,4470 (\sigma: 0,0685) * \log(x) + 3.2408 (\sigma: 0,1719)$
Jueves	$t_{-0,1}(y) = -1,4454 (\sigma: 0,0655) * \log(x) + 3.1702 (\sigma: 0,1687)$
Viernes	$t_{-0,1}(y) = -1,5419 (\sigma: 0,0767) * \log(x) + 3.6556 (\sigma: 0,1866)$
Sábado	$t_{-0,1}(y) = -1,6767 (\sigma: 0,075) * \log(x) + 3.9612 (\sigma: 0,1806)$
Domingo	$t_{-0,1}(y) = -1,29 (\sigma: 0,0541) * \log(x) + 3.2344 (\sigma: 0,1354)$

Tabla 14. Modelos de la transformación de Box-Cox.

En la Fig. 36 se puede ver un ejemplo de aplicación de la transformación de Box-Cox en un día concreto, como es el 21 de Junio, tratándose en este caso de un domingo, donde fueron accedidos un total de 28 vídeos con un total de 201 peticiones. Los modelos aplicados han sido el modelo general y el modelo general que

representaría el comportamiento de la demanda de un domingo, ambos especificados en la Tabla 14. Además, a esta figura se ha incorporado el modelo que se ha obtenido para el propio día ($a = -1,2509$, $b = 3,1818$, $\lambda = -0,1$), y que se ha denotado como modelo específico. Como se puede observar, el resultado parece lógico, pues el modelo específico es el que mejor ajusta ($R^2 = 0,9768$), seguido del modelo del domingo ($R^2 = 0,9766$), y por último el modelo general ($R^2 = 0,8655$).

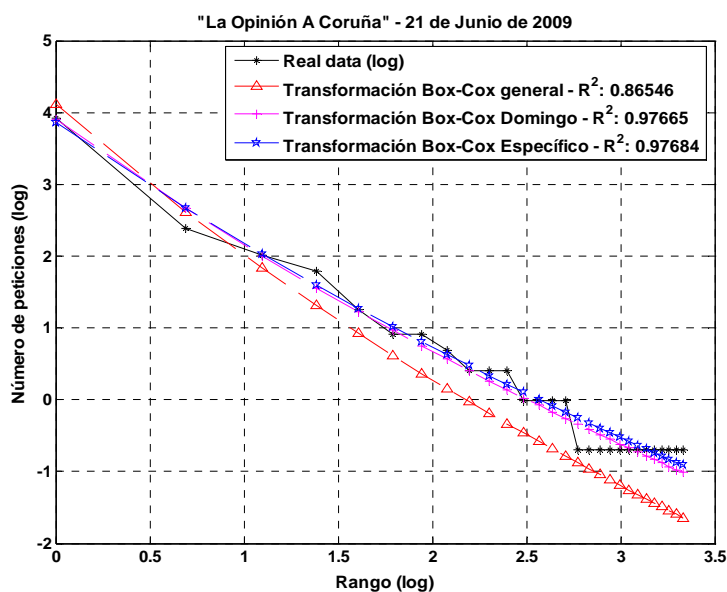


Figura 36. Transformación de Box-Cox con tres modelos.

Desde el punto de vista de las secciones, sólo la sección S3 podría presentar un posible modelado, con un total de 125 días (69%) sobre un total de 182 días, para el resto de las secciones el número de días modelados fue inferior a los 61 días ($\leq 33\%$). Los datos obtenidos para esta sección se encuentran especificados en el Anexo V. En la Fig. 37 se encuentra un ejemplo de aplicación de la transformación de Box-Cox, para modelar los accesos realizados el día 21 de Junio de 2009 en la sección S3, con el fin de hacer una comparación entre el modelado de los accesos totales y los realizados sobre dicha sección. Así pues, en la Fig. 37 junto con los accesos, se representa el modelo general, el modelo obtenido para el domingo, y finalmente el modelo específico obtenido para dicho día ($a = -1,3779$, $b = 3,2583$, $\lambda = -0,1$), donde lógicamente los cuatro conjuntos de datos anteriores estarían centrados en la sección S3. Al igual que ocurrió en el caso anterior el modelo específico es el que aporta un mayor grado de ajuste lineal ($R^2 = 0,9719$), en segundo lugar el modelo del domingo ($R^2 = 0,9335$), y por último el modelo general ($R^2 = 0,8980$). Pero a la vista de los resultados, todos los

ajustes lineales han bajado su valor. Una posible causa podría ser debida a un descenso en el número de accesos, lo cual al disponer de un menor número de datos se traduce en un empeoramiento de los ajustes. De hecho, el número de accesos totales fue de 201 sobre 28 vídeos, y los realizados en la sección S3 fueron 173 sobre 19 vídeos, es decir, un 14% menos.

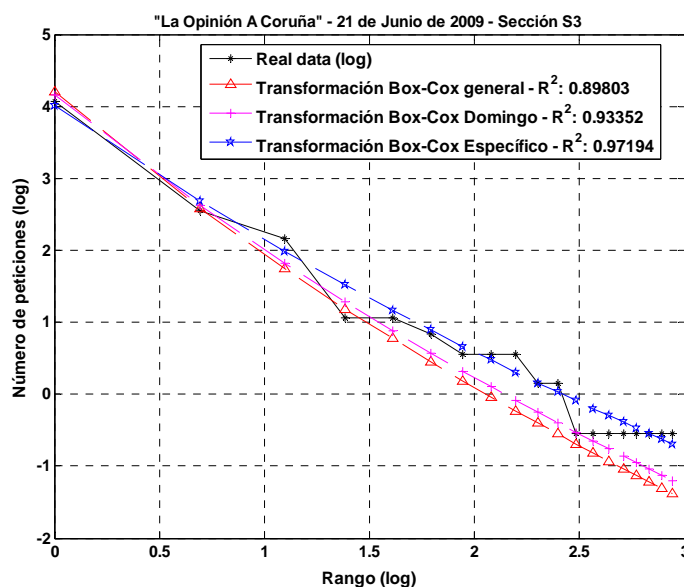


Figura 37. Transformación de Box-Cox con tres modelos en la sección S3.

5.6.2. Periódico "Faro de Vigo"

El número de días modelados con éxito ha sido bastante elevado con un total de 165 días (91%). Igual que en casos anteriores, se ha tratado de buscar un modelo general para el conjunto de parámetros, pero al contrario de lo que ha ocurrido en el periódico "La Opinión A Coruña", el test F-Anova ha resultado ser negativo en el parámetro "a", en los pares correspondientes al (viernes, domingo) y (viernes, sábado), tal y como se aprecia en la Fig. 38, por el contrario en los parámetros "b" y " λ " ha sido positivo. Esto origina el planteamiento de dos modelos generales, uno que caracterizaría la demanda del lunes al viernes, y otro modelo para los días sábado y domingo conjuntamente.

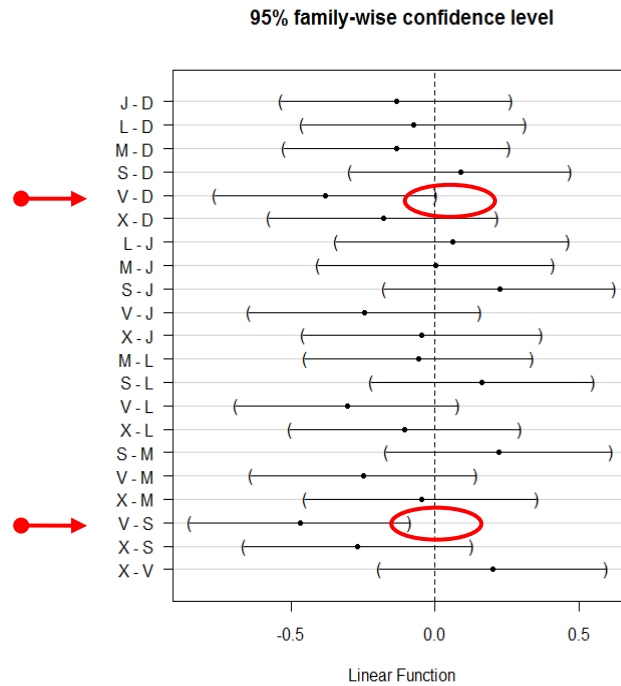


Figura 38. Test F-Anova para el parámetro "a" entre pares de días de la semana.

Un ejemplo de modelado se presenta en la Fig. 39 donde se caracteriza la demanda sobre los vídeos accedidos el martes día 6 de Enero de 2009, con un total de 81 peticiones sobre 16 vídeos. Al igual que las gráficas presentadas en el apartado 5.6.1, se presenta los accesos sobre dicho día, el modelo general que describiría el comportamiento del lunes al viernes, el modelo general del martes y el modelo concreto del día ($a = -0.7241$, $b = 2.2448$, $\lambda = -0,3$). El modelo específico presentó el mejor ajuste lineal ($R^2 = 0,9264$), seguido del modelo del martes ($R^2 = 0,8759$), y en último lugar el modelo general ($R^2 = 0,8554$). En los dos últimos casos el ajuste resultante no fue muy bueno.

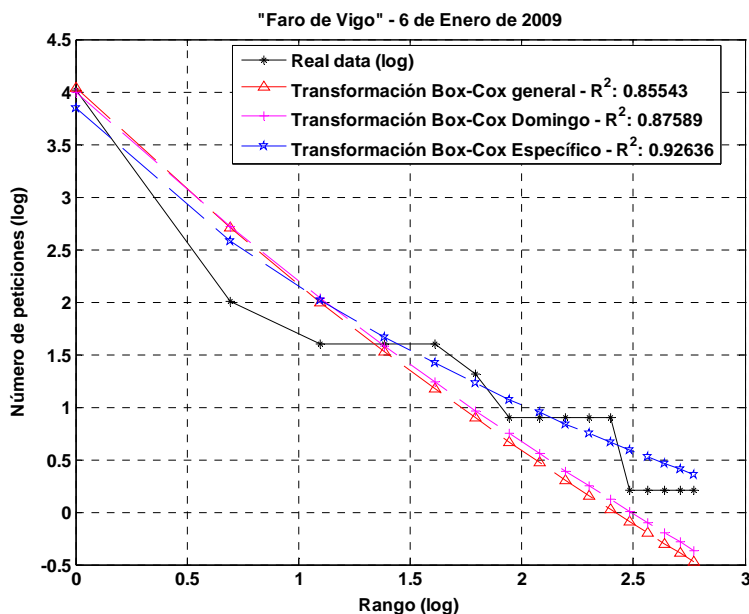


Figura 39. Transformación de Box-Cox con tres modelos.

Las secciones que han presentado un mayor número de días modelados con éxito han sido S3, S6, S8 y S10, con un total de 177, 112, 122 y 153 días respectivamente, el resto de las secciones presentan un número inferior a los 25 días. En todas las secciones el test F-Anova resultó ser positivo para todos los parámetros del modelo, por lo tanto se podría establecer un modelo general para cada una de las secciones, sin embargo la desviación típica resulta ser muy elevada para cada uno de los parámetros, y por lo tanto los modelos no se pueden establecer como válidos ni apropiados. Una de las causas de este suceso podría ser debido al escaso número de vídeos que son accedidos en cada una de las secciones, siendo en la mayoría de los casos inferior a 10, salvo casos puntuales.

5.6.3. Periódico "La Provincia"

El número de días modelados han sido 127 sobre un total de 182 días, lo cual supone un 73% de cobertura. A la hora de buscar un patrón de comportamiento de la demanda para el periodo en su totalidad, se aplicó el test F-Anova, al igual que en casos anteriores, donde para este periódico resultó ser positivo en todos los parámetros del modelo. En la Fig. 40 se puede observar otro ejemplo de ajuste sobre la demanda realizada el domingo 1 de Febrero de 2009 con un total de 1.826 peticiones sobre 32 vídeos, donde el modelo específico del día ($a = -2,183$, $b = 3,7378$, $\lambda = -0,1$) ha

conseguido un grado de ajuste lineal R^2 de 0,9594, y el modelo general del periodo completo y el general para los domingos, cuyos valores están especificados en el Anexo V, han obtenido un grado de ajuste lineal bastante bajo de 0,7827 y 0,8511 respectivamente, como se puede apreciar.

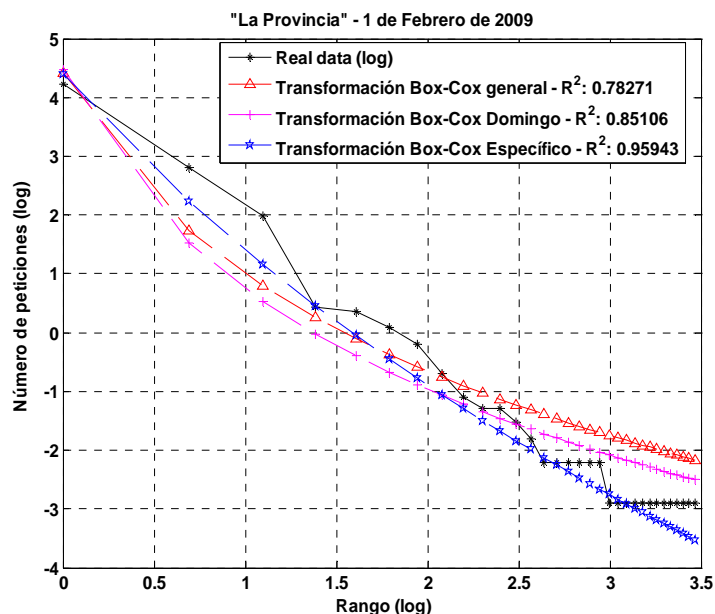


Figura 40. Transformación de Box-Cox con tres modelos.

En general, el ajuste lineal (R^2) global medio se presenta bajo, con un valor que está en torno a 0,92, el cual es inferior respecto a los periódicos estudiados en los apartados 5.6.1. y 5.6.2., los cuales oscilan alrededor de 0,94, pues en estos estudios una diferencia de una centésima es muy apreciable.

Tanto la sección S3 como S8 presentan un grado de cobertura bastante amplio, con un total de 149 días (86%) y 160 días (88%), en el resto de secciones el grado de cobertura fue igual o inferior al 50%. Sin embargo, aunque los parámetros, en ambas secciones, pasaron el test de F-Anova, su desviación típica es alta sobre todo en la sección S8, hecho que también ocurrió en el periódico "Faro de Vigo". En el caso de la sección S8 podría ser debido al bajo número de vídeos que son accedidos, que origina una concentración muy alta sobre el primero de los vídeos y muy baja en el resto, mientras que si el número de vídeos accedidos es alta, aunque siga existiendo más concentración de accesos sobre los primeros vídeos, en general el número de accesos se dispersa más sobre el conjunto de vídeos. En la Fig. 41 se puede observar una comparación entre el número medio de vídeos que son accedidos cada día de la semana,

donde existe una notable diferencia entre la sección S3 y la sección S8. La sección S3 presenta un mayor número de vídeos y una desviación típica menor en sus parámetros (Anexo V), y en la sección S8 ocurre el efecto contrario.

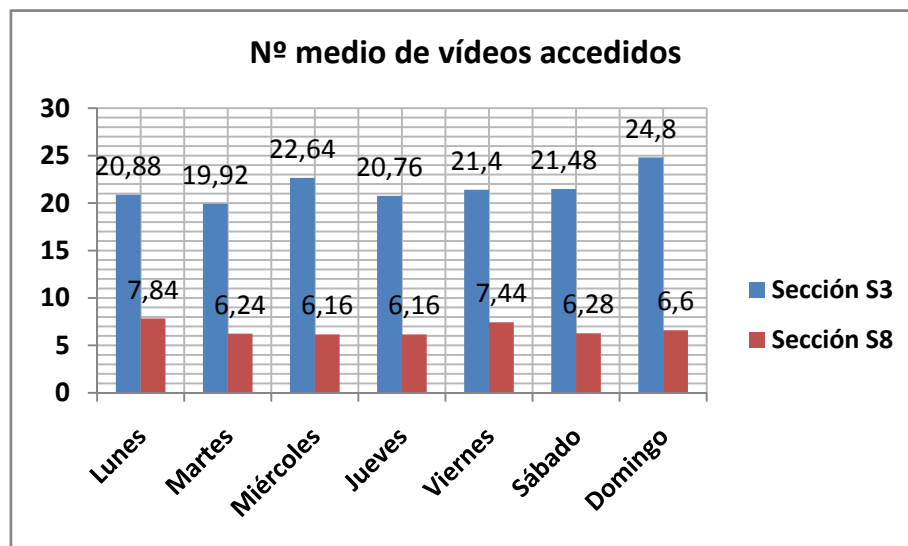


Figura 41. Comparación del número medio de vídeos accedidos entre dos secciones.

5.6.4. Periódico "Levante-EMV"

El grado de cobertura que se ha obtenido en el modelado de los accesos sobre este periódico es muy bajo, siendo el peor de todos los periódicos, donde tan sólo 39 días (22%) sobre un total de 175 días han pasado con éxito el test de contraste de hipótesis. Por lo tanto, el número de días modelados para cada día de la semana es muy bajo oscilando entre 4 y 8. Sin embargo, el grado medio de ajuste lineal R^2 ha sido muy elevado, siendo el mejor de todos los periódicos, con un valor en torno al 0,97 para todos los días de la semana, excepto para el sábado y domingo cuyo valor fue de 0,96. Por otro lado, aunque el test F-Anova resultó ser positivo para todos los parámetros y en teoría podría establecerse un modelo general único, el jueves presentaba un valor muy dispar respecto a los restantes, junto con una desviación típica muy elevada en cada uno de ellos, provocando una distorsión en la media general, por este motivo se ha decidido plantear un modelo general donde sus valores se obtienen teniendo en cuenta todos los días de la semana excepto el jueves. La corroboración de este hecho puede observarse en la Fig. 42, donde se representa el modelado sobre la demanda realizada el miércoles día 11 de Febrero de 2009 sobre un total de 38 vídeos y 1.220 peticiones. Si el modelo

que se aplicase fuese el global, no se consigue ningún ajuste, con un R^2 de 0,5553, sin embargo si se aplica el modelo específico del día ($a = -2,111$, $b = 5,3704$, $\lambda = 0,2$) se obtiene un R^2 de 0,97, o bien si se aplica el modelo general obtenido sin tener en cuenta el jueves y el modelo general que se aplicaría para los miércoles, el ajuste R^2 es de 0,93.

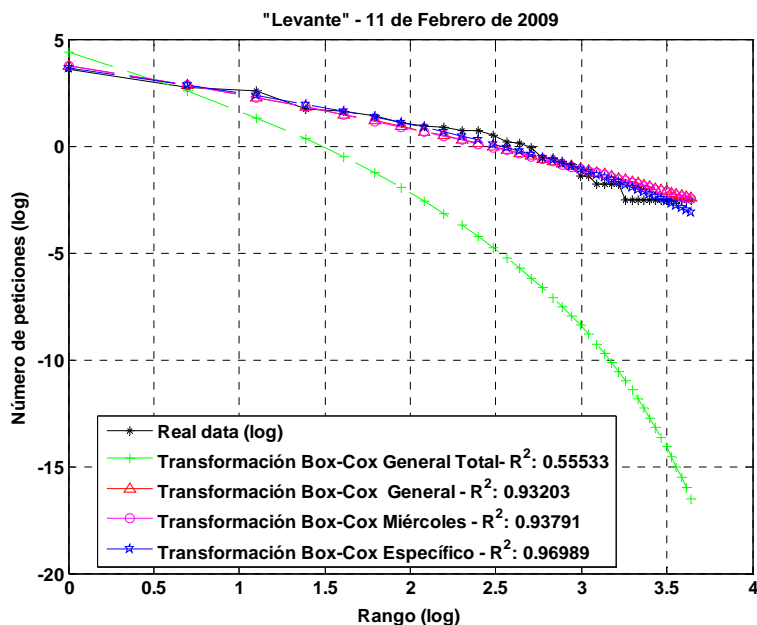


Figura 42. Transformación de Box-Cox con cuatro modelos.

Desde el punto de vista de las secciones, las secciones más destacadas son la sección S1 con 136 días, la sección S3 con 171 días, la sección S4 con 154 días, la sección S6 con 121 días, la sección S8 con 160 días, y por último la sección S10 con 172 días. En el resto de las secciones el número de días osciló entre los 47 días y los 83 días. Sin embargo, en las secciones que podrían parecer significativas, tal y como se puede observar en el Anexo V, la desviación típica de los parámetros fue elevada, e incluso en algunos casos demasiado elevada. Por lo tanto, al igual que ocurrió en los periódicos "Faro de Vigo" y "La Provincia", también parece lógico no plantear ningún modelo.

5.6.5. Periódico "La Nueva España"

El número de días ajustados en este periódico ha sido bastante bajo, al igual que en el periódico "Levante-EMV" aunque levemente superior, con un total de 60 días (33%). Los parámetros "a" y "b" del modelo cumplieron el test de F-Anova, pero no el

parámetro λ , según se aprecia en la Fig. 43, donde el par (sábado, martes) no lo cumpliría, y existirían otra serie de parejas que estarían en estado crítico, las cuales serían (sábado, jueves), (martes, lunes), (viernes, martes) y (miércoles, sábado).

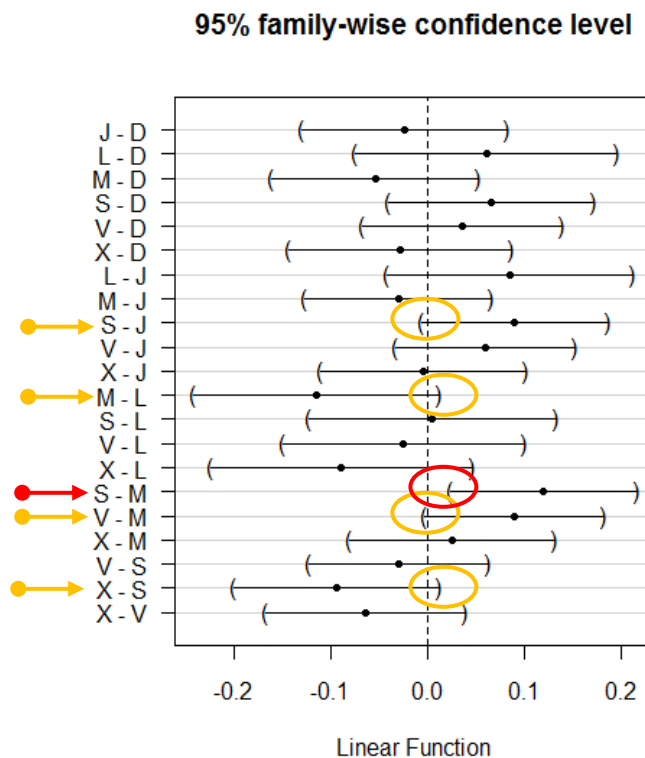


Figura 43. Test F-Anova para el parámetro " λ " entre pares de días de la semana.

Por lo tanto, se ha planteado un modelo general que caracterizaría la demanda de todos los días de la semana menos el sábado, y otro para el sábado exclusivamente.

Desde el punto de vista de las secciones, todas ellas modelan un elevado número de días, donde a excepción de la sección S9 que modela 65 días (36%), el resto de las secciones modelan como mínimo 139 días (76%), tal es el caso de la sección S5, y como máximo 182 días (100%) en el caso de la sección S1. Sin embargo, sólo sería admisible plantear un modelo general para la sección S3, y para la sección S7 siempre y cuando no se considere el lunes, para el resto sería totalmente inapropiado, dado que la desviación típica de los parámetros para cada uno de los días ha resultado ser bastante elevada en general.

5.6.6. Periódico "Superdeporte"

El número de días modelados ha sido bastante elevado, siendo un total de 140 días (77%), y se podría plantear un único modelo que caracterizaría la demanda de los vídeos presentados por este periódico, formado exclusivamente por la sección relativa a deportes.

5.7. Variabilidad del periodo

La longitud del intervalo de tiempo que es considerado a la hora de calcular los patrones de demanda es importante [15, 60, 61]. Lógicamente, a medida que aumenta la longitud del periodo el número de vídeos considerados populares es mayor, al igual que el número de vídeos con una popularidad similar. En esta tesis el estudio se ha planteado de forma diaria, dado que todos los días se producen nuevas incorporaciones de vídeos a los periódicos. Sin embargo, en este apartado se pretende ofrecer una pequeña aportación al estudio de la variación de los parámetros en los distintos modelos con periodos de dos días, tres días, hasta alcanzar periodos quincenales. En la Fig. 44 se puede ver el efecto de la distribución de las peticiones ante distintos periodos de tiempo en el periódico "La Opinión A Coruña" durante el mes de Junio de 2009. Los periodos de tres a nueve días presentan vídeos claramente diferenciadores y bastante populares, los periodos con 12 y 15 días son similares al de 9 días en la zona de la cabeza, y finalmente el periodo con una longitud de 30 días parece iniciar el descenso de la cabeza, y por lo tanto reflejaría el aumento de la presencia de vídeos con una popularidad similar.

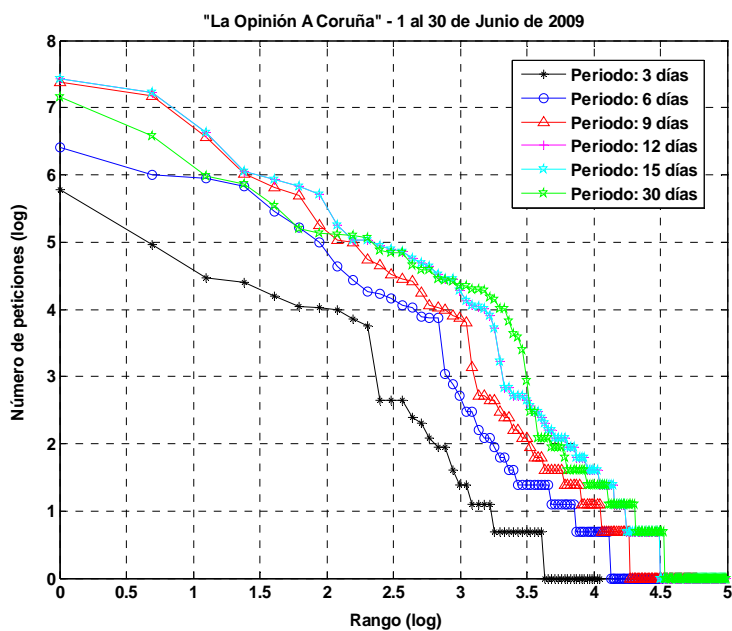


Figura 44. Distribución de peticiones con distintos intervalos de tiempo.

5.7.1. Distribución Zipf-like

En todos los periódicos, salvo "La Provincia", el valor del parámetro θ se incrementa a medida que aumenta la longitud del periodo, según se observa en la Tabla 21, donde el incremento entre un intervalo y el siguiente suele ser bajo. De hecho, si se tienen en cuenta todos los periódicos, excepto "La Provincia", y se analiza como fue el incremento entre un intervalo y el siguiente, con un total de 65 casos, se puede comprobar que en el 14% de las ocasiones (9 casos) fue de un 0%, en un 55% (36 casos) fue de un 1%, en un 21% (14 casos) fue de un 2%, y el 10% restante (6 casos) representa algún caso donde el incremento oscila entre un 3% y un 7%, o bien por el contrario se trataría de un descenso. En este último, en el "Faro de Vigo" del periodo de siete días al de ocho días θ pasó de 1,8744 a 1,7743, es decir, descendió casi en un 6%, pero se trata de casos muy puntuales.

Longitud Intervalo (N° días)	Valor de θ					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante"	"La Nueva España"	"Super-deporte"
2	1,6437 (σ : 0,2087)	1,6423 (σ : 0,3086)	x	1,8274 (σ : 0,1916)	1,6089 (σ : 0,1383)	1,8188 (σ : 0,3827)
3	1,6714 (σ : 0,1679)	1,7213 (σ : 0,2794)	2,8099 (σ : 0,9254)	1,8652 (σ : 0,1619)	1,6448 (σ : 0,1306)	1,8410 (σ : 0,3776)
4	1,7030 (σ : 0,1616)	1,7622 (σ : 0,2839)	2,3249 (σ : 0,155)	1,8652 (σ : 0,1619)	1,6684 (σ : 0,1264)	1,8861 (σ : 0,3611)
5	1,7161 (σ : 0,1258)	1,8087 (σ : 0,2659)	1,5846 (σ : 0,6095)	1,9195 (σ : 0,1499)	1,6976 (σ : 0,1267)	1,9003 (σ : 0,3125)
6	1,7346 (σ : 0,1171)	1,8421 (σ : 0,2350)	2,7633 (σ : 0,8109)	1,9247 (σ : 0,1524)	1,7137 (σ : 0,1205)	1,9201 (σ : 0,3207)
7	1,7622 (σ : 0,1194)	1,8744 (σ : 0,2365)	3,0367 (σ : 0,8044)	1,9503 (σ : 0,1433)	1,7341 (σ : 0,1189)	1,9464 (σ : 0,3282)
8	1,7743 (σ : 1359)	1,7743 (σ : 0,1359)	2,8422 (σ : 1,0527)	1,9609 (σ : 0,1411)	1,7454 (σ : 0,1056)	1,9872 (σ : 0,3238)
9	1,7843 (σ : 0,1194)	1,7843 (σ : 0,1194)	2,2102 (σ : 0,2543)	1,9717 (σ : 0,1211)	1,7602 (σ : 0,1044)	1,9910 (σ : 0,2932)
10	1,7975 (σ : 0,1067)	1,9134 (σ : 0,2112)	2,5851 (σ : 0,2671)	1,9733 (σ : 0,1504)	1,7735 (σ : 0,0985)	2,0071 (σ : 0,3219)
11	1,8022 (σ : 0,0958)	1,9337 (σ : 0,2029)	2,2344	1,9961 (σ : 0,1088)	1,7798 (σ : 0,1146)	2,0179 (σ : 0,3101)
12	1,8173 (σ : 0,0935)	1,9479 (σ : 0,1921)	2,5630 (σ : 0,3008)	1,9880 (σ : 0,1477)	1,7935 (σ : 0,1053)	2,0501 (σ : 0,2986)
13	1,8342 (σ : 0,1049)	1,9556 (σ : 0,1706)	2,4041 (σ : 0,4265)	2,0063 (σ : 0,1051)	1,8011 (σ : 0,1086)	2,0724 (σ : 0,2581)
14	1,8346 (σ : 0,0902)	1,9915 (σ : 0,1627)	2,0262 (σ : 0,6182)	2,0128 (σ : 0,1270)	1,8148 (σ : 0,1087)	2,0856 (σ : 0,3132)
15	1,8447 (σ : 0,0895)	1,9716 (σ : 0,1797)	1,8994 (σ : 0,8493)	2,0096 (σ : 0,1153)	1,8152 (σ : 0,0984)	2,1007 (σ : 0,2482)

Tabla 21. Variación del parámetro θ con periodos de distinta longitud.

Por otro lado, el periódico "La Provincia" es un caso excepcional, ya que dado un intervalo de tiempo i -ésimo se consiguió modelar un bajísimo número de grupos formados por la demanda efectuada durante un conjunto de " i " días, donde la diferencia

con el resto de los periódicos es notable. Por lo tanto, al disponer de tan pocos datos, tal y como aparece reflejado en la Tabla 22, el valor del parámetro θ no sería muy fiable.

Longitud Intervalo (Nº días)	Nº de grupos modelados					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"	"Superdeporte"
2	91	91	0	88	91	68
3	60	60	2	58	60	52
4	45	45	2	44	45	42
5	36	36	2	35	36	34
6	30	30	3	30	30	29
7	26	26	3	25	26	25
8	22	22	3	22	22	21
9	20	20	3	20	20	20
10	18	18	2	18	18	18
11	16	16	1	16	16	16
12	15	15	2	15	15	15
13	14	14	2	14	14	14
14	13	13	3	13	13	13
15	12	12	3	12	12	12

Tabla 22. Número de grupos modelados con periodos de distinta longitud.

En conclusión, en periodos con una longitud igual o inferior a los quince días, la aparición de nuevos vídeos con una popularidad similar no es significativa, dado que a medida que el tamaño del intervalo crece el valor del parámetro debería decrecer, y sin embargo ocurre el efecto contrario.

5.7.2. Distribución Mandelbrot

En el caso de la distribución Mandelbrot, a medida que se incrementa el tamaño del intervalo, la aparición de vídeos con un popularidad similar se pondría de manifiesto con el crecimiento del valor θ (efecto contrario a la distribución Zipf-like), y el decrecimiento del parámetro k . El único periódico donde se refleja un descenso claro del valor del parámetro θ a medida que aumenta la longitud del intervalo es "La Opinión A Coruña", tal y como se observa en la Tabla 23.

Longitud Intervalo (Nº días)	Valor de θ					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"	"Super-deporte"
2	4,0608 (σ : 2,1032)	3,6603 (σ : 3,6250)	5,1536 (σ : 4,1830)	3,0543 (σ : 1,4598)	2,3314 (σ : 0,9435)	5,8729 (σ : 4,1938)
3	3,5822 (σ : 1,3895)	4,2278 (σ : 3,4809)	5,7968 (σ : 4,2639)	2,5834 (σ : 0,9425)	2,019 (0,5835)	6,0949 (σ : 4,0721)
4	3,1033 (σ : 1,3419)	3,7991 (σ : 3,0222)	5,7052 (σ : 5,2041)	2,5834 (σ : 0,9425)	1,8382 (σ : 0,5595)	5,6372 (σ : 4,0249)
5	2,9205 (σ : 1,1912)	4,7961 (σ : 3,6996)	4,9282 (σ : 4,5724)	2,2126 (σ : 0,6727)	1,7828 (σ : 0,4805)	5,6654 (σ : 3,9105)
6	2,7187 (σ : 0,8671)	4,9193 (σ : 3,6178)	5,1956 (σ : 4,8679)	2,1197 (σ : 0,6673)	1,7093 (σ : 0,3667)	6,2645 (σ : 4,3378)
7	2,5669 (σ : 0,8315)	4,3327 (σ : 2,8896)	4,205 (σ : 3,0859)	2,1216 (σ : 0,6677)	1,6469 (σ : 0,3462)	5,1784 (σ : 3,3007)
8	2,2259 (σ : 0,6593)	4,1845 (σ : 2,5481)	5,2037 (σ : 4,8219)	1,9282 (σ : 0,5344)	1,5795 (σ : 0,2781)	4,6228 (σ : 2,8083)
9	2,1665 (σ : 0,6535)	4,569 (3,0372)	5,2694 (σ : 5,3229)	1,8495 (σ : 0,4733)	1,5835 (σ : 0,3369)	4,618 (σ : 2,9219)
10	1,9422 (σ : 0,6625)	4,3311 (σ : 2,4579)	4,9986 (σ : 4,5246)	1,8772 (σ : 0,5126)	1,4833 (σ : 0,2721)	4,1622 (σ : 3,1103)
11	2,0137 (σ : 0,5338)	4,7675 (σ : 3,4855)	5,9938 (σ : 5,3497)	1,6719 (σ : 0,4347)	1,5262 (σ : 0,2603)	4,4894 (σ : 2,5849)
12	1,8387 (σ : 0,5059)	4,0147 (σ : 2,1109)	6,1518 (σ : 5,6350)	1,7553 (σ : 0,3977)	1,4547 (σ : 0,2911)	4,9013 (σ : 3,2868)
13	1,77 (σ : 0,5982)	3,8171 (σ : 2,0692)	5,85 (σ : 5,4295)	1,7414 (σ : 0,4409)	1,5193 (σ : 0,2905)	4,2128 (σ : 2,7167)
14	1,7269 (σ : 0,4342)	3,9623 (σ : 1,2484)	3,8708 (σ : 3,2099)	1,7246 (σ : 0,5115)	1,49 (σ : 0,2397)	3,8615 (σ : 2,3527)
15	1,695 (σ : 0,4243)	3,1833 (σ : 1,4356)	4,222 (σ : 3,8478)	1,5633 (σ : 0,3779)	1,4267 (σ : 0,2877)	4,0758 (σ : 2,2624)

Tabla 23. Número de grupos modelados con periodos de distinta longitud.

En la Tabla 24 se puede observar que los distintos valores para el parámetro k presentan desviaciones típicas altas, donde además dentro de cada periódico tampoco se discierne una tendencia clara de su evolución, pues de un intervalo al siguiente existen

fluctuaciones de ascenso y/o descenso. Y por lo tanto el establecimiento de una pauta de comportamiento general se plantea difícil.

Longitud Intervalo (N° días)	Valor de k					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"	"Super-deporte"
2	8,2143 (σ : 6,3215)	2,5989 (σ : 5,0422)	1 (σ : 2,3412)	7,0454 (σ : 6,1319)	4,5165 (σ : 4,9517)	5,8809 (σ : 6,5499)
3	8,9333 (σ : 6,3121)	4,1833 (σ : 5,9332)	2,5682 (σ : 5,2128)	6,7845 (σ : 5,7423)	4,175 (σ : 4,4358)	7,3246 (σ : 6,7965)
4	8,1444 (σ : 6,5084)	3,8778 (σ : 5,9774)	1,8095 (σ : 3,5266)	6,7845 (σ : 5,7423)	3,8444 (σ : 4,4972)	6,9535 (σ : 6,8858)
5	8,1528 (σ : 6,0152)	5,8333 (σ : 6,7295)	2,2954 (σ : 4,6564)	6,9286 (5,6153)	4,5417 (σ : 5,1677)	7,1857 (σ : 6,6720)
6	8,2167 (σ : 6,0110)	6,5167 (σ : 6,6274)	1,8437 (σ : 4,0936)	6,8333 (σ : 5,8990)	4,1167 (σ : 4,1162)	8,7931 (σ : 6,6137)
7	8,4231 (σ : 6,1542)	5,75 (σ : 5,8775)	1,5278 (σ : 3,7)	7,58 (σ : 5,9313)	4,25 (σ : 4,6027)	7,68 (σ : 6,4386)
8	6,8182 (σ : 6,1384)	6,0227 (σ : 5,8482)	2 (σ : 3,7859)	6,6818 (σ : 5,4719)	3,6364 (σ : 3,1136)	7,3571 (σ : 5,9564)
9	7,5 (σ : 6,5765)	7,25 (σ : 6,6283)	2,8125 (5,2753)	6,275 (σ : 5,1362)	4,55 (σ : 4,5332)	7,35 (σ : 6,3102)
10	5,7222 (σ : 6,1434)	7,4167 (σ : 6,5310)	2,4643 (σ : 4,6221)	7,5278 (σ : 6,0645)	3,6667 (σ : 4,0147)	6 (σ : 5,9136)
11	7,0312 (σ : 6,3152)	7,4687 (σ : 6,1899)	3,2692 (σ : 4,4703)	5,9375 (σ : 5,1117)	4,5 (σ : 4,4907)	8,875 (σ : 6,2783)
12	6,1667 (σ : 6,2839)	7,3333 (σ : 5,8878)	4,3182 (σ : 5,1926)	7,6667 (σ : 5,8084)	4,1333 (σ : 4,5883)	9,6 (σ : 6,4204)
13	5,75 (σ : 6,1729)	6,2143 (σ : 5,6149)	3,7273 (σ : 5,4331)	7,5 (σ : 6,0415)	5,5357 (σ : 5,1233)	8,5 (σ : 6,8134)
14	6 (σ : 6,2216)	7,7308 (σ : 4,7020)	1,5 (σ : 4,1341)	7,8846 (σ : 6,5133)	5,2692 (σ : 5,0729)	7,5 (σ : 6,4646)
15	5,5 (σ : 5,7366)	5,5417 (σ : 5,2763)	2 (σ : 4,5643)	6,25 (σ : 5,4543)	4,5 (σ : 4,5776)	8,8333 (σ : 6,3437)

Tabla 24. Número de grupos modelados con periodos de distinta longitud.

5.7.3. Distribución Stretched

En líneas generales, se podría afirmar que los valores de los parámetros "a" y "c" incrementan a medida que aumenta la longitud del intervalo estudiado, hecho que también ha sido confirmado en [61], aunque existen casos puntuales donde no se cumple. Por ejemplo, tal y como se puede observar en las Tablas 25 y 26, en los periódicos "La Opinión A Coruña", "Faro de Vigo" y "Superdeporte" el valor de los parámetros a y c sufre un leve descenso en el intervalo de tres días respecto al intervalo anterior de dos días. En contraposición, en el periódico "La Provincia" los valores no siguen una tendencia específica, y es que al igual que ocurrió en la distribución Zipf-like el número de casos que se consiguió modelar en cada intervalo fue muy bajo, como mínimo fueron 0 y como máximo 5, lo que hace que dicho número no sea estadísticamente consistente como para extraer un valor medio válido.

Por otra parte, no se observa que dichos valores tiendan hacia una constante a medida que el intervalo de tiempo aumenta, tal y como se afirma en [60, 61], lo cual pueda ser debido a que la longitud máxima del intervalo fue de quince días, y quizás el estudio de intervalos con una amplitud superior pudiese corroborar esta afirmación.

Longitud Intervalo (Nº días)	Valor de c					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"	"Super-deporte"
2	0,0462 (σ : 0,0637)	0,0258 (σ : 0,0448)	x	0,0878 (σ : 0,0841)	0,0463 (σ : 0,0477)	0,0072 (σ : 0,0241)
3	0,0429 (σ : 0,0528)	0,0223 (σ : 0,0421)	0,0417 (σ : 0,0722)	0,1153 (σ : 0,0855)	0,0569 (σ : 0,0499)	0,0062 (σ : 0,0234)
4	0,0511 (σ : 0,0562)	0,0274 (σ : 0,0446)	0	0,1153 (σ : 0,0855)	0,0657 (σ : 0,0539)	0,0089 (σ : 0,0249)
5	0,0515 (σ : 0,0509)	0,0229 (σ : 0,0399)	0,0937 (σ : 0,1127)	0,1505 (σ : 0,0741)	0,0693 (σ : 0,0521)	0,0069 (σ : 0,0239)
6	0,0642 (σ : 0,0640)	0,0246 (σ : 0,0356)	0,0098 (σ : 0,0169)	0,1639 (σ : 0,0789)	0,0777 (σ : 0,0575)	0,0085 (σ : 0,0192)
7	0,0841 (σ : 0,0652)	0,0243 (σ : 0,0342)	0,1066 (σ : 0,1198)	0,1703 (σ : 0,0692)	0,0849 (σ : 0,0621)	0,01 (σ : 0,0243)
8	0,0841 (σ : 0,0652)	0,0304 (σ : 0,0344)	0,125 (σ : 0,25)	0,2112 (σ : 0,0775)	0,087 (σ : 0,0559)	0,0152 (σ : 0,0309)
9	0,1096 (σ : 0,0604)	0,0324 (σ : 0,0379)	0,0781 (σ : 0,1183)	0,1914 (σ : 0,0661)	0,1016 (σ : 0,0593)	0,0183 (σ : 0,0469)
10	0,1249 (σ : 0,0669)	0,0349 (σ : 0,0392)	0,0937 (σ : 0,1362)	0,2187 (σ : 0,0771)	0,1102 (σ : 0,0698)	0,0177 (σ : 0,0334)
11	0,1391 (σ : 0,0657)	0,0431 (σ : 0,0467)	0	0,2239 (σ : 0,0691)	0,1127 (σ : 0,0623)	0,0317 (σ : 0,0546)
12	0,1561 (σ : 0,0669)	0,0393 (σ : 0,0495)	0,2793 (σ : 0,2549)	0,2229 (σ : 0,0634)	0,1151 (σ : 0,0695)	0,0347 (σ : 0,0533)
13	0,1869 (σ : 0,0709)	0,0313 (σ : 0,0416)	0,1849 (σ : 0,2742)	0,2254 (σ : 0,0461)	0,1199 (σ : 0,0675)	0,0358 (σ : 0,0446)
14	0,1749 (σ : 0,0647)	0,0394 (σ : 0,0461)	0,1406 (σ : 0,2414)	0,2392 (σ : 0,0233)	0,1271 (σ : 0,0744)	0,0484 (σ : 0,0658)
15	0,2015 (σ : 0,0779)	0,0419 (σ : 0,0573)	0,0208 (σ : 0,0361)	0,2461 (σ : 0,0542)	0,1243 (σ : 0,0748)	0,0413 (σ : 0,0511)

Tabla 25. Número de grupos modelados con periodos de distinta longitud.

Longitud Intervalo (Nº días)	Valor de a					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"	"Super-deporte"
2	-0,0862 (σ : 0,1259)	-0,0473 (σ : 0,0893)	x	-0,1554 (σ : 0,1492)	-0,0688 (σ : 0,0689)	-0,0156 (σ : 0,0532)
3	-0,0754 (σ : 0,0941)	-0,0429 (σ : 0,0878)	-0,1248 (σ : 0,2162)	-0,1965 (σ : 0,1483)	-0,0837 (σ : 0,0711)	-0,0123 (σ : 0,0491)
4	-0,0897 (σ : 0,1015)	-0,0534 (σ : 0,0926)	0	-0,1965 (σ : 0,1483)	-0,0947 (σ : 0,0742)	-0,0193 (σ : 0,0557)
5	-0,0874 (σ : 0,0876)	-0,0456 (σ : 0,0816)	-0,3683 (σ : 0,5686)	-0,2373 (0,1089)	-0,0997 (σ : 0,0706)	-0,0148 (σ : 0,0522)
6	-0,1093 (σ : 0,0749)	-0,0475 (σ : 0,0719)	-0,0255 (σ : 0,0442)	-0,2578 (σ : 0,1337)	-0,1093 (σ : 0,0749)	-0,0170 (σ : 0,0388)
7	-0,1409 (σ : 0,1091)	-0,0477 (σ : 0,0689)	-0,2728 (σ : 0,3108)	-0,2541 (σ : 0,0917)	-0,1174 (σ : 0,0798)	-0,0201 (σ : 0,0487)
8	-0,1757 (σ : 0,1161)	-0,0591 (σ : 0,0685)	-0,2749 (σ : 0,5498)	-0,3110 (σ : 0,1289)	-0,1199 (σ : 0,0707)	-0,0307 (σ : 0,0622)
9	-0,1783 (σ : 0,0971)	-0,0625 (σ : 0,0752)	-0,1584 (σ : 0,2191)	-0,2722 (σ : 0,0782)	-0,1366 (σ : 0,0713)	-0,0357 (σ : 0,0898)
10	-0,2 (σ : 0,1031)	-0,0668 (σ : 0,0743)	-0,2251 (σ : 0,3174)	-0,2946 (σ : 0,0931)	-0,1434 (σ : 0,0819)	-0,0348 (σ : 0,0644)
11	-0,2176 (σ : 0,0963)	-0,0833 (σ : 0,0898)	0	-0,2987 (σ : 0,0702)	-0,1460 (σ : 0,0734)	-0,0607 (σ : 0,1017)
12	-0,2422 (σ : 0,0975)	-0,0751 (σ : 0,0943)	-0,6575 (σ : 0,6196)	-0,2929 (σ : 0,0787)	-0,1472 (σ : 0,0758)	-0,0683 (σ : 0,1021)
13	-0,2830 (σ : 0,1039)	-0,0603 (σ : 0,0808)	-0,5298 (σ : 0,7853)	-0,2961 (σ : 0,0488)	-0,1514 (σ : 0,0750)	-0,0705 (σ : 0,0866)
14	-0,2621 (σ : 0,0855)	-0,0753 (σ : 0,0862)	-0,4159 (σ : 0,7207)	-0,3076 (σ : 0,0211)	-0,1567 (σ : 0,0813)	-0,0912 (σ : 0,1169)
15	-0,2951 (σ : 0,0997)	-0,0798 (σ : 0,1062)	-0,0259 (σ : 0,0449)	-0,3059 (σ : 0,0466)	-0,1533 (σ : 0,0771)	-0,0805 (σ : 0,0973)

Tabla 26. Número de grupos modelados con periodos de distinta longitud.

5.8. Concentración de la demanda

La velocidad de crecimiento de la popularidad puede proporcionar otro indicio sobre las posibles necesidades de recursos. Por este motivo es importante establecer el concepto de tiempo de vida de un vídeo, lo que en la literatura se conoce como "*lifetime*" [70]. En este contexto se interpreta como el número de días que un vídeo recibe peticiones desde el primer día de su publicación. Por otra parte, se ha analizado la velocidad a la que un vídeo gana popularidad durante su primer día de vida para cada uno de los servicios. En la Fig. 45 se representa que porcentaje de nuevos vídeos recibieron entre un 10% y un 100% sobre el total de sus peticiones en su primer día de vida. El mejor porcentaje de nuevos vídeos dentro de cada periódico resultó ser el siguiente: en "La Opinión A Coruña" el 30% (6 de 20) de vídeos alcanzaron el 30% de peticiones, en el "Faro de Vigo" el 30% (3 de 10) de vídeos recibieron el 90% de peticiones, en "La Provincia" el 17% de los vídeos recibieron el 10%, en "Levante-EMV" el 15% de los vídeos recibieron el 70% de las peticiones, en "La Nueva España" el 18% de los vídeos recibieron el 100%, y finalmente en "Superdeporte" fue el más balanceado, dado que cada porcentaje de peticiones comprendidas en [10, 30, 40, 50, 60, 80] fueron acaparadas por el 12% de los vídeos. Además, en los periódicos "La Provincia", "Levante-EMV", "La Nueva España" y "Superdeporte" un alto porcentaje de vídeos recibieron más del 80% de las peticiones en su primer día de vida, donde el porcentaje de los mismos fue de un 24%, 24%, 38% y 23% respectivamente. Por lo tanto, a la vista de los resultados todos los periódicos presentarían el mismo nivel de consumo de recursos.

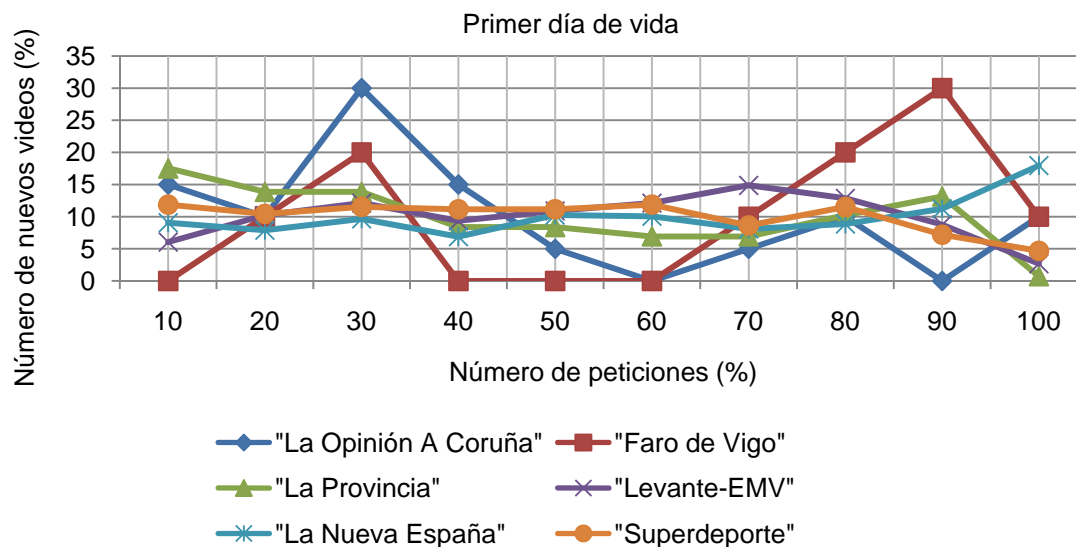


Figura 45. Número de peticiones (%) destinadas a nuevos vídeos (%) en su primer día de vida.

Tras el estudio de la concentración de los accesos sobre estos seis periódicos digitales se ha comprobado que en general las peticiones se concentran sobre unos pocos vídeos, tal y como indica el Principio de Pareto, y donde el resto de las peticiones se encuentran más diseminadas, lo cual puede ser debido al mayor número de vídeos presentes en este tipo de sistemas, este hecho también fue corroborado en [40, 44]. Los resultados son expresados con más detalle en la Tabla 27, donde se representa para cada periódico digital el porcentaje de vídeos que han recibido un porcentaje de accesos determinado, siendo este último presentado en intervalos de diez.

Nº Accesos (%)	Nº medio de vídeos accedidos por servicio (%)					
	"La Opinión A Coruña"	"Faro de Vigo"	"La Provincia"	"Levante-EMV"	"La Nueva España"	"Superdeporte"
[1-10]	19	17	15	18	24	13
[11-20]	15	12	13	18	17	11
[21-30]	14	11	12	16	15	11
[31-40]	14	11	10	16	15	10
[41-50]	13	11	12	14	13	11
[51-60]	10	11	12	11	8	10
[61-70]	9	8	12	4	4	11
[71-80]	6	10	9	0	4	10
[81-90]	0	9	5	3	0	8
[91-100]	0	0	0	0	0	5

Tabla 27. Porcentaje de accesos sobre el conjunto de vídeos de cada periódico digital.

Así pues, el servicio "La Nueva España" es el que presenta el mayor porcentaje de vídeos, por término medio, con menor porcentaje de accesos, siendo un 24% de los vídeos los que reciben un número de peticiones igual o inferior al 10%, seguido de "La Opinión A Coruña", "Levante-EMV", y "Faro de Vigo" con un 19%, 18% y 17% respectivamente, después "La Provincia" con un 15% y finalmente "Superdeporte" con un 13%. El servicio "La Nueva España" es el que sigue a la cabeza si se tiene en cuenta un porcentaje igual o inferior al 50% de los accesos, los cuales van destinados al 84% de los vídeos, seguido de "Levante-EMV" con un 82%, en tercer lugar "La Opinión A Coruña" con un 75%, "Faro de Vigo" y "La Provincia" estarían empatados en cuarta posición con un 62%, y finalmente de nuevo "Superdeporte" con un 56%. Además, como se puede observar, las peticiones están bastante distribuidas en todos los periódicos digitales, de hecho el reparto de las peticiones comprendidas entre el 1% y el 80% fue el siguiente:

- ✓ En "La Opinión A Coruña" el intervalo [1-10] tiene un 19% de vídeos accedidos, y entre el 11% y el 60% de los accesos el porcentaje de vídeos de cada intervalo osciló entre el 10% y el 15%. El 9% de los vídeos acapararon entre el 61% y el 70%, y por último entre el 71% y el 80% de los accesos fueron acaparados por el 6% de los vídeos.

- ✓ En el "Faro de Vigo" cada uno de los ocho intervalos que representan entre el 1% y el 80% de las peticiones fueron dirigidas como mínimo a un 8% y como máximo al 17% de los vídeos.
- ✓ En "La Provincia" los intervalos [71-80] y [31-40] tienen un porcentaje de vídeos accedidos de un 9% y de un 10% respectivamente, para el resto de los intervalos fue de un 12%, un 13% o un 15%.
- ✓ En "Levante-EMV" no hubo ningún vídeo que recibiese entre un 71% y un 80% de las peticiones, un 4% de los vídeos recibió entre un 61% y un 70% de las peticiones, y en el resto de los intervalos el porcentaje de vídeos accedidos osciló entre el 11% y el 18%.
- ✓ En "La Nueva España" entre el 51% y el 80% de los accesos fueron acaparados por 4% y el 8% de los vídeos, para el resto de los intervalos el porcentaje de vídeos accedidos osciló entre el 13% y el 24%.
- ✓ En "Superdeporte" todas las clases que se corresponden con un porcentaje de accesos entre el 11% y el 80%, tiene un porcentaje de vídeos accedidos de un 10% o un 11%, a excepción del intervalo [1-10] que tiene un 13%.

En definitiva, tras el análisis de los datos, la mayoría de los periódicos cumplen el Principio de Pareto (80-20) o (90-10), que significaría que aproximadamente el 80% o el 90% de los accesos estarían concentrados sobre el 20% o el 10% de los vídeos más populares respectivamente. En nuestro ámbito se observó que en tres de los periódicos digitales el número medio de los vídeos más populares estaría comprendido entre un 3% y un 9%, con un porcentaje de peticiones diarias comprendido entre el 81% y el 90%. El servicio "Superdeporte" sería el único donde un 5% de los vídeos más populares recibirían entre un 91% y un 100% de las peticiones, y para el resto de los periódicos ninguno de los vídeos recibió más del 90% de solicitudes.

Los datos anteriormente presentados englobarían la totalidad de los vídeos presentes en los periódicos digitales, pero cabría preguntarse como es la concentración de la demanda sobre los vídeos nuevos que fueron incorporados a los periódicos desde Enero a Junio de 2009. Los estudios realizados han extraído los siguientes resultados:

- ✓ En "La Opinión A Coruña" sólo 20 de los 358 vídeos creados recibieron solicitudes, de los cuales 17 pertenecen a la sección "Autonómico" (S3), 2 a la sección del "Tiempo" (S9), y 1 a "Otros" (S10), con un número de accesos de 2.019 (3,6%) , 501 (29%) y 170 (20%) respectivamente, dicho número aparece acompañado con el porcentaje de accesos que representaría frente al número total de accesos que recibió dicha sección. Además, cabe añadir que en la sección S3 hubo vídeos que como mínimo fueron accedidos durante un día, y como máximo 27 días, siendo la media de 10 días (σ : 2); en la sección S9 uno de los vídeos fue accedido durante 7 días y el otro durante 9 días, y en la sección S10 el vídeo creado fue accedido durante 7 días.
- ✓ En "El Faro de Vigo" sólo tres de los 98 vídeos creados recibieron solicitudes, de los cuales uno pertenece a la sección S2, otro a la S5 y el tercero a la sección S6. El número de accesos que acapararon fue de 147 (2%), 1.411 (26%) y 1.515 (10%) durante un periodo de 12, 90 y 59 días respectivamente.
- ✓ En "La Provincia" el número de vídeos creados han sido un total de 263, donde prácticamente el 98% (258 vídeos) recibieron peticiones. Todas las secciones recibieron solicitudes sobre alguno de los vídeos nuevos que fueron incorporados a las mismas. En la Tabla 28 se presenta información del número de vídeos nuevos que han recibido solicitudes, el número de días mínimo y máximo que un vídeo recibió solicitudes, el número medio de solicitudes con su desviación típica (σ), y el número total de accesos sobre los mismos.

Sección	Nº vídeos accedidos	Mínimo número días	Máximo número días	Nº medio de días	Nº accesos
S1	7	9	39	19 (σ : 2,3)	2.823
S2	3	11	87	39 (σ : 3,79)	2.707
S3	162	2	154	23 (σ : 19,65)	84.509
S4	3	6	9	7 (σ : 0,52)	289
S5	1	6	6	6 (σ : 0,25)	293
S6	21	6	46	24 (σ : 4,96)	17.862
S7	9	9	113	27 (σ : 5,21)	4.288
S8	31	6	93	28 (σ : 8,13)	34.212
S9	3	16	60	36 (σ : 2,94)	8.146
S10	18	3	72	19 (σ : 4,27)	8.585

Tabla 28. Datos relativos a los vídeos creados.

- ✓ En "Levante-EMV" fueron creados un total de 1.076 vídeos, de los cuales 1.046 vídeos (97%) recibieron peticiones. Al igual que en el periódico "La Provincia" en la Tabla 29 se proporciona información sobre los mismos.

Sección	Nº vídeos accedidos	Mínimo número días	Máximo número días	Nº medio de días	Nº accesos
S1	130	1	22	5 (σ : 1,55)	22.382
S2	79	1	15	6 (σ : 1,26)	8.074
S3	112	1	38	10 (σ : 2,57)	62.977
S4	242	1	14	4 (σ : 1,37)	34.290
S5	83	1	18	4 (σ : 0,99)	12.725
S6	75	1	35	6 (σ : 1,42)	14.470
S7	42	1	17	7 (σ : 1,11)	4.494
S8	87	1	34	11 (σ : 2,67)	29.432
S9	121	1	7	3 (σ : 0,82)	27.169
S10	74	1	101	19 (σ : 4,86)	37.282

Tabla 29. Datos relativos a los vídeos creados.

- ✓ En "La Nueva España" fueron creados un total de 1.819 vídeos, de los cuales el 26% (472 vídeos) recibieron accesos. Algunos datos sobre los mismos se resumen en la Tabla 30.

Sección	Nº vídeos accedidos	Mínimo número días	Máximo número días	Nº medio de días	Nº accesos
S1	39	1	26	4 (σ : 0,63)	5.460
S2	0	0	0	0	0
S3	16	1	57	7 (σ : 0,94)	3.347
S4	169	1	159	5 (σ : 2,94)	30.979
S5	78	1	47	7 (σ : 1,74)	24.021
S6	26	1	22	4 (σ : 0,5)	3.213
S7	64	1	63	8 (σ : 2,06)	13.697
S8	0	0	0	0	0
S9	20	1	34	5 (σ : 0,64)	1.785
S10	0	0	0	0	0

Tabla 30. Datos relativos a los vídeos creados.

- ✓ En "Superdeporte" se crearon 232 vídeos, de los cuales recibieron accesos un total de 219 (94%). El menor número de días que un vídeo recibió accesos fue uno y el máximo número de días fue de 104, donde por término medio un vídeo recibió accesos durante 11 días (σ : 8,76), siendo el total de accesos de 158.381.

5.9. Conclusiones

A lo largo de la literatura se puede encontrar distintas distribuciones estadísticas aplicadas para la caracterización de la popularidad de contenido multimedia. En este capítulo se ha abordado el estudio de las distribuciones más destacadas hasta la fecha, como son Zipf-like, two Zipf-like, Mandelbrot y Stretched, con el fin de comprobar la distribución más adecuada al servicio NoD que nos atañe. Además, se añade al estudio una nueva distribución denominada "transformación de Box-Cox", como una nueva aportación en este campo, que en alguno de los periódicos mejora los resultados de la caracterización.

El estudio se realiza día a día, dado que es el periodo de tiempo donde el contenido del periódico se mantiene estable. Para cada distribución se buscará el modelo general que mejor ajuste el comportamiento general diario de la popularidad. Sin embargo, en algunos casos no fue posible, dado que en función del día de la semana el comportamiento de la demanda era muy diferente del resto, lo cual ha llevado en

algunos casos a establecer modelos distintos asociados a un día de la semana o a varios. Por otro lado, también se han establecido modelos generales asociados a cada día de la semana. Si además lo que se pretende es modelar el comportamiento de la demanda por secciones, en la mayoría de las ocasiones sólo fue posible plantear un modelo para la sección S3 relativa al ámbito autonómico.

Aunque esta tesis está centrada en la caracterización de la popularidad sobre la demanda diaria de vídeos, también se ha realizado un breve análisis sobre el comportamiento de la popularidad en periodos con una longitud superior. Lógicamente, a medida que la longitud del periodo crece la presencia de vídeos con una popularidad alta y/o similar debería de aumentar. Este efecto se ve reflejado en la variación de los parámetros de los modelos, tal y como es corroborado en [61].

Desde el punto de vista de la concentración de los accesos se ha comprobado que en cuatro periódicos, las peticiones se concentran sobre unos pocos vídeos, tal y como indica el Principio de Pareto, y donde el resto de las peticiones se encuentran más dispersas, lo cual puede ser debido al mayor número de vídeos presentes en este tipo de sistemas.

Capítulo 6.
VALIDACIÓN DE LOS MODELOS

6.1. Introducción

En el estudio de los seis periódicos digitales durante el periodo comprendido entre el 1 de Enero de 2009 y el 1 de Julio de 2009 se han obtenido un serie de modelos tanto generales, como por días de la semana y secciones, los cuales aparecen detallados a lo largo de los distintos anexos (Anexo I, II, III, IV y V). Sin embargo, el hecho de encontrar un modelo que pueda definir el comportamiento de la demanda es una condición necesaria, pero no suficiente, dado que es necesario comprobar si realmente podrían generar un comportamiento de la demanda similar a la que están caracterizando. El proceso seguido para la validación de los modelos viene reflejado en el apartado 6.2.1. Por otro lado, suponiendo que un modelo o una serie de modelos sean considerados como válidos, es necesario comprobar si serían capaces de predecir la demanda sobre nuevos periodos. En nuestro caso ha sido aplicado sobre el periodo comprendido entre el 2 de Julio de 2009 hasta el 30 de Septiembre de 2009, sobre un total de 13 semanas (91 días). El grado de cobertura de los distintos modelos sobre dicho periodo ha sido estudiado en el apartado 6.2.3.

6.2. Validación de los modelos

6.2.1. Generación de poblaciones sintéticas

A partir de los accesos registrados entre el 1 de Enero de 2009 y el 1 de Julio de 2009, con un total de 182 días, se determinará el número de vídeos que han sido solicitados cada día dentro de dicho periodo, con el fin de obtener un número medio (μ) de vídeos accedidos diariamente, así como su desviación típica (σ). A continuación, para cada uno de los modelos establecidos, se genera una población sintética formada por N días, en nuestro estudio han sido 250 días (debido a limitaciones del paquete estadístico R), donde el número de vídeos accedidos un día dado vendrá establecido a través de una distribución normal $N(\mu, \sigma)$, cuyos parámetros fueron deducidos a partir de los 182 días. El proceso seguido viene representado en la Fig. 46.

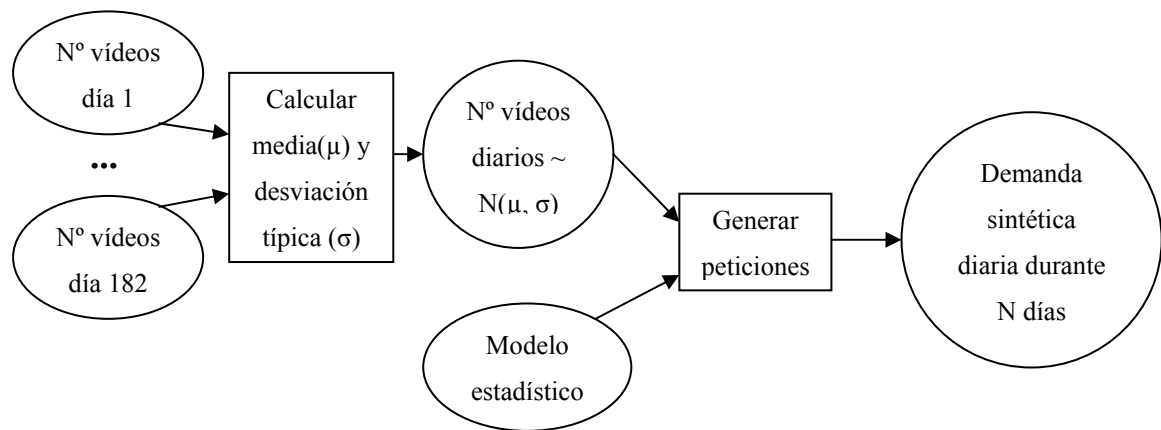


Figura 46. Proceso de generación de poblaciones sintéticas.

6.2.2. Tests de bondad de ajuste

Un test de bondad de ajuste permite saber si es razonable asumir que una muestra aleatoria pueda provenir de una distribución específica, donde la hipótesis nula y alternativa presentan la siguiente forma:

Ho: Los datos de la muestra provienen de la distribución establecida.

H1: Los datos de la muestra no provienen de la distribución establecida.

El test de Chi-cuadrado es el más antiguo, pues data del año 1900 y se debe a Karl Pearson. Se puede aplicar a cualquier distribución de la cual se pueda calcular la función de distribución acumulada. Los datos se dividen en k grupos, y el test estadístico se define de acuerdo a la expresión (32).

$$\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i} \quad (32)$$

O_i es la frecuencia observada del grupo i-ésimo y E_i es la frecuencia esperada. E_i se calcula mediante la función de distribución acumulada. El estadístico χ^2 se distribuye de acuerdo a una variable aleatoria Chi-cuadrado con $(k-p-1)$ grados de libertad, siendo p el número de parámetros estimados para la muestra. Los datos provendrán de la distribución especificada si χ^2 es menor que el valor de la función Chi-cuadrado con $(k-$

p-1) grados de libertad y un nivel de significación α . Este test presenta dos inconvenientes, por un lado si el tamaño de la muestra es pequeño el resultado del test podría no ser válido, y por otro lado es susceptible al número de grupos.

El test de Kolmogorov-Smirnov (K-S) se basa en la comparación entre la función de distribución empírica acumulada (ECDF) y la teórica definida en la ecuación (33), donde $f(y, \theta)$ representa la función de distribución de probabilidad..

$$F(x) = \int_0^x f(y, \theta) dy \quad (33)$$

Dados n puntos ordenados de forma creciente x_1, x_2, \dots, x_n , la función de distribución empírica acumulada (ECDF) se define de acuerdo a la ecuación (34), donde $N(i)$ es el número de puntos menores que x_i .

$$F_n(x_i) = N(i)/n \quad (34)$$

El test estadístico usado viene dado a través de la ecuación (35).

$$D_n = \sup_{1 \leq i \leq n} |F(x_i) - F_n(x_i)| \quad (35)$$

D_n es el extremo máximo de las diferencias en valor absoluto entre ECDF y la función de distribución teórica acumulada. Este test tiene dos características muy importantes. Por un lado, puede ser aplicado cuando el tamaño de la muestra no es muy grande, y por otro lado no necesita que los datos sean divididos en grupos.

Muchos artículos han usado el test de Chi-cuadrado para medir la bondad de ajuste [61, 62], mientras que otros han usado el test de Kolmogorov-Smirnov [40]. En esta tesis se han usado ambos. Un ejemplo de este problema se refleja en la Fig. 47, donde los accesos realizados el domingo día 20 de Septiembre de 2009 en el diario "La Provincia" son modelados a través de tres distribuciones de tipo Stretched, el modelo obtenido para el día en concreto 20 de Septiembre, el modelo general y el modelo general del domingo, los datos sobre los dos últimos modelos se encuentran en el Anexo IV. Tanto en el modelo general como en el modelo del domingo el test de Chi-cuadrado resultó positivo (General: $p_valor=0,9995261$; General domingo: $p_valor=0,9853448$), y por el contrario el test K-S fue negativo ($p_valor < 0,05$), donde los ajustes lineales R^2

resultaron ser de 0,8015 y 0,7703 respectivamente, siendo de 0,8701 el ajuste obtenido para el modelo específico de ese día en concreto.

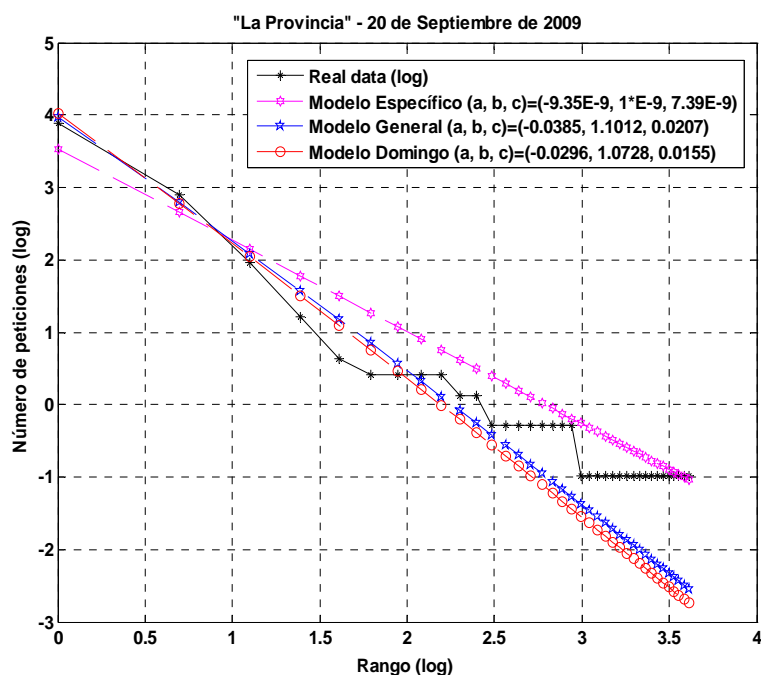


Figura 47. Ajuste con tres distribuciones de tipo Stretched.

Sin embargo, en la Fig. 48 se muestra el ajuste, también mediante la distribución Stretched, sobre los accesos realizados el sábado día 8 de Agosto de 2009, donde tanto el test de Kolmogorov-Smirnov como el de Chi-cuadrado resultaron ser positivos, tanto en el modelo general (Chi-cuadrado: $p_valor=0,940023$; K-S: $p_valor=0,9866454$) como en el general del sábado (Chi-cuadrado: $p_valor=0,94784$; K-S: $p_valor=0,9866454$). El modelo específico, general y el general para el sábado obtuvieron un ajuste lineal R^2 de 0,9134, 0,8997 y 0,9046 respectivamente.

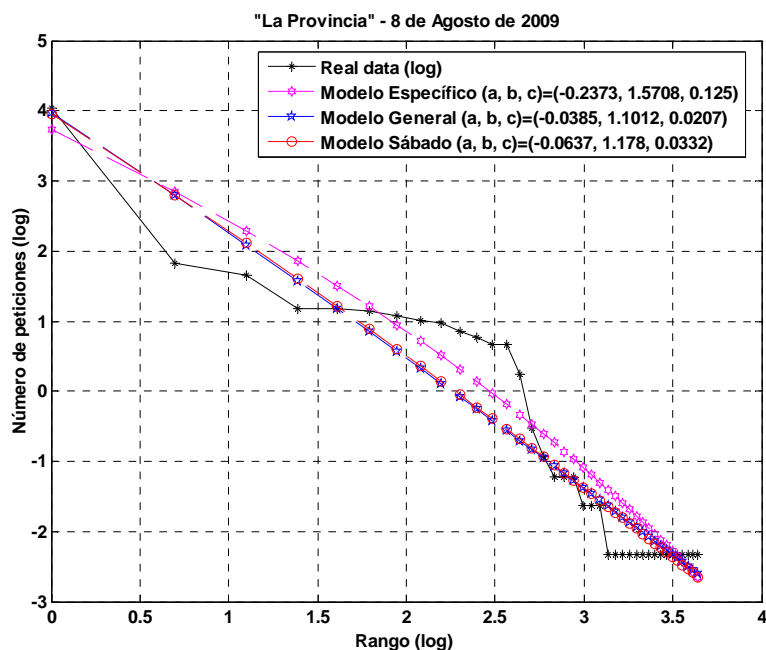


Figura 48. Ajuste con tres distribuciones de tipo Stretched.

Por lo tanto, el test K-S resulta ser más robusto en la validación de los modelos.

6.2.3. Validación de las poblaciones sintéticas

Dados N grupos de peticiones generadas según un modelo, es decir, 250 días como fue comentado en el apartado 6.2.1., y para un modelo estadístico específico (Zipf-like, Mandelbrot, Stretched o Box-Cox), se calcularán los valores de los parámetros de dicho modelo, con el objetivo de obtener el mejor ajuste al conjunto de datos correspondiente al grupo de peticiones sintéticas correspondiente. Al final se obtendrá un total de N modelos. Después, se hace necesario un proceso de filtrado, con el fin de eliminar aquellos modelos que no pasen un grado de ajuste aceptable, el cual será medido a través de un test de contraste de hipótesis como es Kolmogorov-Smirnov, de acuerdo a los motivos explicados en el apartado 6.2.2., siendo necesario para su aceptación que el p -valor sea mayor o igual a 0,05. También serán eliminados aquellos modelos donde el valor del parámetro vaya en contra de la naturaleza del mismo, es decir, si el valor de un parámetro representase la pendiente de una recta negativa, y el valor del mismo hubiese resultado ser positivo, dicho modelo debería ser descartado. También serán eliminados aquellos modelos donde el valor de alguno de sus parámetros

resulte ser atípico, es decir, se aleje del valor establecido como su media más allá de lo indicado por su desviación típica. A partir de los modelos considerados válidos, se procede a su incorporación con el resto de los modelos obtenidos para las clases correspondientes a cada uno de los días de la semana. En último lugar, se comparan los valores de los parámetros de todos los modelos agrupados por clases con los modelos obtenidos a partir de las poblaciones sintéticas generadas, los cuales serán considerados iguales o parecidos entre sí, siempre y cuando el test de F-Anova resulte ser positivo. En caso afirmativo se concluirá que el modelo inicial, a partir del cual se ha generado la población sintética, es un modelo válido que reflejaría el comportamiento de la demanda. Este proceso viene reflejado en la Fig. 49.

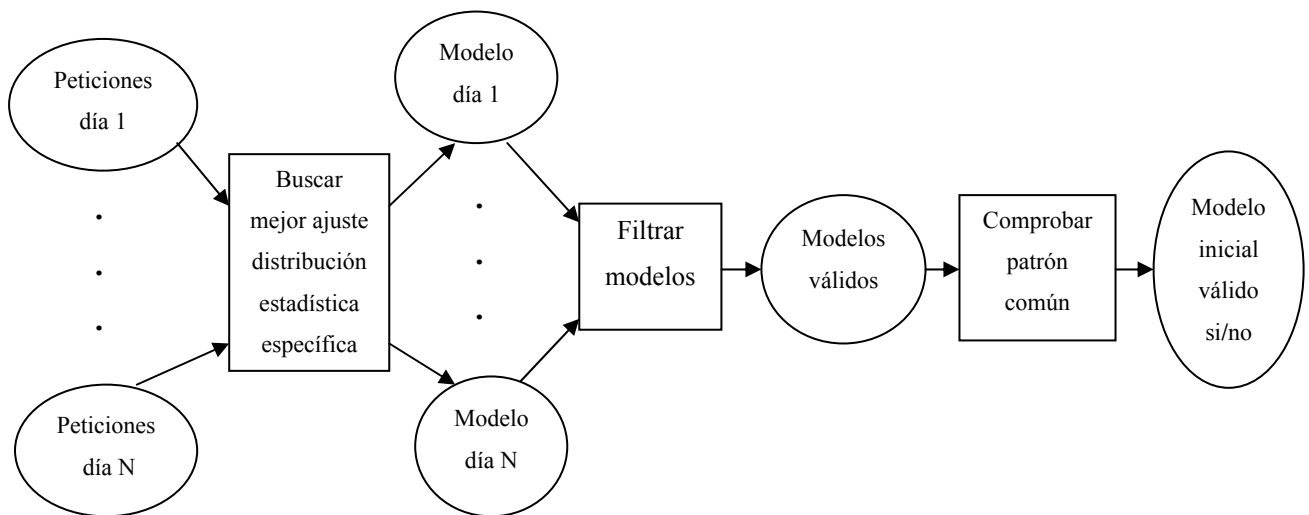


Figura 49. Proceso de validación de modelos.

A continuación se muestra un ejemplo de su aplicación sobre el periódico digital "Superdeporte" y la "transformación de Box-Cox", donde los resultados del test F-Anova para los parámetros a y b aparecen reflejados en la Fig. 50 (a) y (b), donde se comparan los parámetros de los modelos obtenidos desde el 1 de Enero de 2009 al 1 de Julio de 2009 con los modelos obtenidos en el ajuste de la población sintética, generada a partir del modelo general (Anexo V), con el fin de validar dicha población. Tal y como se puede observar el test resultó ser positivo en ambos casos.

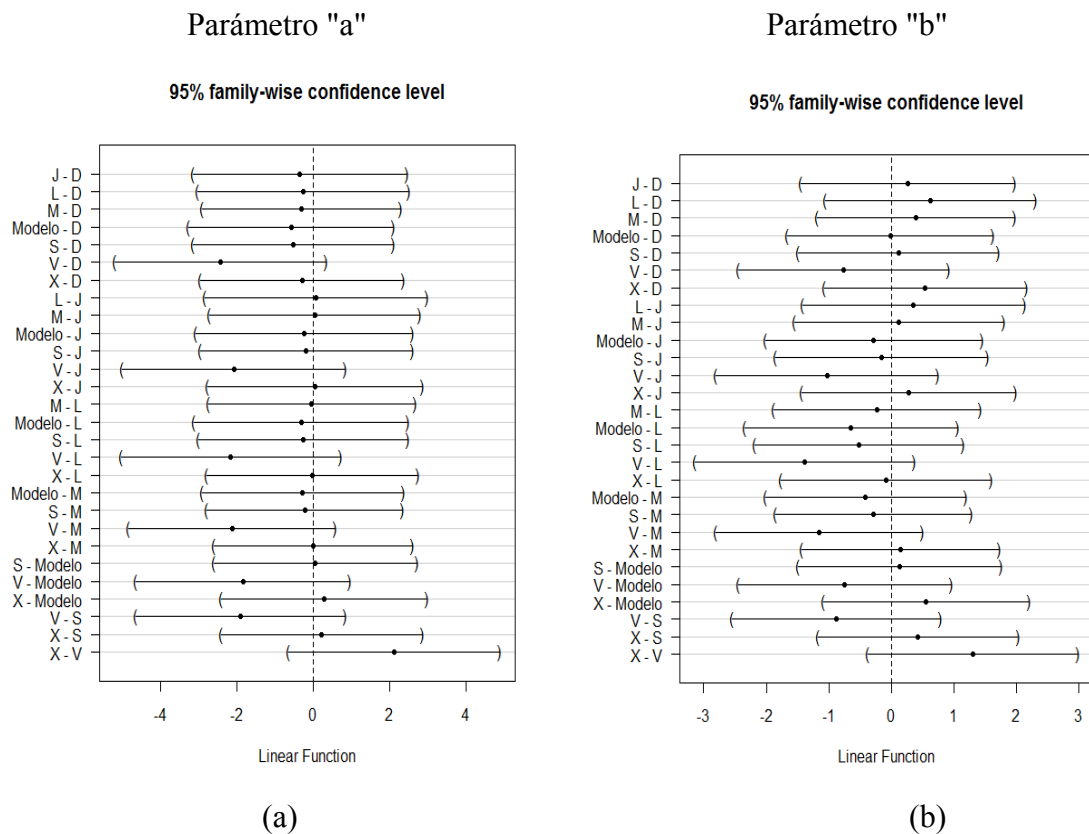


Figura 50. Test F-Anova para los parámetros a y b de la distribución de Box-Cox.

Este proceso ha sido llevado a cabo en cada uno de los periódicos, y los modelos obtenidos para cada uno de ellos se reflejan en los distintos Anexos.

6.3. Validación de los modelos

La validación de los distintos modelos se ha realizado teniendo en cuenta la demanda que se produjo entre el 2 de Julio de 2009 y el 30 de Septiembre de 2009, es decir, sobre un total de 91 días. Los resultados obtenidos son presentados en la Tablas 31 y 32. Cada una de las entradas de la tablas contiene tres números, el primero representa el número de días en los que tanto el test de Chi-cuadrado como el de K-S resultaron ser positivos simultáneamente, el segundo número representa el número de días que tuvo éxito el test Chi-cuadrado y el tercer número representa el número de días que el test K-S tuvo éxito.

En la Tabla 31 se presenta la validación de los modelos tanto generales como los obtenidos para cada día de la semana. Si nos centramos sólo en los resultados obtenidos

para el test de K-S, en términos generales se puede observar que la "transformación de Box-Cox" fue la que presentó los mejores resultados de cobertura o bien con una mínima diferencia respecto a otra distribución, aunque también existe algún caso excepcional como el modelo general obtenido para la "transformación de Box-Cox" en el periódico "Levante-EMV", el cual no consigue validar ningún día de los 91. En cada periódico digital se ha resaltado en verde la casilla que se corresponde con aquella distribución que consiguió validar el mayor número de días, y en color naranja la segunda mejor distribución.

Periódicos digitales	Modelo General			
	Zipf-like	Mandelbrot	Stretched	Box-Cox
"La Opinión A Coruña"	64 / 71 / 75	17 / 18 / 40	34 / 37 / 53	36 / 37 / 77
"Faro de Vigo"	59 / 69 / 66	14 / 16 / 31	68 / 75 / 74	74 / 72 / 74
"La Provincia"	34 / 52 / 51	12 / 18 / 32	35 / 49 / 46	57 / 66 / 57
"Levante-EMV"	53 / 66 / 67	35 / 59 / 37	68 / 81 / 76	0
"La Nueva España"	62 / 88 / 62	44 / 72 / 48	83 / 91 / 83	78 / 91 / 82
"Superdeporte"	42 / 55 / 54	34 / 46 / 52	34 / 45 / 40	55 / 52 / 54
Periódicos digitales	Modelos por días de la semana			
	Zipf-like	Mandelbrot	Stretched	Box-Cox
"La Opinión A Coruña"	67 / 68 / 79	34 / 32 / 61	45 / 51 / 59	45 / 54 / 78
"Faro de Vigo"	59 / 65 / 70	18 / 20 / 44	55 / 56 / 62	70 / 75 / 70
"La Provincia"	36 / 48 / 47	19 / 28 / 34	36 / 46 / 46	60 / 63 / 60
"Levante-EMV"	52 / 71 / 67	44 / 66 / 44	65 / 83 / 71	62 / 69 / 50
"La Nueva España"	61 / 89 / 60	46 / 79 / 49	79 / 91 / 78	79 / 91 / 83
"Superdeporte"	36 / 49 / 49	19 / 26 / 40	22 / 34 / 29	52 / 53 / 53

Tabla 31. Número de días con el test de Chi-cuadrado / Kolmogorov-Smirnov positivo.

A la vista de los resultados, desde el punto de vista del modelo general Box-Cox se consigue establecer como la mejor distribución, a excepción de "Levante-EMV" como fue comentado anteriormente, y en el periódico "La Nueva España" ocuparía la segunda posición a tan sólo un día de diferencia respecto a la distribución Stretched. Desde el punto de vista de los modelos generales establecidos para cada día de la semana, Box-Cox vuelve a ocupar la primera posición, excepto en el periódico "Levante-EMV" que ocuparía la tercera posición con un total de 50 días cubiertos con éxito, y respecto al periódico "La Opinión A Coruña" ocuparía el segundo puesto con tan sólo un día de diferencia. En definitiva, se puede observar un incremento en el número de días cuya demanda sea ajustada a través del modelo específico del día de la semana correspondiente, frente al número de días que son ajustados con éxito cuando se usa el modelo general, sobre todo en el periódico "Levante-EMV" los resultados han sido muy significativos pasando de cero días a un total de cincuenta. Lo lógico sería pensar que la aplicación del modelo específico en cada clase sería lo más adecuado y lo más eficaz, sin embargo a veces los resultados obtenidos frente al modelo general tampoco aportan una mejora sustancial, pues en la mayoría de los casos la diferencia es mínima. Así por ejemplo, en el periódico digital "La Opinión A Coruña" en las clases del miércoles, viernes y sábado ambos modelos (general y específico) consiguieron ajustar el mismo número de días, y para el lunes se consiguió el ajuste sobre dos días más, para el jueves un día más, para el domingo cuatro días más, pero para el martes el modelo específico empeora el ajuste produciéndose una pérdida de tres días. En conclusión, desde un punto de vista general, los modelos específicos asociados a un día concreto aumentarían el número de días que podrían ser ajustados a través de dicho modelo frente al modelo general, pero como contrapartida se pasaría de tener un único modelo a tener varios.

Otra forma de modelado sería por secciones, así pues en la Tabla 32 se presenta el resultado de cobertura de los modelos generales obtenidos para la caracterización de la popularidad de las secciones correspondientes a los distintos periódicos digitales. En general, la sección S3 pudo ser modelada con las cuatro distribuciones (Zipf-like, Mandelbrot, Stretched y Box-Cox), siendo las distribuciones Stretched y Zipf-like las más destacadas, junto con Mandelbrot en el caso de "Faro de Vigo" y "Levante-EMV". En el resto de secciones la distribución predominante fue Mandelbrot.

Periódicos digitales	Modelo general por secciones				
	Sección	Zipf-like	Mandelbrot	Stretched	Box-Cox
"La Opinión A Coruña"	S3	46 / 47 / 83	14 / 14 / 53	39 / 43 / 74	37 / 37 / 79
"Faro de Vigo"	S3	x	9 / 10 / 67	x	x
	S6	x	x	x	x
	S8	x	x	x	x
	S10	x	x	x	x
"La Provincia"	S3	14 / 14 / 58	16 / 18 / 48	16 / 17 / 82	16 / 16 / 44
	S8	x	12 / 18 / 79	x	x
"Levante-EMV"	S3	x	6 / 8 / 60	x	x
	S8	x	8 / 14 / 77	x	x
	S10	x	18 / 22 / 63	x	x
"La Nueva España"	S1	x	4 / 5 / 12	x	x
	S3	45 / 45 / 74	26 / 27 / 51	24 / 24 / 71	22 / 22 / 60
	S4	x	11 / 13 / 82	x	x
	S5	x	9 / 18 / 90	x	x
	S6	x	6 / 17 / 91	x	x
	S10	61 / 73 / 64	17 / 19 / 28	41 / 45 / 78	x

Tabla 32. Número de días con el test de Kolmogorov-Smirnov y Chi-cuadrado positivo en el modelo general.

En la Tabla 33 se presenta el resultado de cobertura de los modelos establecidos para cada día de la semana, en la caracterización de la popularidad de las secciones correspondientes a los distintos periódicos digitales. Al igual que en el caso anterior, la sección S3 pudo ser modelada con las cuatro distribuciones (Zipf-like, Mandelbrot, Stretched y Box-Cox), siendo las distribuciones Stretched y Zipf-like las más destacadas, junto con Mandelbrot en el caso de "Faro de Vigo" y "Levante-EMV". En el resto de secciones la distribución predominante fue Mandelbrot.

Periódicos digitales	Modelos por días de la semana en las secciones				
	Sección	Zipf-like	Mandelbrot	Stretched	Box-Cox
"La Opinión A Coruña"	S3	55 / 55 / 80	15 / 15 / 57	45 / 47 / 73	20 / 20 / 53
"Faro de Vigo"	S3	x	14 / 20 / 71	x	x
"La Provincia"	S3	21 / 24 / 65	15 / 15 / 48	19 / 20 / 75	15 / 15 / 48
	S8	x	18 / 24 / 81	x	x
"Levante-EMV"	S3	x	8 / 11 / 69	x	x
	S8	x	7 / 15 / 79	x	x
	S10	x	19 / 21 / 64	x	x
"La Nueva España"	S1	x	8 / 11 / 25	x	x
	S3	47 / 48 / 72	33 / 34 / 62	26 / 27 / 70	22 / 22 / 54
	S4	x	6 / 10 / 59	x	x
	S5	x	7 / 17 / 90	x	x
	S6	x	11 / 21 / 89	x	x
	S10	56 / 67 / 61	28 / 34 / 42	43 / 37 / 64	x

Tabla 33. Número de días con el test de Kolmogorov-Smirnov y Chi-cuadrado positivo en el modelo asociado a cada día de la semana.

6.4. Conclusiones

En los resultados de la validación, según el test de K-S, la transformación de Box-Cox fue la que presentó los mejores resultados de cobertura o bien con una mínima diferencia respecto a la distribución Stretched o Zipf-like, tanto desde el punto de vista del modelo general como en los modelos correspondientes a cada día de la semana. Sin embargo, como contrapartida en el periódico "Levante-EMV" no fue posible establecer un modelo general, y respecto a los modelos diarios tuvo un comportamiento peor que en Stretched o Zipf-like. En conclusión, desde un punto de vista global, los modelos específicos que caracterizan la popularidad de la clase correspondiente a un día concreto

de la semana son mejores, dado que aumentan el número de días que podrían ser ajustados a través de dicho modelo frente al modelo general.

Desde el punto de vista de las secciones, la distribución Mandelbrot sería la distribución predominante. Sin embargo, desde el punto de vista de la sección S3 en el caso de los periódicos "La Opinión A Coruña", "La Provincia" y "La Nueva España" parece mejor una distribución Zipf-like o Stretched. Por lo tanto, el planteamiento de un modelo por sección no sería adecuado dado que no se puede validar la mayoría de dichos modelos.

Capítulo 7.
CARACTERIZACIÓN DE LA
POPULARIDAD MEDIANTE
REDUNDANCIAS PARCIALES Y
GLOBALES

7.1. Introducción

El contenido de tipo streaming se ha convertido en uno de los principales componentes característicos en el tráfico vía Internet. Por lo tanto, el diseño de un sistema de distribución de vídeo bajo demanda debería ser lo más eficaz y eficiente posible, tanto en la gestión como en la planificación de sus recursos (caching, capacidad de almacenamiento, ancho de banda, etc). Un conocimiento previo sobre la tendencia que podría seguir el patrón de comportamiento de los usuarios en sus peticiones, es un pilar fundamental para conseguir alcanzar dicho objetivo. Este capítulo estará centrado en el análisis y modelado de patrones de redundancia presentes en secuencias de peticiones sobre los vídeos, es decir, en que medida se concentran las peticiones sobre los diferentes vídeos. Así, cuando las peticiones se concentran sobre unos pocos vídeos se trataría de una redundancia fuerte, en el caso contrario cuando las peticiones se encuentran más dispersas se trataría de una redundancia débil. El estudio fue realizado con las peticiones efectuadas a lo largo de todo el año 2010, en los vídeos presentes en los seis periódicos digitales que han sido objeto de estudio a lo largo de esta tesis.

El grado de redundancia ha sido calculado mediante un método de redundancia global (gR) y otro parcial (pR) [66]. La aplicación de estos métodos es nueva en este tipo de servicios. La principal contribución de este análisis ha sido la obtención de un modelo de redundancia parcial global para cada periódico digital, el cual podría predecir el grado de peticiones de los vídeos que tiene una alta probabilidad de repetirse en un futuro cercano. Una aportación importante es que el modelo resultó ser independiente tanto del usuario como del tiempo. La validación obtuvo excelentes resultados, dado que el modelo obtenido predice entre un 1% y un 6% de peticiones redundantes, con un nivel de seguridad entre un 88% y un 100%.

7.2. Métodos para el estudio de redundancias

En la redundancia global se mide el porcentaje de repeticiones sobre los vídeos que han ocurrido en un periodo completo, y se calcula mediante la relación entre el número de vídeos únicos que fueron accedidos y total de peticiones realizadas, de acuerdo a la expresión (36). Esta medida proporciona una idea del grado de dispersión de las peticiones a lo largo de los vídeos. En el caso de que todos los vídeos recibiesen

una sola petición se diría que no existe redundancia ($gR=0$). Sin embargo, a medida que el número de vídeos que acaparan un gran cantidad de solicitudes tienda a ser menor, la redundancia global incrementará considerablemente.

$$gR = 100 - 100 * \left(\frac{\# \text{vídeos únicos}}{\# \text{total peticiones}} \right) \quad (36)$$

Por otro lado, el conjunto de peticiones se ordenan cronológicamente en función de su fecha de solicitud. Dicho conjunto se divide en secuencias de igual longitud T , donde T representará un porcentaje sobre el total de peticiones del conjunto. A continuación, se comparan pares de secuencias entre sí, que se denotará por (Q_i, Q_j) , siendo ambas con igual longitud T . De esta forma, dadas dos secuencias Q_i y Q_j , la relación entre el número de vídeos de la primera secuencia Q_i que son referenciados en la segunda secuencia Q_j respecto al número total de peticiones en la primera secuencia Q_i , es una métrica que se denomina redundancia parcial (pR), y se calculará de acuerdo a la expresión (37).

$$pR(Q_i, Q_j) = 100 * \left(\frac{\# \text{vídeos únicos } Q_i \in Q_j}{\# \text{peticiones } Q_i} \right) \quad (37)$$

El par de secuencias en estudio pueden ser consecutivas, o bien tener una serie de secuencias entre ellas, que se representará mediante el parámetro delta (Δ). En la Fig. 51. se encuentra representado los pares de secuencia en estudio. En primer lugar, pR se estudia con pares de secuencias consecutivas ($pR(Q_1, Q_2)$, $pR(Q_3, Q_4)$, ...; $\Delta = 0$). En segundo lugar, pR se estudiará entre pares de secuencias con una secuencia de separación entre ellas ($pR(Q_1, Q_3)$, $pR(Q_2, Q_4)$, ...; $\Delta = 1$). A continuación, cada par de secuencias tendrán dos secuencias entre ellas ($pR(Q_1, Q_4)$, $pR(Q_2, Q_5)$, ...; $\Delta = 2$), y así sucesivamente. La máxima distancia de secuencias entre pares de secuencias y la longitud máxima de una secuencia T tendrán que ser establecidas inicialmente.

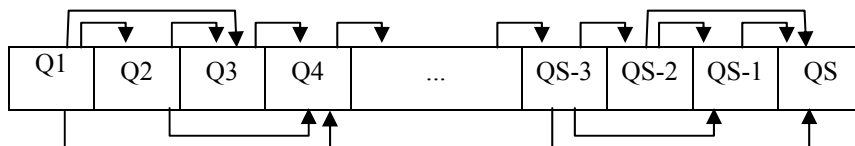


Figura 51. Pares de secuencias que forman parte del cálculo de pR .

Al final, el conjunto de peticiones será testeado con diferentes longitudes de secuencias T y diferentes valores de delta Δ . Para cada par de valores (T, Δ) se calculará un valor final para pR , el cual se obtiene a partir de la media de todos los valores de pR calculados a partir de pares de secuencias con el mismo tamaño T y la misma separación entre ellas. Por ejemplo, la redundancia parcial para secuencias con un tamaño T y delta igual a 0 será $pR(T, 0) = \text{media}(pR(Q_1, Q_2), pR(Q_2, Q_3), \dots, pR(Q_{S-1}, Q_S))$, si delta es igual a uno se calculará como $pR(T, 1) = \text{media}(pR(Q_1, Q_3), pR(Q_2, Q_4), \dots, pR(Q_{S-2}, Q_S))$, y así sucesivamente. El algoritmo será según la Tabla 34.

```

Funcion CalcularRedundanciaParcial(Peticiones): matriz[[T1,T2, ...,Tm],[0,Δ1,Δ2,..., Δq]]
//Entrada: Secuencia de peticiones sobre los vídeos 'Peticiones'
//Salida: Matriz 'redundanciaParcialMedia' de 'mxq', donde cada celda(i,j) representa
//      la media calculada a partir de todos pares de secuencias con longitud 'i' y
//      una distancia de 'j' secuencias entre ellos.

Peticiones={Item1, Item2, ..., ItemN} // Secuencia de peticiones vídeos, donde el Itemi
//representa el vídeo elegido en i-ésima posición.
N = Número de peticiones en un cierto periodo de tiempo
redundanciaParcialMedia[[T1, T2, ..., Tm], [0, Δ1, Δ2, ..., Δq]] = 0

Para cada secuencia T ∈ [T1, T2, ..., Tm], donde Ti representa un porcentaje
//comprendido entre un 1% y un 50% de N
L = N*T //Número de items en la secuencia
S = N/L //Número de secuencias con longitud L

//Construcción de S secuencias con longitud L cada una
posicionInicial = 1;
Para i desde 1 hasta S
    Q(i) = Peticiones[posicionInicial ... (posicionInicial + L - 1)];
    posicionInicial = posicionInicial + L;
FinPara i

Para cada valor delta (Δ) ∈ [0, Δ1, Δ2, ..., Δq]
    Para j desde 1 hasta (S-delta-1)
        redundanciaParcial(j) = CalcularRedundanciaParcial(Qj, Q(j+delta+1))
    FinPara j
    redundanciaParcialMedia(T,delta)=media(redundanciaParcial(1...(S-delta-1)))
FinPara delta
FinPara T

retorna redundanciaParcialMedia
Fin Funcion

```

Tabla 34. Algoritmo para calcular la redundancia parcial.

El periodo completo fue dividido en bloques con el mismo tamaño B , los cuales han sido numerados de uno en adelante. La función expresada en la Tabla 34 será aplicada sobre cada bloque de peticiones B , obteniendo para cada bloque una matriz con información relacionada con pares (T_i, Δ_j) . A partir de las matrices de los bloques impares se construye una única matriz, y a partir de los bloques pares se construye otra, donde ambas representarán modelos finales. El modelo obtenido a partir de los bloques pares será usado para validar el modelo de los bloques impares. En la Tabla 35 se encuentra expresado el algoritmo para construir ambos modelos.

```

Peticiones = {Item1, Item2, ..., ItemM} //Secuencia de peticiones de los videos para
// todo el periodo de tiempo con longitud M
P = M / B; //Número de bloques con longitud B

//Construcción de P bloques con longitud B
posicionInicial = 1;
Para i desde 1 hasta P
    B(i) = Peticiones[posicionInicial ... (posicionInicial + B - 1)];
    posicionInicial = posicionInicial+B;
FinPara i

//Calculo de pR para cada bloque
Para i desde 1 hasta P
    redundanciaParcialBlock_i(m, q) = CalcularRedundanciaParcial(Bi)
FinPara i

//Calculo del modelo final
Para i desde 1 hasta P
    Si(i es impar)
        redundanciaParcialFinalImpar(m, q) = redundanciaParcialFinalImpar(m, q)
            + redundanciaParcialBlock_i(m, q)
    else
        redundanciaParcialFinalPar(m, q) = redundanciaParcialFinalPar(m, q)
            + redundanciaParcialBlock_i(m, q)
    Fin
FinPara i

media = P/2
redundanciaParcialFinalPar(m, q) = redundanciaParcialFinalPar(m, q) / media

Si(P es impar)
    redundanciaParcialFinalImpar(m, q) = redundanciaParcialFinalImpar(m, q) / (media+1)
sino
    redundanciaParcialFinalImpar(m, q) = redundanciaParcialFinalImpar(m, q) / media
Fin
    
```

Tabla 35. Algoritmo para el cálculo de la redundancia parcial final.

7.3. Análisis de la redundancia

Inicialmente, el problema se abordó tomando en conjunto los accesos de los seis primeros meses del año 2.010, de Enero a Junio, para el cálculo de un modelo de redundancia parcial en cada periódico digital siguiendo el algoritmo representado en el apartado 7.2. en la Tabla 34. El proceso de validación se realizó con los accesos de los seis meses restantes, de Julio a Diciembre. El número de vídeos y accesos en cada uno de los semestres fueron los mostrados en la Tabla 36.

		Periódicos digitales					
Año 2010	Datos	"La Opinión a Coruña"	"Faro de Vigo"	"Levante-EMV"	"La Provincia"	"La Nueva España"	"Superdeporte"
Enero a Junio	# Vídeos	420	420	833	632	1.957	869
	# Accesos	9.704	27.824	65.158	130.001	271.975	183.876
Julio a Diciembre	# Vídeos	350	396	705	705	1.623	1.161
	# Accesos	19.949	20.625	32.646	81.753	137.632	234.456

Tabla 36. Número de accesos y vídeos en los dos semestres del año 2.010.

El conjunto de peticiones de Enero a Junio fue estudiado con secuencias de peticiones T, con tamaños pertenecientes a $[0,0055; 0,011; 0,042; 0,08; 0,17; 0,25; 0,33; 0,5]$. Así por ejemplo, si T es igual a 0,5 se estaría estudiando la proporción de vídeos del primer trimestre que son accedidos en el segundo trimestre. Aunque se consiguió encontrar al menos un modelo para Δ igual a 0 en cada uno de los periódicos, dichos modelos no consiguieron validar el periodo de Julio a Diciembre en cada uno de los periódicos correspondientes. Este hecho se puede observar en las Figs. 52 y 53, en el periódico digital "La Nueva España". En consecuencia, el estudio de la redundancia en periódicos digitales donde la inserción de nuevo contenido es diaria no se debería realizar en periodos demasiado grandes, pues como se ha podido constatar fue inviable en todos los casos.

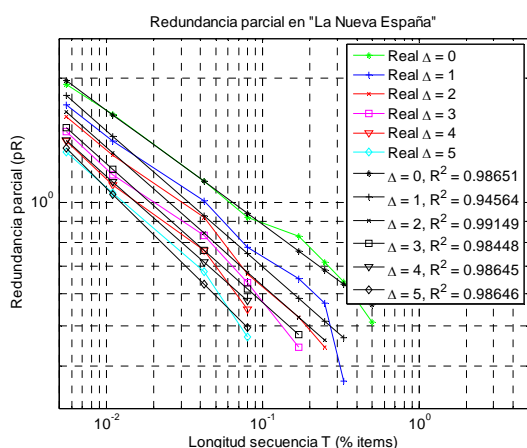


Figura 52. pR en "La Nueva España"

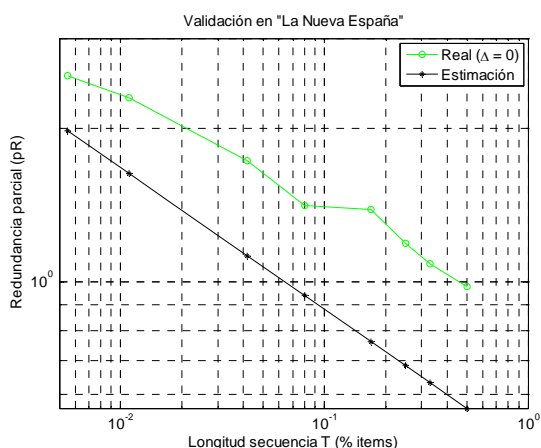


Figura 53. Validación con $\Delta = 0$.

Por lo tanto, es un hecho que el número de peticiones de un día al siguiente o de un periódico a otro sufre fluctuaciones, las cuales se pueden agudizar con la aparición de picos o carencia de peticiones en ciertos periodos de tiempo (vacaciones, fines de semana, días laborales, etc). Con el fin de alcanzar una primera aproximación en la eliminación del factor tiempo y disminuir la vulnerabilidad de las fluctuaciones en las peticiones, se ha considerado una distribución uniforme de las peticiones diarias N ($\#Peticiones/365$) en los seis periódicos digitales. En consecuencia, en cada periódico se ha considerado un análisis homogéneo. Por otro lado, hay que tener en cuenta que estos periódicos tienen una incorporación diaria de nuevo contenido, y por lo tanto sería deseable hacer predicciones sobre el porcentaje de peticiones entre intervalos próximos en el tiempo. La secuencia de peticiones de los vídeos, en el periodo comprendido desde el 1 Enero hasta el 31 de Diciembre de 2010, ha sido dividida en bloques del mismo tamaño B ($2*N$) en cada periódico digital. En la Tabla 37 se muestra el tamaño del bloque B que ha sido elegido para cada uno de los seis periódicos.

Block size	Periódicos digitales					
	"La Opinión A Coruña"	"Faro de Vigo"	"Levante-EMV"	"La Provincia"	"La Nueva España"	"Superdeporte"
B	162	264	534	1.160	2.244	2.292

Tabla 37. Tamaño B en los periódicos digitales.

De acuerdo a la división por bloques de la secuencia de peticiones, estos son numerados de uno en adelante. Por un lado, el conjuntos de bloques impares serán usados para calcular un modelo de predicción del porcentaje de vídeos que serán solicitados de nuevo en un futuro cercano, mientras que el conjunto de bloques pares serán usados para la validación de dicho modelo.

Desde el punto de vista de la redundancia global, esta fue calculada en cada uno de los bloques B de manera individual. A partir del valor gR de cada uno de los bloques impares se calcula una gR media, que representará la gR media global del modelo. De forma análoga, si se tiene en cuenta el conjunto de bloques pares, se calcula gR media global para la parte que será usada para la validación. Si comparamos los datos de la Tabla 36 con los de la Tabla 38 se puede observar que los dos periódicos digitales con el menor número de peticiones por bloque, como son "La Opinión A Coruña" y "Faro de Vigo", alcanzaron la redundancia global más alta. Por el contrario, el periódico digital con el mayor número de elementos por bloque, como es "Superdeporte", alcanzó la redundancia global más baja. En términos probabilísticos, dado que el estudio se basa en el valor medio de gR, en bloques pequeños el número de vídeos que se repiten en cada bloque B es menor, mientras que a medida que el bloque aumenta su tamaño el número de repeticiones aumenta.

Periódicos digitales						
gR	"La Opinión A Coruña"	"Faro de Vigo"	"Levante-EMV"	"La Provincia"	"La Nueva España"	"Superdeporte"
Bloques impares	11,2856 (s.d. 13,7464)	1,6889 (s.d. 10,1124)	8,1461 (s.d. 7,6353)	4,7263 (s.d. 3,7878)	4,9519 (s.d. 3,4169)	2,5368 (s.d. 1,3655)
Bloques pares	11,1247 (s.d. 13,3415)	11,476 (s.d. 10,3493)	7,5236 (s.d. 7,241)	4,7046 (s.d. 3,8428)	4,9353 (s.d. 3,5705)	2,785 (s.d. 1,62)

Tabla 38. Resultados de gR.

Además, con el fin de estudiar en profundidad la consistencia temporal en las secuencias de peticiones, se ha analizado secuencias con distinta longitud T dentro del bloque B, tal y como se ha explicado en el algoritmo de la Tabla 34. Por otro lado, la longitud de la secuencia T representa un porcentaje sobre el total de peticiones del bloque B, y no un valor absoluto. En nuestro análisis han sido considerados valores para T pertenecientes al siguiente intervalo [1%, 2%, ..., 50%]. En la Fig. 54 se muestran los resultados obtenidos para pR en los seis periódicos digitales para Δ igual a 0 y con

distintas longitudes de secuencia para T. El eje de abscisas representa la longitud de la secuencia $T \in [1\%, 2\%, \dots, 50\%]$, y el eje de ordenadas representa la media de pR obtenida a partir del valor de pR de los bloques impares, donde ambos ejes están expresados en escala logarítmica.

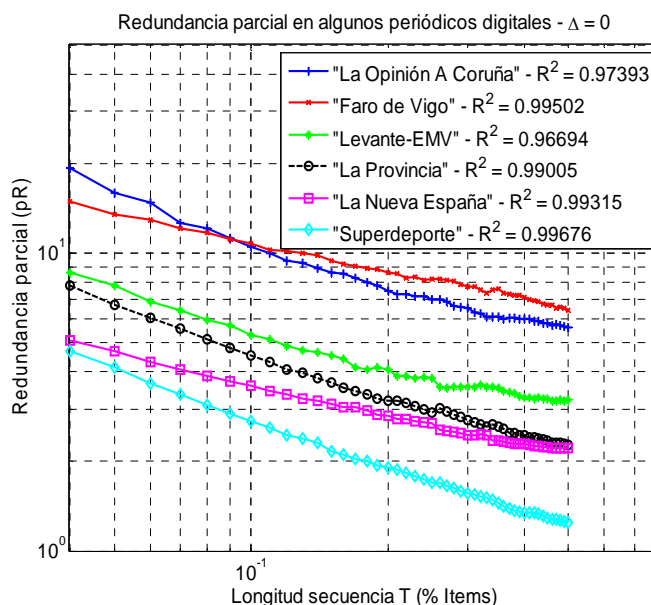


Figura 54. pR con $\Delta = 0$.

En la Fig. 54 se observa que el valor de la redundancia parcial decrece linealmente a medida que se incrementa el tamaño de la secuencia, donde el coeficiente de regresión lineal resultó ser excelente en todos los casos, con un valor superior o igual a 0,96. Este comportamiento lineal parece tener lógica. De hecho, cuando el tamaño de la secuencia es pequeña, las peticiones sobre los mismos vídeos están más cercanas unas de otras en el tiempo. Especialmente, si se trata de vídeos que han sido publicados recientemente. A medida que la longitud de la secuencia aumenta, las peticiones sobre el mismo vídeo son más distantes en el tiempo, y pR disminuye. Además, con la incorporación diaria, los nuevos vídeos siempre tendrán más probabilidad de ser seleccionados, junto con los correspondientes a noticias de impacto, que los publicados con anterioridad. Por lo tanto, dependiendo del tamaño de la secuencia T, pR se puede expresar de acuerdo a la ecuación lineal expresada en (38) en escala logarítmica.

$$\log(pR(T)) = -a * \log(T) + \log(b) \quad (38)$$

En escala lineal la ecuación (38) es equivalente a la ecuación (39).

$$pR(T) = b * T^{-a} \quad (39)$$

El valor de b es una constante para un ajuste de escala, el cual representa cuanto vale la redundancia parcial cuando el tamaño de la secuencia es de un 1% sobre el tamaño del bloque B , y a es la pendiente de la línea en la representación logarítmica. En la Tabla 39 están representados los resultados para cada uno de los periódicos digitales de acuerdo a la ecuación (38).

Parámetros	Periódicos digitales					
	"La Opinión A Coruña"	"Faro de Vigo"	"Levante-EMV"	"La Provincia"	"La Nueva España"	"Superdeporte"
a	-0,4454	-0,3319	-0,4382	-0,5041	-0,3680	-0,5471
log(b)	1,3568	1,6388	0,8752	-0,4059	-0,4713	-0,2121
R²	0,9739	0,9950	0,9669	0,9900	0,9931	0,9968
Error Standard	0,0488	0,0259	0,0805	0,0492	0,0437	0,0365

Tabla 39. Parámetros específicos para cada periódico digital.

Tal y como ha sido explicado en el apartado 7.2., la distancia (Δ) entre pares de secuencias (Q_i, Q_j) será otro parámetro que es necesario tener en cuenta en el cálculo de la redundancia parcial. A medida que la distancia entre pares de secuencias aumenta, el número de vídeos que tiene en común Q_i con Q_j disminuya. En la Fig. 55 se puede observar los resultados que fueron obtenidos para la pR en el periódico "La Nueva España" con $\Delta \in [0, 1, 2, 3, 5, 10]$ en escala logarítmica. Por un lado, estos resultados reflejan que cuando dos secuencias estaban cercanas en el tiempo la pR era alta. De hecho, el máximo valor de pR se alcanza cuando no existe distancia entre dos secuencias ($\Delta = 0$). Por otro lado, para un valor específico de T , pR decrece a medida que Δ aumenta. De hecho, la diferencia entre la pendiente de la recta del modelo lineal asociado a un valor determinado Δ con la pendiente del modelo lineal asociado a $(\Delta+1)$ decrece de forma constante. Este mismo fenómeno ocurre con el valor de origen entre los distintos modelos lineales. Este hecho puede ser observado en la Fig. 56, donde el conjunto de valores de las distintas pendientes y valores en el origen asociados a los distintos modelos lineales de la Fig. 55, decaen de forma constante como se ha

comentado, y donde cada uno de dichos parámetros sigue un modelo lineal con un coeficiente de regresión superior a 0,97 en ambos casos.

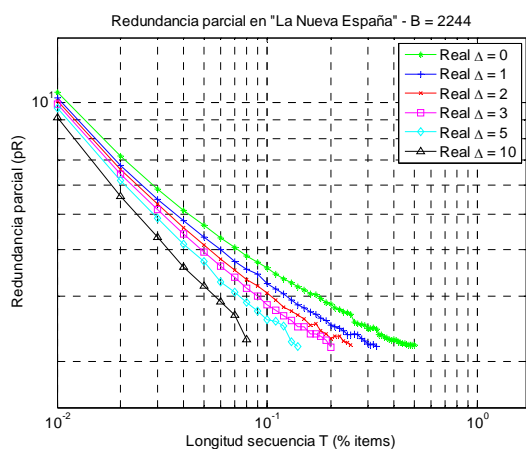


Figura 55. pR en "La Nueva España".

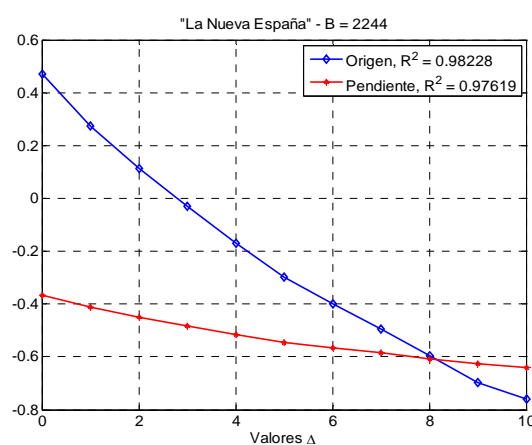


Figura 56. Valores pendiente y origen.

En definitiva, el parámetro delta (Δ) debería formar parte de la ecuación (39), y la expresión final se correspondería con la ecuación (40).

$$pR(T, \Delta) = (b - \lambda\Delta) * T^{-(a+\beta\Delta)} \quad (40)$$

El valor inicial para Δ es siempre 0, pero su último valor estará asociado con el último modelo válido ($p_value > 0,05$). Este valor varía de un periódico digital a otro. En la Tabla 40 se muestran los distintos valores de γ , β y el conjunto de valores válidos para Δ en cada periódico digital.

Parámetro	Periódicos digitales					
	"La Opinión A Coruña"	"Faro de Vigo"	"Levante-EMV"	"La Provincia"	"La Nueva España"	"Superdeporte"
Valores Δ	[0, 5]	[0, 14]	[0, 5]	[0, 10]	[0, 10]	[0, 10]
γ	-0,1103 ($R^2 = 0,95$)	-0,0780 ($R^2 = 0,99$)	-0,1711 ($R^2 = 0,97$)	-0,0751 ($R^2 = 0,91$)	-0,1209 ($R^2 = 0,98$)	-0,1077 ($R^2 = 0,92$)
β	-0,0381 ($R^2 = 0,93$)	-0,0200 ($R^2 = 0,99$)	-0,0443 ($R^2 = 0,96$)	-0,0174 ($R^2 = 0,86$)	-0,0268 ($R^2 = 0,98$)	-0,0226 ($R^2 = 0,86$)

Tabla 40. Valores de Δ , γ y β en los seis periódicos digitales

En las Figs. 57, 58, 59, 60, 61 y 62 ha sido representado el resultado del ajuste entre los datos reales con los modelos obtenidos en cada periódico digital. En todos los casos el coeficiente de regresión resultó ser igual o superior 0,95. El modelo confirma

con una alta probabilidad que para un valor específico de Δ , el valor de pR disminuye a medida que T aumenta. Este hecho es cierto en términos probabilísticos, dado que el estudio se ha basado en el valor medio de pR.

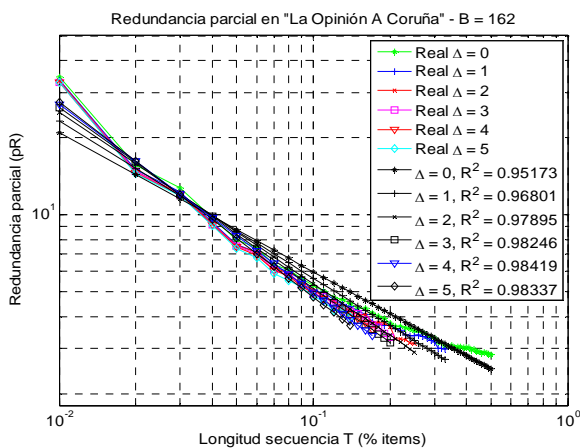


Figura 57. pR "La Opinión A Coruña", B=162.

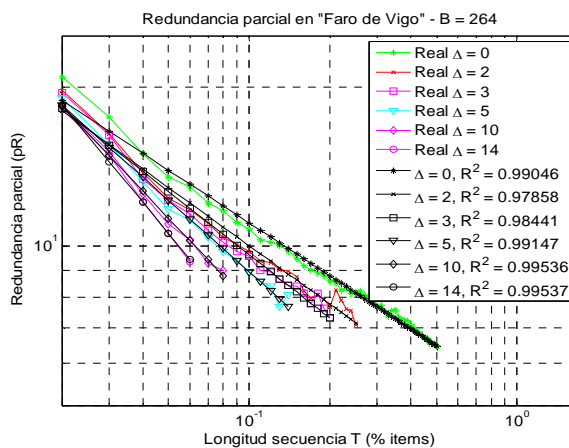


Figura 58. pR "Faro de Vigo", B=264.

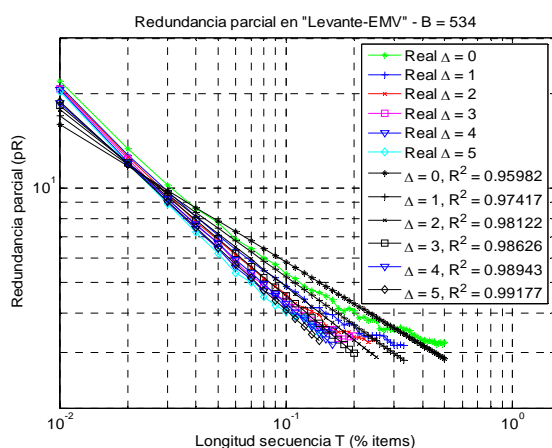


Figura 59. pR "Levante-EMV", B=534.

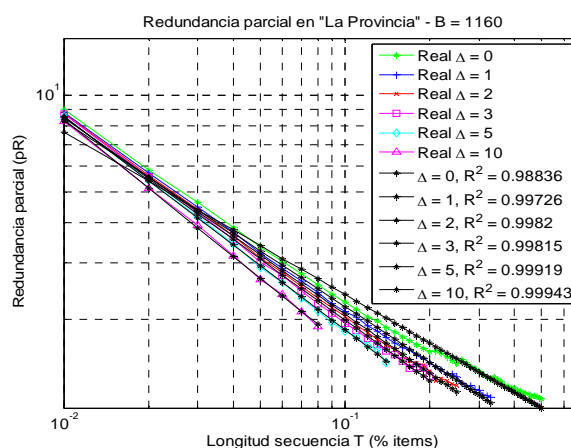


Figura 60. pR "La Provincia", B=1.160.

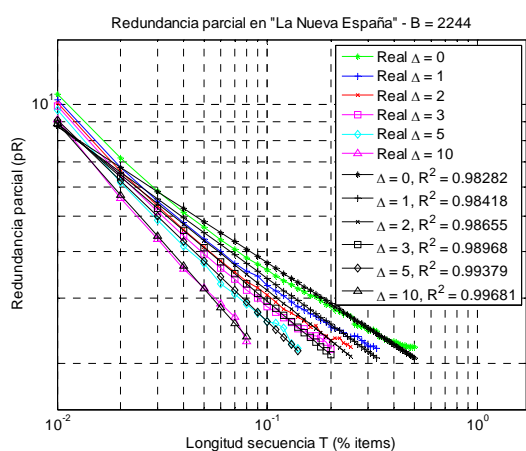


Figura 61. pR "La Nueva España", B=2.244.

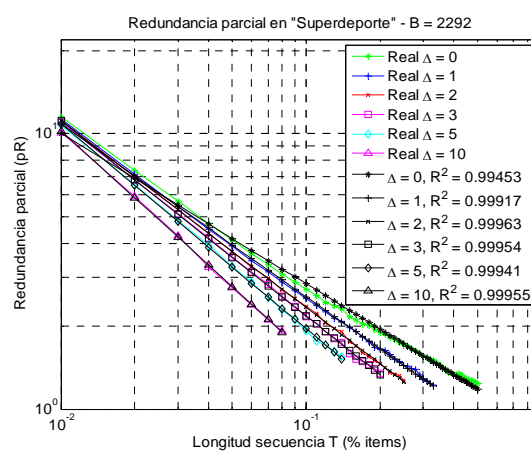


Figura 62. pR "Superdeporte", B=2.292.

7.4. Validación

Tal y como se mencionó en el apartado 7.2., los bloques pares han sido usados para el proceso de validación. Los diferentes modelos que han sido propuestos para cada uno de los periódicos digitales en el apartado 7.3. consiguen un alto grado de ajuste con los datos reales. De hecho, todos los modelos pasaron con éxito el test de contraste de hipótesis ($p_value > 0,05$), lo cual significa que no hubo diferencias significativas entre el modelo y los datos reales. La calidad del ajuste se puede observar en las Figs. 63, 64, 65, 66, 67 y 68 para cada uno de los diferentes periódicos. A veces se aprecian pequeñas diferencias entre el modelo y los datos reales que parecen mayores de lo que son debido a su representación en escala logarítmica.

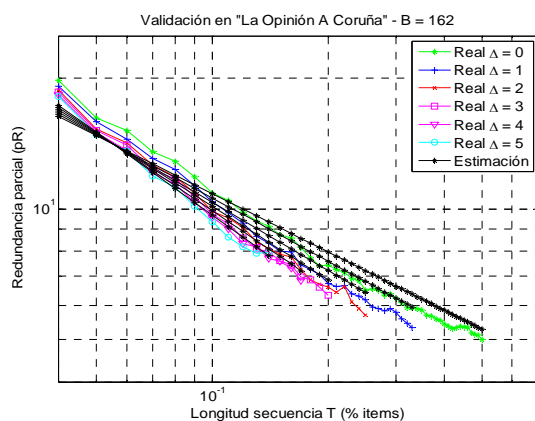


Figura 63. pR "La Opinión A Coruña", B=162.

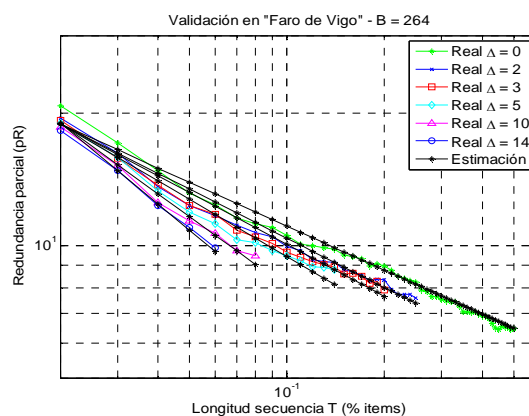


Figura 64. pR "Faro de Vigo", B=264.

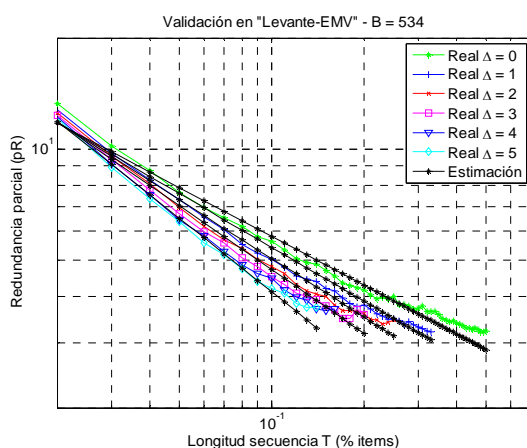


Figura 65. pR "Levante-EMV", B=534.

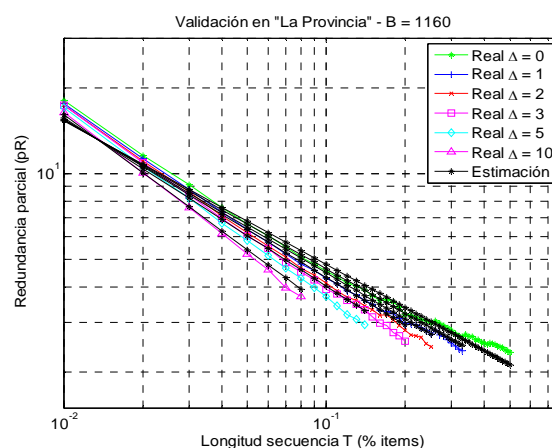


Figura 66. pR "La Provincia", B=1.160.

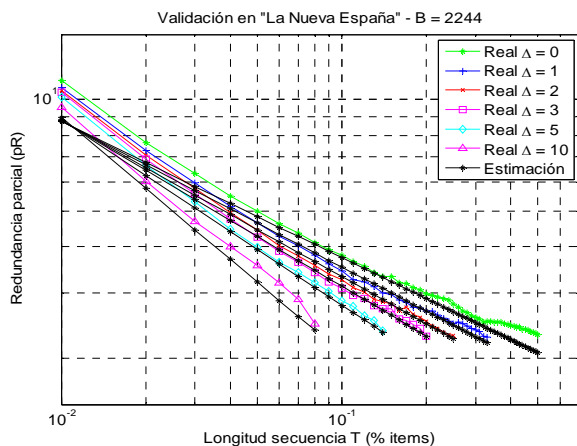


Figura 67. pR "La Nueva España", B=2.244.

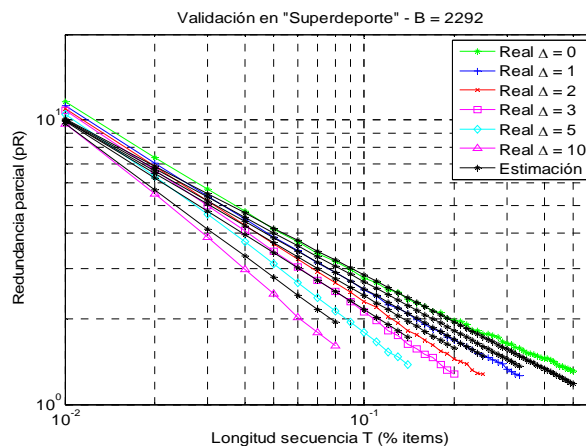


Figura 68. pR "Superdeporte", B=2.292.

Por ejemplo, si se desea estimar la pR de un secuencia de longitud T igual a 0,5 a la secuencia inmediatamente contigua ($pR = f(0,5; 0)$) a lo largo de los distintos periódicos digitales, los resultados que se conseguirían serían los siguientes: en "La Opinión A Coruña" el valor del modelo fue de 5,288 y el valor real fue de 5,006, en el "Faro de Vigo" el valor del modelo fue de 6,481 mientras que el valor real fue de 6,476, en "Levante-EMV" el valor del modelo fue de 2,872 y el valor real fue de 3,221, en "La Provincia" el modelo obtuvo un valor de 2,128 y el valor real fue de 2,355, en "La Nueva España" el modelo tuvo un valor de 2,067 y en el valor real de 2,315, y finalmente en "Superdeporte" el valor del modelo fue de 1,182 y el valor real de 1,31. Por lo tanto, se detectaron pequeñas diferencias entre el modelo y los datos reales, donde la mínima diferencia fue de 0,005 en "Faro de Vigo" y la máxima fue de 0,282 en "La Opinión a Coruña", siendo "Levante-EMV" el periódico con una diferencia ligeramente superior a los demás con un valor de 0,349.

En definitiva, los resultados de la validación determinan que la redundancia más baja está en el periódico "Superdeporte" con un 1% y un nivel de exactitud del 89%, siendo el valor más alto en el "Faro de Vigo" con un 6% y un nivel de exactitud del 100%. Estos resultados resaltan que en periódicos donde el número de vídeos es menor, como es el caso de "La Opinión A Coruña" o "Faro de Vigo", los vídeos tienen menos competencia unos con otros y tienen más probabilidad de ser seleccionados con más frecuencia. Considerando algunas de las características de un sistema de distribución de vídeo, parece conveniente la combinación entre una política de caching y otra de copia

de contenido en los lugares más próximos al usuario, con el fin de reducir el tiempo de latencia [4]. En consecuencia, en periódicos digitales con un elevado número de vídeos sería más adecuado el uso de políticas de caching, y en el caso contrario sería más adecuado las políticas de copia.

7.5. Conclusiones

En este capítulo ha sido analizado el grado de redundancia presente en las distintas secuencias de peticiones registradas en los seis periódicos digitales. Dado que el número de peticiones de un día al siguiente y de un periódico a otro está sujeto a variaciones, se ha considerado una distribución uniforme diaria de las mismas.

La presencia de la regularidad ha sido estudiada con un método de redundancia global y otro parcial. En primer lugar, los resultados de la redundancia global muestran que la presencia de peticiones redundantes disminuye a medida que la longitud del conjunto de peticiones en estudio es menor. En segundo lugar, la redundancia parcial ha sido estudiada entre pares de secuencias de peticiones de la misma longitud T , bien contiguas o con una cierta separación entre las mismas (Δ). A partir de los parámetros T y Δ se ha conseguido establecer una función $f(\Delta, T)$ que resultó ser independiente tanto del usuario como del tiempo. Lógicamente, el máximo valor de la función se obtiene cuando no hay separación entre los pares secuencias ($\Delta = 0$). Gracias a este modelo se podrían realizar predicciones sobre el grado de redundancia de las peticiones en un futuro cercano.

Además, en los resultados de la validación se consiguió predecir entre un 1% y un 6% de la redundancia con un nivel de seguridad que osciló entre el 88% y el 100%. Por lo tanto, se puede decir que los diferentes modelos obtenidos para la redundancia parcial proporcionan excelentes resultados.

Capítulo 8.
CONCLUSIONES Y LINEAS DE
TRABAJO FUTURAS

8.1. Introducción

Al final de cada capítulo se han mencionado las aportaciones generales de cada uno de los temas tratados. En este capítulo final se realizará un resumen de las principales aportaciones, así como de las líneas de trabajo futuras.

8.2. Conclusiones

La tesis se inició con el estudio de la popularidad de los vídeos en uno de los primeros diarios digitales de noticias en España, Asturias.com, el cual ofrecía información en lengua asturiana. La frecuencia de incorporación de nuevos vídeos aproximadamente era de uno por semana, característica que lo hacía singular, pues el hecho de introducir contenidos de forma tan espaciada en el tiempo hacía que vídeos con una gran demanda siguiesen siendo solicitados pese a la introducción de nuevos contenidos. Esta situación no se daría en otros periódicos digitales, en los que la información es introducida diariamente, haciendo que vídeos anteriores pasen a ser desapercibidos con rapidez. En su estudio se pudo comprobar que el valor del parámetro θ de la distribución Zipf-like es un buen estimador de la popularidad del servicio en periodos de estabilidad del contenido.

A continuación, se realizó un estudio sobre periódicos digitales de ámbito autonómico donde la inserción de contenido era diaria. En concreto, en el estudio fueron elegidos los seis periódicos digitales con el mayor número de accesos. A pesar de su naturaleza regional publican bastantes vídeos relativos a noticias con una temática variada, los cuales son presentados al usuario clasificados por categorías en función de su temática. Por lo tanto, con el objetivo de establecer un marco común de referencia entre los distintos periódicos se buscaron las categorías comunes y se establecieron diez secciones (nacional, internacional, autonómica, etc).

Tanto desde el punto de vista del número de vídeos presentes por sección, como desde el punto de vista del número de accesos, la sección correspondiente al ámbito autonómico resultó ser la más destacada, hecho que resalta más el carácter regional de este tipo de periódicos. En el caso específico de los deportes, el periódico "Superdeporte" sobresale con bastante superioridad respecto a la sección "Deportes" del

resto de los periódicos. Por otro lado, en función del periódico unas secciones publican más vídeos nuevos que otras, siendo la sección "Nacional", "Autonómica" y "Sociedad" las más destacadas.

En esta tesis se ha realizado la caracterización de la popularidad de los vídeos publicados en seis periódicos digitales de ámbito autonómico. El conocimiento del patrón de demanda es uno de los factores a tener en cuenta a la hora de realizar predicciones sobre las posibles necesidades futuras de recursos, con el fin de conseguir un sistema de distribución de contenido eficaz y/o eficiente.

Las aportaciones de esta tesis se pueden sintetizar en los siguientes apartados:

- ✓ Dado que las inserciones se realizan diariamente, el estudio se ha realizado día a día, dado que es el periodo de tiempo donde el contenido del periódico se mantiene estable.
- ✓ Se establecieron modelos generales para la caracterización de la popularidad en cada uno de los seis periódicos digitales y para cada una de las distribuciones (Zipf-like, two Zipf-like, Mandelbrot, Stretched y Box-Cox) siempre que fue posible.
- ✓ Se establecieron modelos generales asociados a cada día de la semana.
- ✓ Si además lo que se pretende es modelar el comportamiento de la demanda por secciones, en la mayoría de las ocasiones sólo fue posible plantear un modelo para la sección S3 relativa al ámbito autonómico.
- ✓ Aunque esta tesis está centrada en la caracterización de la popularidad sobre la demanda diaria de vídeos, también se ha realizado un breve análisis sobre el comportamiento de la popularidad en periodos con una longitud superior. Lógicamente, a medida que la longitud del periodo crece la presencia de vídeos con una popularidad alta y/o similar debería de aumentar. Este efecto se ve reflejado en la variación de los parámetros de los modelos, tal y como es corroborado en [61].
- ✓ Desde el punto de vista de la concentración de los accesos se ha comprobado que en cuatro periódicos las peticiones se concentran sobre unos pocos vídeos,

tal y como indica el Principio de Pareto, y donde el resto de las peticiones se encuentran más dispersas, lo cual puede ser debido al mayor número de vídeos presentes en este tipo de sistemas.

- ✓ En los resultados de la validación, según el test de K-S, la transformación de Box-Cox fue la que presentó los mejores resultados de cobertura o bien con una mínima diferencia respecto a la distribución Stretched o Zipf-like, tanto desde el punto de vista del modelo general como en los modelos correspondientes a cada día de la semana. Sin embargo, como contrapartida en el periódico "Levante-EMV" no fue posible establecer un modelo general, y respecto a los modelos diarios tuvo un comportamiento peor que Stretched o Zipf-like.
- ✓ Desde el punto de vista de las secciones, la distribución Mandelbrot sería la distribución predominante. Sin embargo, desde el punto de vista de la sección S3 es mejor una distribución Zipf-like, Stretched o Mandelbrot.
- ✓ Se ha conseguido establecer un modelo de redundancia parcial para cada uno de los periódicos digitales, el cual permitiría predecir la proporción de redundancia de peticiones de vídeos de un periodo de tamaño T en otro periodo del mismo tamaño T' ($T=T'$) posterior en el tiempo. Los modelos predicen una redundancia entre un 1% y un 6% con un nivel de seguridad comprendido entre el 88% y el 100%. Este modelo ha resultado ser independiente del usuario y del tiempo.
- ✓ A partir de los resultados de la redundancia parcial y considerando algunas de las características de un sistema de distribución de vídeo, parece adecuada la combinación entre una política de caching y otra de copia de contenido en los lugares más próximos al usuario, con el fin de reducir el tiempo de latencia [4]. En consecuencia, en periódicos digitales con un elevado número de vídeos sería más adecuado el uso de políticas de caching, y en el caso contrario sería más adecuado las políticas de copia.

8.3. Trabajos futuros

En esta tesis se ha puesto de manifiesto que un servicio de NoD presenta diferencias significativas frente a un servicio genérico de VoD, tales como la duración

de los vídeos, vídeos vinculados a noticias de actualidad, inserción diaria de noticias, etc). El establecimiento de la caracterización de la popularidad de los distintos periódicos digitales se hace necesaria para garantizar una buena calidad de acceso al servicio de los usuarios. Sin embargo, todavía quedan aspectos por cubrir y afianzar, los cuales pasan a ser numerados a continuación:

- ✓ Estudiar la viabilidad de la "transformación de Box-Cox" en un mayor número de periódicos digitales.
- ✓ Analizar en distintos escenarios el comportamiento de los parámetros que caracterizan la calidad de un servicio de VoD, a partir de los modelos que caracterizan la popularidad en los distintos periódicos.
- ✓ Analizar periodos de tiempo mayores, con el fin de obtener modelos más fiables para la caracterización de la popularidad.
- ✓ Dado que no todos los periódicos son igual de demandados y no todas las secciones tienen la misma popularidad, se debería de analizar posibles algoritmos de distribución de vídeos, de forma que permita al administrador de la red mejorar la planificación de los recursos necesarios, así como la distribución de contenido.
- ✓ Aplicar técnicas de clustering, con el fin de detectar los grupos de noticias que resultan ser de interés para los usuarios cada día. Este tipo de estudios podría ayudar a seleccionar de forma automática el conjunto de noticias [71] que podrían presentarse en portada al día siguiente, o incluso los vídeos publicitarios más adecuados.

8.4. Publicaciones derivadas de esta tesis

- [1] M. G. Aparicio, et al., "Measuring temporal redundancy in sequences of video requests in a News-on-Demand service", *Journal of Computer Communications*, 2012, con un JCR de 0,815 y categoría C, de acuerdo al último ranking de CORE "Computing Research & Education" de Febrero de 2010. En revisión.

- [2] M. G. Aparicio, *et al.*, "A wider view of video popularity characterization centered on News-on-Demand", *Journal of Computer Networks*, 2012, con un JCR de 1,176 y categoría A, de acuerdo al último ranking de CORE "Computing Research & Education" de Febrero de 2010. En revisión.
- [3] M. G. Aparicio, *et al.*, "Box-Cox transformation as an alternative method for modeling video-on-demand popularity," en *Globecom: Ubiquitous Computing and Networks*, Miami, Florida, 2010, pp. 1798-1802. Indexado en Google Scholar y en IEEEExplore digital library con ISBN 978-1-4244-8863-6.
- [4] M. G. Aparicio, *et al.*, "Influencia de la incorporación de nuevos contenidos en la popularidad de servicios de vídeo bajo demanda," en *VII Jornadas de Ingeniería Telemática* Alcalá de Henares, 2008, pp. 201-208.

8.5. Otras publicaciones

- [1] M. G. Aparicio, *et al.*, "Find the information you want on the web, how likely could be get it successfully?", en el Congreso Internacional IADIS e-Society 2006, Dublin, Irlanda, 2006, pp. 275-279.
- [2] M. G. Aparicio, *et al.*, "A first approximation to estimate the usability of a software application task", en el Congreso Internacional IADIS WWW/Internet 2004, Madrid, España, 2004, pp. 969-972.
- [3] M. G. Aparicio, *and Javier Tuya*, "Verificación formal de un servicio informatizado para la gestión de impagados de un sector", en el Congreso Nacional SIT'2001, A Coruña, España, 2001, pp. 237-246.
- [4] M. G. Aparicio *et al.*, "Inserción de la Programación por Contrato en Ficheros Objeto de la JVM (.class)", en el Congreso Nacional seid' 2000, Ourense, España, pp. 551-559.

Capítulo 9.

REFERENCIAS

9.1. Referencias

- [1] T. Wauters, *et al.*, "Optical network design for video on demand services," in *9th Conference on optical network design and modelling*, Milan, Italy, 2005, pp. 251-260.
- [2] X. Zhou and C. Xu, "Optimal video replication and placement on a cluster of video-on-demand servers," in *Parallel Processing* Vancouver, Canada, 2002, pp. 547-555.
- [3] J. Merwe, *et al.*, "Streaming Video Traffic: Characterization and Network Impact," in *Web Content Caching and Distribution*, Boulder, Colorado, 2002.
- [4] S. Bakiras and T. Loukopoulos, "Combining replica placement and caching techniques in content distribution networks," *Computer Communications*, vol. 28, pp. 1062-1073, 2005.
- [5] T. Palpanas and A. Mendelzon, "Web Prefetching Using Partial Match Prediction," in *Web Caching Workshop*, San Diego, CA, 1999, pp. 1-21.
- [6] S. Jin and A. Bestavros, "Generating Internet Streaming Media Objects and Workloads," *Web Content Delivery*, pp. 177-196, 2005.
- [7] W. Tang, *et al.*, "Medisyn: A synthetic streaming media service workload generator," 2003, pp. 12-21.
- [8] P. Barford and M. Crovella, "Generating representative web workloads for network and server performance evaluation," in *International Conference on Measurement and Modeling of Computer Systems*, 1998, pp. 151-160.
- [9] R. García, *et al.*, "Probabilistic analysis and interdependence discovery in the user interactions of a video news on demand service," *Computer Networks*, vol. 53, pp. 2038-2049, 2009.
- [10] C. Griwodz, *et al.*, "Long-term movie popularity models in video-on-demand systems: or the life of an on-demand movie," in *Multimedia*, Seattle, USA, 1997, pp. 349-357.
- [11] K. P. Gummadi, *et al.*, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," presented at the Proceedings of the nineteenth ACM symposium on Operating systems principles, Bolton Landing, NY, USA, 2003.
- [12] M. Cha, *et al.*, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," in *Internet Measurement Conference*, San Diego, California, 2007, pp. 1-14.
- [13] S. Krug, *No me hagas pensar. Una aproximación a la usabilidad en la Web.*, Segunda ed.: Pearson Education, S. A., 2006.
- [14] J. Nielsen, *Usabilidad. Diseño de sitios Web.*: Alhambra Editorial, 2000.
- [15] L. Cherkasova and M. Gupta, "Analysis of enterprise media server workloads: access patterns, locality, content evolution, and rates of change," *IEEE/ACM Transactions on Networking (TON)*, vol. 12, pp. 781-794, 2004.
- [16] X. Kang, *et al.*, "Measurement, modeling, and analysis of internet video sharing site workload: A case study," in *International Conference on Web Services*, 2008, pp. 278-285.
- [17] X. Kang, *et al.*, "Understanding internet video sharing site workload: a view from data center design," *Journal of Visual Communication and Image Representation*, vol. 21, pp. 129-138, 2010.
- [18] J. Ni, *et al.*, "Hierarchical content routing in large-scale multimedia content delivery network," 2003, pp. 854-859.

- [19] F. Thouin and M. Coates, "Video-on-demand networks: design approaches and future challenges," *IEEE Network*, vol. 21, pp. 42-48, 2007.
- [20] J. M. Almeida, *et al.*, "Analysis of educational media server workloads," presented at the Proceedings of the 11th international workshop on Network and operating systems support for digital audio and video, Port Jefferson, New York, United States, 2001.
- [21] A. Dan, *et al.*, "Scheduling policies for an on-demand video server with batching," in *Multimedia'94*, 1994, pp. 15-23.
- [22] K. A. Hua and S. Sheu, "Skyscraper Broadcasting: A New Broadcasting Scheme for Metropolitan Video-on-Demand Systems," in *SIGCOMM*, Cannes, France, 1997, pp. 89-100.
- [23] J. Segarra and V. Cholvi, "Distribution of Video-on-Demand in Residential Networks," presented at the Proceedings of the 8th International Workshop on Interactive Distributed Multimedia Systems, 2001.
- [24] Z. Ge, *et al.*, "Design and Analysis of A Demand Adaptive and Locality Aware Streaming Media Server Cluster," in *Multimedia Systems* vol. 13, September ed: ACM/Springer, 2007, pp. 235-249.
- [25] Z. G. K., Ed., *Human Behavior and the Principle of Least-Effort*. Cambridge, MA: Addison Wesley, 1949.
- [26] A. Bestavros, "WWW Traffic Reduction and Load Balancing through Server-Based Caching," *Parallel and Distributed Technology*, vol. 5, pp. 56-67, 1997.
- [27] S. Glassman, "A caching relay for the World Wide Web," *Comput. Netw. ISDN Syst.*, vol. 27, pp. 165-173, 1994.
- [28] C. Almeida, *et al.*, "Characterizing reference locality in the WWW," in *Parallel and Distributed Information Systems*, Miami Beach, Florida, 1996, pp. 92-103.
- [29] L. Breslau, *et al.*, "Web caching and Zipf-like distributions: Evidence and implications," in *INFOCOM*, 1999, pp. 126-134.
- [30] S. Jin and A. Bestavros, "Sources and Characteristics of Web Temporal Locality," in *Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, San Francisco, CA, 2000, pp. 28-35.
- [31] C. Cunha, *et al.*, "Characteristics of WWW client-based traces," 1995.
- [32] R. P. Doyle, *et al.*, "The Trickle-Down Effect: Web Caching and Server Request Distribution," *Computer Communications*, vol. 25, pp. 345-356, June 2002.
- [33] B. Wang, *et al.*, "Optimal Proxy Cache Allocation for Efficient Streaming Media Distribution," in *INFOCOM*, New York, USA, 2002, pp. 1726-1735.
- [34] E. Balafoutis and I. Stavrakakis, "Proxy Caching and Video Segmentation Based on Request Frequencies and Access Costs," in *ICT*, Tahiti, French Polynesia, 2003, pp. 1367-1372.
- [35] NUSSBAUMER, *et al.*, *Networking requirements for interactive video on demand* vol. 13. New York, NY, ETATS-UNIS: Institute of Electrical and Electronics Engineers, 1995.
- [36] S. Acharya, *et al.*, "Characterizing user access to videos on the world wide web," in *Multimedia Computing and Networking* San Jose, CA, 2000, pp. 130-141.
- [37] M. Chesire, *et al.*, "Measurement and analysis of a streaming-media workload," in *Symposium on Internet Technologies and Systems*, 2001, pp. 1-12.
- [38] L. Cherkasova and M. Gupta, "Characterizing locality, evolution, and life span of accesses in enterprise media server workloads," in *NOSSDAV'02*, Miami, Florida, USA, 2002, pp. 33-42.

- [39] C. P. Costa, *et al.*, "Analyzing client interactivity in streaming media," presented at the Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, 2004.
- [40] H. Yu, *et al.*, "Understanding user behavior in large-scale video-on-demand systems," *SIGOPS Oper. Syst. Rev.*, vol. 40, pp. 333-344, 2006.
- [41] J. Almeida, *et al.*, "A hybrid caching strategy for streaming media files," in *Multimedia Computing and Networking*, San Jose, 2001.
- [42] S. González, *et al.*, "A case study of load sharing based on popularity in distributed VoD systems," *IEEE transactions on Multimedia*, vol. 8, pp. 1299-1304, 2006.
- [43] J. Wang and P. Yu, "Fragmental proxy caching for streaming multimedia objects," *IEEE transactions on Multimedia*, vol. 9, pp. 147-156, 2007.
- [44] P. Gill, *et al.*, "Youtube traffic characterization: a view from the edge," presented at the Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, San Diego, California, USA, 2007.
- [45] X. Cheng, *et al.*, "Understanding the characteristics of internet short video sharing: YouTube as a case study," *ArXiv e-prints*, vol. 707, 2008.
- [46] C. Huang, *et al.*, "Can internet video-on-demand be profitable?," presented at the Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications, Kyoto, Japan, 2007.
- [47] E. Veloso, *et al.*, "A hierarchical characterization of a live streaming media workload," *Transactions on Networking*, vol. 14, pp. 133-146, 2006.
- [48] F. T. Johnsen, *et al.*, "Analysis of Server Workload and Client Interactions in a News-on-Demand Streaming System," presented at the International Symposium on Multimedia, San Diego, California, 2006.
- [49] F. T. Johnsen, *et al.*, "Workload characterization for news-on-demand streaming services," in *Performance, Computing, and Communications Conference*, New Orleans, LA, 2007, pp. 314-323.
- [50] E. Nilsen, "Analysis of News-On-Demand Characteristics and Client Access Patterns," Informatics, University of Oslo, Oslo, 2005.
- [51] X. Pañeda, *et al.*, "Popularity analysis of a video-on-demand service in a digital newspaper: influence of the subject, video characteristics and new content publication policy," *Advanced. Media and Communication*, vol. 1, pp. 369-385, 2007.
- [52] Y. J. Kim, *et al.*, "Clustered Multimedia NOD: Popularity-Based Article Prefetching and Placement," in *Mass Storage Systems*, San Diego, CA, USA, 1999, pp. 194-202.
- [53] W. Tang, *et al.*, "Long-term streaming media server workload analysis and modeling," Citeseer2003.
- [54] W. Tang, *et al.*, "Modeling and generating realistic streaming media server workloads," *Computer Networks*, vol. 51, pp. 336-356, 2007.
- [55] S. Chen, *et al.*, "SProxy: A caching infrastructure to support internet streaming," *IEEE transactions on Multimedia*, vol. 9, pp. 1062-1072, 2007.
- [56] T. Yu, *et al.*, "Internal Popularity of Streaming Video and its Implication on Caching " in *Advanced Information Networking and Applications*, Vienna, Austria, 2006.
- [57] K. Sripanidkulchai, *et al.*, "An analysis of live streaming workloads on the internet," presented at the Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, Taormina, Sicily, Italy, 2004.

- [58] C. K. d. S. Rodrigues and R. M. M. Leao, "Bandwidth usage distribution of multimedia servers using Patching," *Computer Networks* vol. 51, pp. 569-587, 2007.
- [59] O. Saleh and M. Hefeeda, "Measurement and caching of peer-to-peer traffic," in *ICNP*, Santa Barbara, California, 2006, pp. 249-258.
- [60] L. Guo, *et al.*, "Does Internet media traffic really follow Zipf-like distribution?," 2007, p. 360.
- [61] L. Guo, *et al.*, "The stretched exponential distribution of internet media access patterns," presented at the Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing, Toronto, Canada, 2008.
- [62] L. Guo, *et al.*, "Analyzing patterns of user content generation in online social networks," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.
- [63] R. Fonseca, *et al.*, "On the Intrinsic Locality Properties of Web Reference Streams," in *INFOCOM*, San Francisco, 2003, pp. 448-458.
- [64] M. F. Arlitt and C. L. Williamson, "Web Server Workload Characterization: The Search for Invariants," in *SIGMETRICS*, Philadelphia, PA, 1996, pp. 126-137.
- [65] F. Duarte, *et al.*, "Locality of Reference in an Hierarchy of Web Caches," in *Networking*, Coimbra, Portugal, 2006, pp. 344-354.
- [66] L. Lancieri and N. Durand, "Internet User Behavior: Compared Study of the Access Traces and Application to the Discovery of Communities," *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans*, vol. 36, pp. 208-219, January 2006.
- [67] N. Durand and L. Lancieri, "Study of the Regularity of the Users' Internet Accesses," in *Intelligent Data Engineering and Automated Learning*, Manchester, UK, 2002, pp. 173-178.
- [68] M. G. Aparicio, *et al.*, "Influencia de la incorporación de nuevos contenidos en la popularidad de servicios de vídeo bajo demanda," in *VII Jornadas de Ingeniería Telemática* Alcalá de Henares, 2008, pp. 201-208.
- [69] Z. Yang, "Predicting a Future Lifetime through Box-Cox Transformation," *Lifetime Data Analysis*, vol. 5, pp. 265-279, 1999.
- [70] F. Figueiredo, *et al.*, "The Tube over Time: Characterizing Popularity Growth of YouTube Videos," presented at the Web Search and Data Mining, Hong Kong, China, 2011.
- [71] Y.-L. Chen and T.-L. Yu, "News Classification based on experts' work knowledge," in *Networking and Information Technology*, Singapore, 2011, pp. 100-106.

ANEXO I.

DATOS ESPECÍFICOS DE LA DISTRIBUCIÓN ZIPF-LIKE

1. PERIÓDICO "La Opinión A Coruña"

1.1. Sin k-transformada

1.1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	26 / 26	23 / 26	21 / 26	20 / 26	23 / 26	23 / 26	22 / 26
Total días	26 + 23 + 21 + 20 + 23 + 23 + 22 = 158 (87%)						

Tabla 1. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Bartlett/Levene (Varianza)	F-Anova (Media)
θ	0,969	0,9974	0,05
R	< 0,05	0,7922	0,1636

Tabla 2. Test de normalidad, varianza y media para θ y R.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,5548	0,2126	0,9485	0,0223
Martes	1,6234	0,2056	0,9460	0,0184
Miércoles	1,6656	0,2375	0,9478	0,0195
Jueves	1,6483	0,2268	0,9363	0,0242
Viernes	1,5750	0,2198	0,9401	0,0186
Sábado	1,6483	0,2164	0,9487	0,0188
Domingo	1,4769	0,2212	0,9511	0,0159
Media	1,5971	0,2242	0,9457	0,02

Tabla 3. Media y desviación típica para θ y R por días de la semana.

1.1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/4	0	16/26	0	0/1	0	1/9	0/5	0/2	0/3
Martes	0/4	0	13/26	0	0	0	0/7	0/2	0	1/6
Miércoles	0/4	0	16/26	0/1	0/3	0/2	1/7	2/7	0	0/6
Jueves	1/4	0/1	9/26	0/1	0/1	0	0/9	1/6	0	0/3
Viernes	0/1	0	12/26	0	0/1	0	0/7	0/4	0/1	1/6
Sábado	0/3	0	12/26	0	0/2	0/1	1/6	0/7	0	0/4
Domingo	0/2	0	16/26	0	0	0	2/11	0/8	0/1	0/5
Total días	1	0	94	0	0	0	5	3	0	2

Tabla 4. N° de días modelados con éxito/sin éxito por días y sección.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Barlett/Levene (Varianza)	F-Anova (Media)
θ	0,127	0,1133	0,2047
R	< 0,05	0,7791	0,3877

Tabla 5. Test de normalidad, varianza y media para θ y R para la sección S3 en su totalidad.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,8041	0,2876	0,9576	0,0165
Martes	1,7247	0,2781	0,9513	0,0188
Miércoles	1,7725	0,2441	0,9432	0,0196
Jueves	1,8138	0,1016	0,9552	0,0101
Viernes	1,6107	0,2922	0,9460	0,0326
Sábado	1,8185	0,2121	0,9544	0,0141
Domingo	1,6316	0,2946	0,9542	0,0188
Media	1,7334	0,2656	0,9520	0,0198

Tabla 6. Media y desviación típica para θ y R por días de la semana para la sección S3.

1.2. Con k-transformada

1.2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	26 / 26	23 / 26	22 / 26	21 / 26	24 / 26	25 / 26	24 / 26
Total días	26 + 23 + 22 + 21 + 24 + 25 + 24 = 165 (91%)						

Tabla 7. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,3135	0,4607
kx	< 0,05	0,5917	0,5917
ky	< 0,05	0,6797	0,6797
R	< 0,05	0,8262	0,2202

Tabla 8. Test de normalidad, varianza y media para θ , kx, ky y R.

Días	θ		kx		ky	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	1,5973	0,4001	1,3464	0,7971	1,1154	0,3258
Martes	1,7494	0,2462	1,4348	0,5069	1,0000	0,0000
Miércoles	1,7408	0,3126	1,3182	0,6463	1,0454	0,2132
Jueves	1,9341	0,8603	1,8571	2,4142	1,0000	0,0000
Viernes	2,0601	1,9342	2,2500	3,6742	1,0833	0,2823
Sábado	1,8368	0,5048	1,6800	1,4353	1,1200	0,6000
Domingo	1,5942	0,3160	1,6667	0,9168	1,1250	0,3378
Media	1,7875	0,6535	1,65	1,4844	1,06988	0,2513

Tabla 9. Media y desviación típica para θ , kx y ky por días de la semana.

Días	R	
	Media	s.d.
Lunes	0,9527	0,0239
Martes	0,9383	0,0389
Miércoles	0,9349	0,0444
Jueves	0,9302	0,0336
Viernes	0,9404	0,0274
Sábado	0,9473	0,0298
Domingo	0,9468	0,0348
Media	0,9415	0,0332

Tabla 10. Media y desviación típica para R por días.

1.2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/5	0	17/26	0	0/1	0	3/9	1/5	0/2	1/3
Martes	0/4	0	14/26	0/1	0/1	0/1	0/7	1/3	0	1/7
Miércoles	0/5	0	17/26	0/1	0/3	0/2	1/9	2/7	0	1/7
Jueves	1/5	0/1	11/26	0/1	0/2	0	0/10	2/6	0	0/3
Viernes	0/1	0	14/26	0	0/1	0	2/7	0/5	0/1	2/6
Sábado	0/3	0	16/26	0	0/2	0/1	1/6	0/7	0	0/5
Domingo	1/2	0	18/26	0/1	0	0	3/12	0/8	0/1	1/6
Total días	2	0	107	0	0	0	10	6	0	6

Tabla 11. N° de días modelados con éxito/sin éxito por días y sección.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,4638	0,1711
kx	< 0,05	0,4019	0,298
ky	< 0,05	0,7317	0,7317
R	< 0,05	0,8262	0,2202

Tabla 12. Test de normalidad, varianza y media para θ , kx, ky y R para la sección S3.

Días	θ		kx		ky	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	1,8582	0,4452	1,2941	0,5879	1,1176	0,3321
Martes	1,7658	0,3836	1,2857	0,6112	1,1428	0,3631
Miércoles	1,9886	0,3656	1,5294	0,6243	1,0000	0,0000
Jueves	2,5244	1,5676	2,6364	4,1538	1,0000	0,0000
Viernes	2,0049	1,3561	1,8571	1,7033	1,2143	0,5789
Sábado	2,3126	0,9947	2,6250	3,2016	1,7500	3,0000
Domingo	1,7294	0,5488	1,5555	0,8556	1,7222	2,8244
Media	2,02628	0,8088	1,8262	1,6768	1,2781	1,0141

Tabla 13. Media y desviación típica para θ , kx y ky por días de la semana para la sección S3.

Días	R	
	Media	s.d.
Lunes	0,9611	0,0159
Martes	0,9546	0,0165
Miércoles	0,9446	0,02
Jueves	0,9567	0,0101
Viernes	0,9508	0,0315
Sábado	0,9623	0,0144
Domingo	0,9572	0,0166
Media	0,9553	0,0193

Tabla 14. Media y desviación típica de R por días en la sección S3.

2. PERIÓDICO "Faro de Vigo"

2.1. Sin k-transformada

2.1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	17 / 26	19 / 26	17 / 26	17 / 26	16 / 26	17 / 26	20 / 26
Total días	17 + 19 + 17 + 17 + 16 + 17 + 20 = 123 (68%)						

Tabla 15. Nº de días modelados con éxito/sin éxito por días de la semana.

Test estadísticos		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	< 0,05	< 0,05

Tabla 16. Test de normalidad, varianza y media de θ para todos los días.

	Test estadísticos - (L, M, X, J, V)		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,6736	0,498
R	< 0,05	< 0,05	0,2981

Tabla 17. Test de normalidad, varianza y media para θ y R de lunes a viernes.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,5296	0,3238	0,9523	0,022
Martes	1,5555	0,3815	0,9510	0,0175
Miércoles	1,6756	0,4263	0,9483	0,021
Jueves	1,6765	0,3674	0,9543	0,0129
Viernes	1,7129	0,3225	0,9610	0,0104
Media	1,6273	0,3659	0,9532	0,0175

Tabla 18. Media y desviación típica para θ y R de lunes a viernes.

	Test estadísticos - (S, D)		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,5535	0,6995
R	< 0,05	0,1155	0,05

Tabla 19. Test de normalidad, varianza y media de θ y R para el sábado y domingo.

Días	θ		R	
	Media	s.d.	Media	s.d.
Sábado	1,3545	0,2266	0,9458	0,0176
Domingo	1,3808	0,1843	0,9569	0,0133
Media	1,3687	0,2023	0,9518	0,0162

Tabla 20. Media y desviación típica de θ y R para el sábado y domingo.

2.1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/1	0/2	2/23	x	0/3	0/18	x	1/16	x	1/23
Martes	0/4	1/3	1/25	x	0/4	2/19	x	0/16	x	1/21
Miércoles	0/1	0/1	1/26	x	0/5	2/18	x	2/20	x	1/22
Jueves	0/1	0/2	1/26	x	0/4	2/19	x	1/20	x	0/23
Viernes	0/2	0/4	2/25	x	0/2	0/17	x	2/18	x	1/21
Sábado	0/2	1/1	1/26	x	0/3	1/18	x	0/18	x	0/21
Domingo	1/2	0/1	1/24	x	0/4	1/16	x	2/19	x	1/22
Total días	1	2	9	x	0	8	x	8	x	5

Tabla 21. N° de días modelados con éxito/sin éxito por días y sección.

2.2. Con k-transformada

2.2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
N° días	17 / 26	19 / 26	17 / 26	18 / 26	17 / 26	17 / 26	20 / 26
Total días	17 + 19 + 17 + 18 + 17 + 17 + 20 = 125 (69%)						

Tabla 22. N° de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,1383	< 0,05
kx	< 0,05	0,1637	0,1637
ky	< 0,05	0,05347	0,05347
R	< 0,05	0,7648	0,9066

Tabla 23. Test de normalidad, varianza y media para θ , kx, ky y R.

2.2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/1	0/2	3/24	x	1/3	1/18	x	1/18	x	3/23
Martes	0/5	3/3	3/25	x	1/4	2/20	x	1/18	x	4/21
Miércoles	0/1	0/1	3/26	x	0/5	4/18	x	4/20	x	3/22
Jueves	0/1	0/2	1/26	x	1/4	2/19	x	3/21	x	1/23
Viernes	0/2	0/5	4/26	x	1/2	2/17	x	3/18	x	1/22
Sábado	0/2	1/1	4/26	x	0/3	1/18	x	1/18	x	1/21
Domingo	1/3	0/1	3/24	x	0/4	2/16	x	2/19	x	7/22
Total días	1	4	21	x	4	14	x	14	x	20

Tabla 24. N° de días modelados con éxito/sin éxito por días y sección.

3. PERIÓDICO "La Provincia"

3.1. Sin k-transformada

3.1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
N° días	17 / 25	16 / 25	17 / 25	15 / 25	16 / 25	16 / 25	20 / 25
Total días	17 + 16 + 17 + 15 + 16 + 16 + 20 = 117 (67%)						

Tabla 25. N° de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,9094	0,3870
R	< 0,05	0,707	0,6457

Tabla 26. Test de normalidad, varianza y media para θ y R.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,5266	0,4209	0,9333	0,0317
Martes	1,5488	0,3641	0,9346	0,0247
Miércoles	1,5775	0,3543	0,9244	0,0272
Jueves	1,7381	0,3519	0,9414	0,0291
Viernes	1,7239	0,3031	0,9430	0,0374
Sábado	1,6459	0,3102	0,9304	0,0318
Domingo	1,5324	0,3404	0,9314	0,0328
Media	1,6084	0,3521	0,9338	0,0307

Tabla 27. Media y desviación típica para θ y R por días de la semana.

3.1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/7	0	9/25	0	0	2/11	1/7	2/23	1/2	0/12
Martes	1/6	0/1	7/24	0	0	1/13	0/7	0/19	0/1	0/12
Miércoles	0/4	0	7/25	0	0	0/14	0/5	0/22	0/2	1/9
Jueves	0/4	0/1	7/25	0	0/1	1/10	0/8	0/20	0/1	0/12
Viernes	0/6	0	11/25	0	0	0/12	0/6	0/23	0/1	1/15
Sábado	0/5	0	10/25	0	0	0/12	1/8	1/21	0/1	0/12
Domingo	0/9	0/1	11/25	0	0	1/11	2/6	1/24	0/2	0/12
Total días	1	0	62	0	0	5	4	4	1	2

Tabla 28. Relación del n° de días modelados con éxito/sin éxito por día y sección.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Barlett/Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,4713	0,651
R	0,298	0,9005	0,0771

Tabla 29. Test de normalidad, varianza y media para θ y R para la sección S3 en su totalidad.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,8577	0,7156	0,9429	0,0219
Martes	1,536	0,6782	0,9417	0,0223
Miércoles	1,339	0,3049	0,9297	0,0262
Jueves	1,6386	0,5635	0,9369	0,0189
Viernes	1,541	0,473	0,9566	0,0176
Sábado	1,7077	0,0532	0,9399	0,0223
Domingo	1,6372	0,5244	0,9283	0,0216
Media	1,686	0,5401	0,9399	0,0222

Tabla 30. Media y desviación típica para θ y R por días de la semana para la sección S3.

3.2. Con k-transformada

3.2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	17 / 25	16 / 25	18 / 25	15 / 25	16 / 25	16 / 25	20 / 25
Total días	17 + 16 + 18 + 15 + 16 + 16 + 20 = 118 (67%)						

Tabla 31. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,994	0,8082
kx	< 0,05	0,7927	0,7927
ky	< 0,05	0,4221	0,4221
R	< 0,05	0,5321	0,9193

Tabla 32. Test de normalidad, varianza y media para θ , kx, ky y R.

3.2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/7	0	10/25	0	0	3/11	2/8	2/23	1/2	0/13
Martes	1/8	0/1	8/24	0	0	1/14	0/7	0/20	0/1	1/13
Miércoles	0/4	0	9/25	0	0	0/14	0/5	1/22	1/2	1/10
Jueves	0/5	0/1	9/25	0	0/1	2/10	0/8	2/20	0/1	0/13
Viernes	1/6	0	10/25	0	0	1/12	2/9	3/21	0/1	0/13
Sábado	0/7	0	13/25	0	0	0/14	0/6	2/23	0/1	1/17
Domingo	0/10	0/1	11/25	0	0	2/11	3/6	1/24	0/2	0/13
Total días	2	0	70	0	0	9	7	11	2	3

Tabla 33. N° de días modelados con éxito/sin éxito por días y sección.

4. PERIÓDICO "Levante-EMV"

4.1. Sin k-transformada

4.1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
N° días	24 / 25	25 / 25	24 / 25	23 / 25	24 / 25	25 / 25	25 / 25
Total días	24 + 25 + 24 + 23 + 24 + 25 + 25 = 170 (97%)						

Tabla 34. N° de días modelados con éxito/sin éxito por día de la semana.

Test: θ - Todos los días			Test: R - Todas los días		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,3156	< 0,05	< 0,05	0,0994	0,3131
Test: θ - (L, M, X, J, V)			Test: R - (L, M, X, J, V)		
Shapiro-Wilk (Normal)	Barlett (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
0,2388	0,2010	0,2852	< 0,05	< 0,05	0,3544
Test: θ - (S, D)			Test: R - (S, D)		
Shapiro-Wilk (Normal)	Barlett (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
0,5201	0,723	0,6459	< 0,05	0,3889	0,1502

Tabla 35. Test de normalidad, varianza y media para θ y R para todas los días, de lunes a viernes, y para el par formado por el sábado y el domingo.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,7903	0,1869	0,9558	0,0178
Martes	1,8199	0,1991	0,9482	0,0242
Miércoles	1,7628	0,2355	0,9487	0,0209
Jueves	1,8799	0,1705	0,9544	0,0159
Viernes	1,8305	0,1436	0,9566	0,0119
Media final	1,8162	0,1907	0,9527	0,0187

Tabla 36. Media y desviación típica para θ y R para el conjunto de lunes a viernes.

Días	θ		R	
	Media	s.d.	Media	s.d.
Sábado	1,7051	0,2295	0,9510	0,0229
Domingo	1,6740	0,2469	0,9591	0,0152
Media final	1,6896	0,2364	0,9550	0,0197

Tabla 37. Media y desviación típica para θ y R para el sábado y el domingo.

4.1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/20	0/9	1/24	1/22	0/11	0/19	0/6	0/23	1/8	2/24
Martes	1/20	1/14	1/25	2/21	2/10	1/18	0/4	1/24	1/14	3/25
Miércoles	4/20	0/14	3/25	1/22	0/9	1/17	0/5	1/24	3/13	2/25
Jueves	2/17	1/11	0/23	1/22	1/5	1/15	0/4	1/23	0/16	1/24
Viernes	0/14	0/11	3/24	1/24	1/6	0/13	0/6	2/21	0/4	4/25
Sábado	5/23	0/12	2/25	3/23	0/5	1/14	1/6	2/21	0/11	3/25
Domingo	0/21	0/10	3/25	0/20	0/6	1/19	2/8	2/23	1/2	5/25
Total días	12	2	13	9	4	5	3	9	5	20

Tabla 38. N° de días modelados con éxito/sin éxito por días y sección.

4.2. Con k-transformada

4.2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
N° días	24 / 25	25 / 25	25 / 25	25 / 25	25 / 25	25 / 25	25 / 25
Total días	24 + 25 + 25 + 25 + 25 + 25 + 25 = 174 (99%)						

Tabla 39. N° de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,2346	< 0,05
kx	< 0,05	< 0,05	< 0,05
ky	< 0,05	0,4517	0,4517
R	< 0,05	0,6148	0,4508

Tabla 40. Test de normalidad, varianza y media para θ , kx, ky y R.

4.2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	5/20	1/10	2/24	5/25	0/11	2/19	1/6	0/23	6/8	4/24
Martes	3/20	5/15	2/25	7/21	3/10	2/19	0/4	1/24	8/14	4/25
Miércoles	4/20	2/14	4/25	3/22	2/9	3/18	0/5	2/24	7/13	2/25
Jueves	4/17	3/11	3/23	10/22	2/5	4/15	0/4	3/23	7/16	3/23
Viernes	9/23	2/12	4/25	8/23	3/5	2/15	1/6	4/21	5/11	6/25
Sábado	1/15	4/11	4/24	7/24	1/6	1/14	1/7	4/21	0/4	6/25
Domingo	1/21	1/10	3/25	4/20	0/6	3/20	2/8	2/23	1/2	7/25
Total días	27	18	22	44	11	17	5	16	34	32

Tabla 41. Nº de días modelados con éxito/sin éxito por días y sección.

5. PERIÓDICO "La Nueva España"

5.1. Sin k-transformada

5.1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	26 / 26	26 / 26	26 / 26	26 / 26	26 / 26	26 / 26	26 / 26
Total días	26 + 26 + 26 + 26 + 26 + 26 + 26 = 182 (100%)						

Tabla 42. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Barlett/Levene (Varianza)	F-Anova (Media)
θ	0,1775	0,1061	< 0,05
R	< 0,05	0,9144	0,4824

Tabla 43. Test de normalidad, varianza y media para θ y R.

Test para lo días: L, M, X, J, V, S			
	Shapiro-Wilk (Normal)	Barlett/Levene (Varianza)	F-Anova (Media)
θ	0,3659	0,2033	0,2624
R	< 0,05	0,875	0,4351

Tabla 44. Test de normalidad, varianza y media de θ y R para el conjunto de lunes a sábado.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,5631	0,2098	0,9628	0,0277
Martes	1,5802	0,1927	0,9652	0,0152
Miércoles	1,5809	0,1702	0,9658	0,0128
Jueves	1,5586	0,1795	0,9669	0,0175
Viernes	1,6164	0,1509	0,9726	0,0122
Sábado	1,4982	0,1284	0,9658	0,0098
Media final	1,5570	0,1797	0,9661	0,0168

Tabla 45. Media y desviación típica para θ y R para el conjunto de lunes a sábado.

Día	θ		R	
	Media	s.d.	Media	s.d.
Domingo	1,4429	0,1293	0,9638	0,0161

Tabla 46. Media y desviación típica para θ y R del domingo.

5.1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	3/26	x	21/26	1/25	2/21	0/21	x	x	2/14	23/26
Martes	2/26	x	15/26	0/25	2/20	1/25	x	x	3/12	23/26
Miércoles	1/26	x	19/26	2/24	2/21	3/23	x	x	1/11	21/26
Jueves	1/26	x	23/26	4/25	2/20	1/23	x	x	1/12	23/26
Viernes	0/26	x	20/26	1/23	0/19	1/22	x	x	1/9	24/26
Sábado	3/26	x	17/26	1/24	1/20	1/19	x	x	0/4	26/26
Domingo	4/26	x	21/26	0/26	0/17	1/22	x	x	0/2	25/26
Total días	14	x	136	9	9	8	x	x	8	165

Tabla 47. N° de días modelados con éxito/sin éxito para los días de la semana y por sección.

Test: θ			Test: R		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,4076	< 0,05	< 0,05	0,801	0,6057
Test θ - (L, M, X, J, V, S)			Test: R		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,667	0,3876	< 0,05	0,6995	0,4902

Tabla 48. Test de normalidad, varianza y media para θ y R para todos los días, el conjunto de lunes a sábado de la sección S3.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,3606	0,2717	0,9242	0,0642
Martes	1,5102	0,3355	0,9052	0,0534
Miércoles	1,5088	0,3531	0,9178	0,0482
Jueves	1,4589	0,3447	0,9304	0,0346
Viernes	1,5176	0,3702	0,9233	0,0374
Sábado	1,3992	0,2775	0,9147	0,0402
Media final	1,4592	0,2790	0,9193	0,0463

Tabla 49. Media y desviación típica para θ y R para el conjunto de lunes a sábado de la sección S3.

Día	θ		R	
	Media	s.d.	Media	s.d.
Domingo	1,2695	0,2349	0,9179	0,0448

Tabla 50. Media y desviación típica para θ y R para el domingo de la sección S3.

Test: θ			Test: R		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,2047	< 0,05	< 0,05	0,3518	< 0,05
Test: θ - [L, D] - {V}			Test: R		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,4017	0,2201	< 0,05	0,2634	< 0,05

Tabla 51. Test de normalidad, varianza y media para θ y R para todos los días, el conjunto de lunes a sábado de la sección S10.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,2696	0,381	0,9347	0,0255
Martes	1,279	0,3579	0,9359	0,0238
Miércoles	1,2205	2,2835	0,9392	0,0204
Jueves	1,1347	0,2837	0,9385	0,0246
Sábado	1,3492	0,2181	0,9559	0,0185
Domingo	1,2218	0,2213	0,9455	0,0237
Media	1,2036	0,2992	0,9439	0,0238

Tabla 52. Media y desviación típica para θ y R para el conjunto de lunes a sábado de la sección S10.

Día	θ		R	
	Media	s.d.	Media	s.d.
Viernes	1,4685	0,1963	0,9552	0,021

Tabla 53. Media y desviación típica para θ y R para el domingo de la sección S10.

5.2. Con k-transformada

5.2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	26 / 26	26 / 26	26 / 26	26 / 26	26 / 26	26 / 26	26 / 26
Total días	26 + 26 + 26 + 26 + 26 + 26 + 26 = 182 (100%)						

Tabla 54. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,8554	0,746
kx	< 0,05	< 0,05	< 0,05
ky	< 0,05	0,427	0,427
R	< 0,05	0,7234	0,3358

Tabla 55. Test de normalidad, varianza y media para θ , kx, ky y R.

5.2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	3/26	x	22/26	5/25	4/21	1/21	x	x	4/14	23/26
Martes	4/26	x	17/26	5/25	3/20	3/26	x	x	6/12	23/26
Miércoles	8/26	x	20/26	7/23	4/21	6/23	x	x	7/11	21/26
Jueves	3/26	x	23/26	7/25	4/20	4/23	x	x	2/12	23/26
Viernes	6/26	x	20/26	5/23	3/19	3/22	x	x	4/9	24/26
Sábado	6/26	x	19/26	4/24	5/20	4/19	x	x	0/4	26/26
Domingo	6/26	x	23/26	4/26	4/18	2/22	x	x	0/2	25/26
Total días	36	x	144	37	27	23	x	x	23	165

Tabla 56. N° de días modelados con éxito/sin éxito por días de la semana y por sección.

6. PERIÓDICO "Superdeporte"

6.1. Sin k-transformada

6.1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
N° días	11 / 24	9 / 25	12 / 25	11 / 25	10 / 25	7 / 25	6 / 25
Total días	11 + 9 + 12 + 10 + 11 + 7 + 6 = 66 (36%)						

Tabla 57. N° de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Barlette (Varianza)	F-Anova (Media)
θ	0,952	0,2629	0,07
R	0,09132	0,6067	0,6867

Tabla 58. Test de normalidad, varianza y media para θ y R.

Días	θ		R	
	Media	s.d.	Media	s.d.
Lunes	1,7474	0,3324	0,9379	0,0164
Martes	1,6671	0,3374	0,9391	0,0229
Miércoles	1,8206	0,4994	0,9308	0,0253
Jueves	1,8848	0,5488	0,9273	0,0262
Viernes	1,9803	0,3378	0,9435	0,0163
Sábado	1,7332	0,2754	0,9338	0,0188
Domingo	1,3116	0,2571	0,9310	0,0274
Media final	1,7668	0,4219	0,9348	0,0218

Tabla 59. Media y desviación típica para θ y R por días de la semana.

6.2. Con k-transformada

6.2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	12 / 25	9 / 25	12 / 25	11 / 25	10 / 25	7 / 25	6 / 25
Total días	12 + 9 + 12 + 11 + 10 + 7 + 6 = 66 (38%)						

Tabla 60. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,6022	0,1211
kx	< 0,05	0,3704	0,3704
ky	< 0,05	0,5607	0,5607
R	< 0,05	0,8739	0,5649

Tabla 61. Test de normalidad, varianza y media para θ , kx, ky y R.

ANEXO II.

DATOS ESPECÍFICOS DE LA DISTRIBUCIÓN TWO ZIPF-LIKE

1. PERIÓDICO "La Opinión A Coruña"

1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	11 / 26	9 / 26	4 / 26	4 / 26	6 / 26	6 / 26	9 / 26
Total días	11 + 9 + 4 + 4 + 6 + 6 + 9 = 49 (27%)						

Tabla 1. Nº de días modelados con éxito/sin éxito por días de la semana en la cabeza.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Bartlett/Levene (Varianza)	F-Anova (Media)
Nº vídeos1 (cabeza)	< 0,05	0,6658	0,5824
Nº vídeos2 (cola)	< 0,05	0,621	0,2676
Θ1 (cabeza)	< 0,05	0,4555	0,8636
Θ2 (cola)	< 0,05	0,8673	0,8673
Accesos1 (cabeza)	< 0,05	0,3765	0,3384
Accesos2 (cola)	< 0,05	0,6035	0,547
R1 (cabeza)	0,3899	0,8365	0,08
R2 (cola)	< 0,05	< 0,05	< 0,05

Tabla 2. Test de normalidad, varianza y media.

Días	Θ1		Accesos1		R1	
	Media	σ	Media	σ	Media	σ
Lunes	1,4523	0,2411	330,18	166,87	0,9750	0,0099
Martes	1,6059	0,4739	434,44	374,41	0,9776	0,0083
Miércoles	1,2889	0,7197	340,25	186,67	0,9831	0,0135
Jueves	1,6302	0,0697	408,5	293,28	0,9712	0,0082
Viernes	1,4772	0,3441	395,83	231,99	0,9660	0,0145
Sábado	1,6276	0,1854	221,67	54,11	0,9756	0,0094
Domingo	1,5187	0,6129	205,11	111,45	0,9828	0,0102
Media	1,5184	0,4127	328,33	228,47	0,9762	0,0112

Tabla 3. Media y desviación típica por días de la semana en la cabeza.

Días	$\Theta 2$		Accesos2		R2	
	Media	σ	Media	σ	Media	σ
Lunes	0,1309	0,2983	18,45	12,28	0,0935	1,2152
Martes	0,1418	0,2949	27,11	28,67	0,4436	0,4623
Miércoles	0,2379	0,4758	39,75	52,89	-3,2759	5,1379
Jueves	0	0	16,25	4,99	0,4436	0,4771
Viernes	0,1228	0,3008	30,67	19,62	0,4293	0,9490
Sábado	0,0446	0,1092	12,83	7,78	0,7257	0,0463
Domingo	0,2182	0,3715	22,44	20,61	0,8419	0,1445
Media	0,1354	0,2926	23,14	22,75	x	x

Tabla 4. Media y desviación típica por días de la semana en la cola.

Días	N° vídeos1		N° vídeos2	
	Media	σ	Media	σ
Lunes	16,09	4,3	14,91	4,89
Martes	14,33	5,1	20,22	15,78
Miércoles	11,75	5,19	19,75	13,09
Jueves	16,5	5,92	16,25	4,99
Viernes	17,67	9,16	26,5	17,20
Sábado	13,17	3,19	12,17	6,37
Domingo	13	7,03	14,78	5,02
Media	14,71	5,79	17,45	10,79

Tabla 5. Media y desviación típica por días de la semana.

1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/4	0/0	10/26	0/0	0/1	0/0	0/4	0/2	0/1	0/3
Martes	0/3	0/0	7/26	0/1	0/2	0/1	0/5	0/0	0/0	0/2
Miércoles	0/1	0/0	4/26	1/1	0/2	0/2	0/4	0/2	0/0	0/1
Jueves	0/4	0/1	4/26	0/0	0/0	0/0	0/4	0/3	0/0	0/2
Viernes	0/1	0/0	6/26	0/0	0/0	0/0	0/5	0/2	0/1	0/1
Sábado	0/1	0/0	5/26	0/0	0/1	0/0	0/1	0/4	0/0	1/2
Domingo	0/0	0/0	8/26	0/0	0/0	0/0	0/3	0/4	0/1	0/3
Total días	0	0	44	1	0	0	0	0	0	1

Tabla 6. N° de días modelados con éxito/sin éxito por días de la semana y sección.

2. PERIÓDICO "Faro de Vigo"

2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	7 / 26	9 / 26	12 / 26	12 / 26	13 / 26	10 / 26	14 / 26
Total días	7 + 9 + 12 + 12 + 13 + 10 + 14 = 77 (42%)						

Tabla 7. Nº de días modelados con éxito/sin éxito por días de la semana en la cabeza.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
Nº vídeos1 (cabeza)	< 0,05	0,5101	< 0,05
Nº vídeos2 (cola)	< 0,05	0,7415	0,1176
Θ1 (cabeza)	< 0,05	0,7399	0,3673
Θ2 (cola)	< 0,05	0,1453	< 0,05
Accesos1 (cabeza)	< 0,05	0,2197	0,2902
Accesos2 (cola)	< 0,05	0,1377	< 0,05
R1 (cabeza)	< 0,05	0,5084	< 0,05
R2 (cola)	< 0,05	0,1703	0,1610

Tabla 8. Test de normalidad, varianza y media.

Días	Θ1		Accesos1		R1	
	Media	σ	Media	σ	Media	σ
Lunes	2,9354	1,3159	326,71	324,66	0,9951	0,0107
Martes	1,9528	1,0982	390,11	256,58	0,9697	0,0166
Miércoles	2,6615	1,2011	826,25	1138,54	0,9861	0,0139
Jueves	2,6902	1,38	609,25	821,01	0,9870	0,0169
Viernes	2,2622	1,4109	563,61	852,30	0,9833	0,0148
Sábado	2,6630	1,4076	201,1	105,91	0,9872	0,0135
Domingo	1,8123	1,2883	232,21	359,63	0,9726	0,0180
Media	2,3864	1,3156	462,51	696,92	x	x

Tabla 9. Media y desviación típica por días de la semana en la cabeza.

Días	$\Theta 2$		Accesos2		R2	
	Media	σ	Media	σ	Media	σ
Lunes	0,7908	0,4197	60,86	51,84	0,9017	0,0490
Martes	0,0997	0,2993	17,44	15,22	0,6861	0,18
Miércoles	0,5737	0,3591	41,17	30,27	0,8375	0,0851
Jueves	0,5736	0,4617	46,17	39,77	0,7216	0,3617
Viernes	0,4455	0,4032	28,69	22,48	0,1305	1,8198
Sábado	0,4244	0,46	31,1	27,82	0,6142	0,5911
Domingo	0,2055	0,3651	19,64	23,99	-0,38	1,8514
Media	x	x	x	x	0,5171	1,0423

Tabla 10. Media y desviación típica por días de la semana en la cola.

Días	N° vídeos1		N° vídeos2	
	Media	σ	Media	σ
Lunes	3,4286	2,9358	17,7143	6,8487
Martes	14,2222	6,5912	12,8889	4,8333
Miércoles	7,5	5,4855	18,0833	7,0512
Jueves	7	6,2085	16,9167	6,4167
Viernes	8,3846	5,9936	14,9231	8,3213
Sábado	7,4	5,5817	14,9	7,5491
Domingo	12,0714	5,7708	11	4,8832
Media	x	x	15,0259	6,8786

Tabla 11. Media y desviación típica por días de la semana.

2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/0	0/1	2/23	X	1/1	3/11	X	3/16	X	3/15
Martes	1/3	0/1	6/24	X	0/1	3/14	X	4/15	X	1/15
Miércoles	0/0	0/1	6/23	X	2/4	4/15	X	7/17	X	2/16
Jueves	0/0	0/1	9/22	X	0/0	4/17	X	2/16	X	3/18
Viernes	0/0	0/2	3/19	X	0/1	3/15	X	3/10	X	1/17
Sábado	0/0	0/2	7/19	X	0/1	1/13	X	2/9	X	2/17
Domingo	0/3	0/0	6/24	X	2/2	2/12	X	3/15	X	2/16
Total días	1	0	39	X	5	20	x	24	x	14

Tabla 12. N° de días modelados con éxito/sin éxito por días de semana y sección.

3. PERIÓDICO "La Provincia"

3.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	8 / 25	12 / 25	12 / 25	6 / 25	10 / 25	6 / 25	9 / 25
Total días	8 + 12 + 12 + 6 + 10 + 6 + 9 = 63 (36%)						

Tabla 13. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
Nº vídeos1 (cabeza)	< 0,05	0,7107	0,6635
Nº vídeos2 (cola)	< 0,05	0,7808	0,8018
Θ1 (cabeza)	< 0,05	0,4819	0,4342
Θ2 (cola)	< 0,05	0,4441	0,2141
Accesos1 (cabeza)	< 0,05	0,3313	0,1122
Accesos2 (cola)	< 0,05	0,6578	0,6111
R1 (cabeza)	< 0,05	0,6516	0,5102
R2 (cola)	< 0,05	0,4769	0,4771

Tabla 14. Test de normalidad, varianza y media.

Días	Θ1		Accesos1		R1	
	Media	σ	Media	σ	Media	σ
Lunes	2,2282	1,0386	333,25	321,25	0,9809	0,0211
Martes	3,1066	1,4586	413,42	435,57	0,9912	0,0115
Miércoles	2,2123	1,3940	567,5	424,65	0,9863	0,0149
Jueves	1,7251	0,3899	543,33	347,47	0,9809	0,0159
Viernes	2,5126	1,6054	1042,2	924,91	0,9862	0,0135
Sábado	2,0245	0,7564	423,17	255,77	0,9768	0,0173
Domingo	2,3771	1,4768	739,89	680,04	0,9871	0,0107
Media	2,3916	1,3134	592,33	574,41	0,9852	0,0148

Tabla 15. Media y desviación típica por días de la semana en la cabeza.

Días	$\Theta 2$		Accesos2		R2	
	Media	σ	Media	σ	Media	σ
Lunes	0,3156	0,3427	54,375	61,29	0,4894	0,4868
Martes	0,3659	0,2117	48,25	28,88	0,6972	0,2354
Miércoles	0,5170	0,4212	80,5	102,09	0,6327	0,8004
Jueves	0,1855	0,3152	39,17	44,57	0,9042	0,1005
Viernes	0,5831	0,3398	81,8	69,76	0,9163	0,0387
Sábado	0,2476	0,2757	31	15,41	0,7223	0,2553
Domingo	0,4683	0,4220	58,22	66,65	0,8162	0,1604
Media	0,4089	0,3508	59,41	64,75	0,7219	0,4331

Tabla 16. Media y desviación típica por días de la semana en la cola.

Días	N° vídeos1		N° vídeos2	
	Media	σ	Media	σ
Lunes	11,25	10,73	30,25	22,61
Martes	9,42	11,83	30,83	11,39
Miércoles	7,42	5,79	32,42	18,52
Jueves	15,33	12,27	25,83	16,29
Viernes	11,9	10,77	33,4	16,96
Sábado	14,5	9,16	23,67	8,02
Domingo	14,67	14,63	24,78	11,84
Media	11,46	10,72	29,44	15,50

Tabla 17. Media y desviación típica por días de la semana.

3.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/2	0/0	8/25	0/0	0/0	0/9	0/1	3/22	0/1	0/11
Martes	0/3	0/1	9/24	0/0	0/0	1/10	2/6	7/19	0/0	1/8
Miércoles	0/4	0/0	9/25	0/0	0/0	1/12	1/4	4/21	0/1	2/8
Jueves	0/3	0/1	5/25	0/0	0/1	1/5	0/1	7/19	0/0	1/9
Viernes	0/2	0/0	12/25	0/0	0/0	3/9	1/3	2/23	0/1	3/12
Sábado	1/4	0/0	10/25	0/0	0/0	3/6	0/4	3/20	0/1	3/7
Domingo	0/3	0/0	7/25	0/0	0/0	2/9	0/3	7/20	0/0	1/6
Total días	1	0	59	0	0	11	4	33	0	11

Tabla 18. Relación del n° de días modelados con éxito/sin éxito por categoría y sección.

4. PERIÓDICO "Levante-EMV"

4.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	11 / 25	11 / 25	11 / 25	5 / 24	5 / 25	2 / 25	14 / 25
Total días	11 + 11 + 11 + 5 + 5 + 2 + 14 = 59 (34%)						

Tabla 19. Nº de días modelados con éxito/sin éxito por día de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
Nº vídeos1 (cabeza)	< 0,05	< 0,05	0,1389
Nº vídeos2 (cola)	< 0,05	0,6245	0,863
Θ1 (cabeza)	< 0,05	0,841	0,5528
Θ2 (cola)	< 0,05	0,34	< 0,05
Accesos1 (cabeza)	< 0,05	0,1178	< 0,05
Accesos2 (cola)	< 0,05	0,7606	0,529
R1 (cabeza)	< 0,05	0,468	0,254
R2 (cola)	< 0,05	0,3228	0,325

Tabla 20. Test de normalidad, varianza y media.

Días	Θ1		Accesos1		R1	
	Media	σ	Media	σ	Media	σ
Lunes	1,4650	0,2119	1749,91	442,38	0,9671	0,0147
Martes	1,4039	0,3284	1931,36	1046,79	0,9792	0,0109
Miércoles	1,2685	0,2950	1243,27	739,54	0,9647	0,0194
Jueves	1,4274	0,4121	1512,4	629,82	0,9710	0,0159
Viernes	1,3054	0,2821	1539	622,42	0,9742	0,0131
Sábado	1,7805	0,1459	795,5	355,67	0,9586	0,0163
Domingo	1,3979	0,4454	1021,57	496,70	0,9748	0,0156
Media	1,3951	0,3365	x	x	0,9714	0,0156

Tabla 21. Media y desviación típica por días de la semana en la cabeza.

Días	$\Theta 2$		Accesos2		R2	
	Media	σ	Media	σ	Media	σ
Lunes	0,5754	0,3047	77,91	38,10	0,9259	0,0408
Martes	0,7765	0,2443	130,54	88,30	0,9152	0,0777
Miércoles	0,6913	0,2053	91	35,90	0,8966	0,0437
Jueves	0,4984	0,4789	83	75,12	0,8648	0,0587
Viernes	0,8888	0,1296	131,2	50,21	0,9403	0,0089
Sábado	0,2684	0,3796	51,5	19,09	0,8099	0,1196
Domingo	0,3416	0,4397	83,86	112,26	0,2615	1,5841
Media	x	x	95,63	75,69	0,7701	0,7523

Tabla 22. Media y desviación típica por días de la semana en la cola.

Días	N° vídeos1		N° vídeos2	
	Media	σ	Media	σ
Lunes	21,45	10,29	38	9,16
Martes	14,64	5,18	44,82	11,64
Miércoles	16,82	5,49	42,73	24,18
Jueves	19,6	10,26	35,4	16,55
Viernes	12,6	3,78	43	9,56
Sábado	15	2,83	39	1,41
Domingo	23,36	12,65	35,78	22,63
Media	18,64	9,31	39,86	17,09

Tabla 23. Media y desviación típica por días de la semana.

4.1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	2/13	3/9	3/24	1/18	2/9	6/16	1/1	6/23	0/0	6/24
Martes	4/15	1/8	4/25	4/19	0/5	1/12	0/2	6/23	0/2	4/25
Miércoles	3/17	0/11	5/24	4/21	0/4	4/12	2/2	4/22	0/3	6/25
Jueves	3/16	0/6	5/24	1/20	0/3	2/10	2/4	6/21	2/5	4/24
Viernes	4/16	1/5	9/24	3/17	1/3	2/7	0/3	7/18	0/4	5/25
Sábado	3/12	1/5	4/23	3/17	0/2	3/8	0/5	3/16	0/0	1/2
Domingo	5/12	1/6	6/25	4/19	1/4	3/15	0/5	6/20	0/0	8/25
Total días	24	7	31	20	4	21	5	38	2	34

Tabla 24. N° de días modelados con éxito/sin éxito por días de la semana y sección.

5. PERIÓDICO "La Nueva España"

5.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	19 / 26	19 / 26	18 / 26	18 / 26	24 / 26	22 / 26	24 / 26
Total días	19 + 19 + 18 + 18 + 24 + 22 + 24 = 144 (79%)						

Tabla 25. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
Nº vídeos1 (cabeza)	< 0,05	0,7964	0,7264
Nº vídeos2 (cola)	< 0,05	0,8492	0,5185
Θ1 (cabeza)	< 0,05	0,9155	< 0,05
Θ2 (cola)	< 0,05	0,745	0,3243
Accesos1 (cabeza)	< 0,05	0,555	0,2105
Accesos2 (cola)	< 0,05	0,787	0,6167
R1 (cabeza)	< 0,05	0,6847	0,684
R2 (cola)	< 0,05	0,3122	0,2455

Tabla 26. Test de normalidad, varianza y media.

Días	Θ1		Accesos1		R1	
	Media	σ	Media	σ	Media	σ
Lunes	1,4450	0,2593	2098,63	1755,20	0,9721	0,0148
Martes	1,5080	0,2449	2195,32	1332,24	0,9772	0,0093
Miércoles	1,5254	0,2227	2365,56	1748,50	0,9716	0,0116
Jueves	1,4910	0,1939	2486,61	1599,59	0,9761	0,0160
Viernes	1,4967	0,2118	2476,87	1560,45	0,9737	0,0154
Sábado	1,2963	0,2271	1653,45	1119,89	0,9726	0,0106
Domingo	1,3076	0,2499	1598,5	1004,29	0,9708	0,0127
Media	x	x	2104,92	1463,72	0,9733	0,0130

Tabla 27. Media y desviación típica por días de la semana en la cabeza.

Días	$\Theta 2$		Accesos2		R2	
	Media	σ	Media	σ	Media	σ
Lunes	0,3288	0,3194	109,79	87,49	0,8285	0,1733
Martes	0,1752	0,2687	74,68	71,87	0,7872	0,1958
Miércoles	0,2598	0,3105	84,67	67,48	0,9118	0,0829
Jueves	0,2941	0,3271	95,89	68,78	0,8407	0,1069
Viernes	0,3613	0,3125	107,17	91,73	0,8129	0,3049
Sábado	0,4004	0,2981	115,09	75,93	0,8896	0,0474
Domingo	0,3296	0,2954	103,71	54,67	0,4539	1,4261
Media	0,3121	0,3057	99,64	74,66	0,7791	0,6148

Tabla 28. Media y desviación típica por días de la semana en la cola.

Días	N° vídeos1		N° vídeos2	
	Media	σ	Media	σ
Lunes	37,21	16,99	67,68	46,18
Martes	47,16	16,59	52,37	33,11
Miércoles	40,44	15,97	54,22	21,59
Jueves	42,61	16,69	60	28,81
Viernes	40,29	18,09	58,71	28,03
Sábado	39,36	20,49	66,18	23,06
Domingo	42,87	19,73	66,75	22,56
Media	41,39	17,88	61,14	29,65

Tabla 29. Media y desviación típica por días de la semana.

Días	$\Theta 1$	
	Media	σ
Lunes a Viernes	1,4931	0,2241
Sábado y domingo	1,3022	0,2367

Tabla 30. Media y desviación típica.

5.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	7/26	X	10/26	4/25	2/17	2/15	X	X	1/4	11/26
Martes	1/25	X	7/26	4/25	3/14	9/23	X	X	0/3	14/26
Miércoles	5/26	X	5/26	3/24	4/14	5/17	X	X	1/5	14/26
Jueves	3/26	X	6/26	9/23	5/17	6/21	X	X	1/3	11/26
Viernes	4/26	X	9/26	2/20	1/16	5/19	X	X	0/4	16/26
Sábado	7/26	X	6/26	3/19	5/12	5/14	X	X	0/2	16/26
Domingo	8/26	X	10/26	3/21	2/11	4/16	X	X	0/0	19/26
Total días	35	X	53	28	22	36	X	X	3	101

Tabla 31. N° de días modelados con éxito/sin éxito para los días de la semana y por sección.

6. PERIÓDICO "Superdeporte"

6.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
N° días	7 / 24	10 / 25	4 / 25	5 / 24	4 / 25	9 / 25	5 / 25
Total días	7 + 10 + 4 + 5 + 4 + 9 + 5 = 44 (25%)						

Tabla 32. N° de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
N° vídeos1 (cabeza)	< 0,05	0,1330	0,1378
N° vídeos2 (cola)	< 0,05	0,974	0,9273
Θ1 (cabeza)	< 0,05	0,6927	0,6806
Θ2 (cola)	< 0,05	0,6065	0,6252
Accesos1 (cabeza)	< 0,05	0,5594	0,8564
Accesos2 (cola)	< 0,05	0,85	0,8826
R1 (cabeza)	< 0,05	0,5821	0,5372
R2 (cola)	< 0,05	0,3564	0,3636

Tabla 33. Test de normalidad, varianza y media.

Días	$\Theta 1$		Accesos1		R1	
	Media	σ	Media	σ	Media	σ
Lunes	3,9671	1,4144	643,86	692,52	0,9951	0,0093
Martes	2,8461	0,9221	773,9	849,87	0,9934	0,0102
Miércoles	3,0725	1,1916	453,25	154,48	0,9913	0,0106
Jueves	3,0345	1,5789	1051,8	1171,27	0,9919	0,0087
Viernes	2,7714	1,3176	512,25	146,36	0,9886	0,0140
Sábado	3,6088	2,0115	773	797,6	0,9979	0,0042
Domingo	3,5161	0,7942	457,4	337,26	0,9981	0,0028
Media	3,2918	1,3803	695,70	714,75	0,9943	0,0087

Tabla 34. Media y desviación típica por días de la semana en la cabeza.

Días	$\Theta 2$		Accesos2		R2	
	Media	σ	Media	σ	Media	σ
Lunes	0,4976	0,4607	49,86	49,35	0,5256	0,6212
Martes	0,6089	0,2404	37,4	21,47	0,8890	0,0364
Miércoles	0,4263	0,2974	29,5	15,97	0,6110	0,2711
Jueves	0,4668	0,3772	36,2	28,73	0,3313	1,0558
Viernes	0,2887	0,3334	26,25	22,08	0,5462	0,5141
Sábado	0,5755	0,2657	46,44	43,92	0,7941	0,1149
Domingo	0,6399	0,2249	46,4	19,58	0,8571	0,1114
Media	0,5260	0,3115	40,39	31,93	0,6908	0,4732

Tabla 35. Media y desviación típica por días de la semana en la cola.

Días	N° vídeos1		N° vídeos2	
	Media	σ	Media	σ
Lunes	4,7143	4,8550	22,1429	9,1183
Martes	4,3	2,7908	20	7,4386
Miércoles	6,5	5,1961	21	10,2307
Jueves	5,2	3,1145	19	8,0932
Viernes	10	9,2014	18,75	11,0265
Sábado	3	1,5811	22	7,2111
Domingo	2,8	0,8367	24,4	5,8991
Media	4,75	4,2597	21,1136	7,8034

Tabla 36. Media y desviación típica por días de la semana.

ANEXO III.

DATOS ESPECÍFICOS DE LA DISTRIBUCIÓN MANDELBROT

1. PERIÓDICO "La Opinión A Coruña"

1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	26 / 26	24 / 26	26 / 26	25 / 26	25 / 26	26 / 26	26 / 26
Total días	26 + 24 + 26 + 25 + 25 + 26 + 26 = 178 (98%)						

Tabla 1. Nº de días modelados con éxito/sin éxito por día de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,8699	0,7138
k	< 0,05	0,6745	0,5339

Tabla 2. Test de normalidad, varianza y media para θ y k.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	2,6106	0,7136	3,6413	3,1083
Martes	2,5947	1,2722	3,8203	6,8758
Miércoles	2,6571	1,5498	3,7668	7,3213
Jueves	2,3680	0,6179	2,0853	1,8017
Viernes	2,6591	1,4011	3,5012	4,0525
Sábado	3,1015	2,1485	5,0953	9,6133
Domingo	3,3377	4,5322	8,7446	26,6412
Media	2,7659	2,1426	4,4033	11,6168

Tabla 3. Media y desviación típica para a, b por día de la semana.

1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	2	0	26	0	0	0	3	0	1	0
Martes	3	0	26	0	0	0	4	0	0	0
Miércoles	0	0	26	1	0	0	2	1	0	0
Jueves	1	1	26	0	0	0	2	2	0	1
Viernes	1	0	25	0	0	0	3	2	0	0
Sábado	0	0	26	0	0	0	0	3	0	1
Domingo	0	0	26	0	0	0	1	1	0	0
Total días	7	1	181	1	0	0	14	9	1	2

Tabla 4. Nº de días modelados con éxito por día de la semana y sección.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,2636	0,2147
k	< 0,05	0,3735	0,3272

Tabla 5. Test de normalidad, varianza y media para θ y k en la sección S3.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	2,6868	1,0886	2,3999	3,6533
Martes	2,4615	0,7951	2,1706	3,1006
Miércoles	2,7391	1,2271	2,6583	3,6979
Jueves	2,6223	1,0343	2,0317	2,6661
Viernes	2,6534	1,2928	2,6922	3,4942
Sábado	3,9757	4,5567	6,9586	18,1547
Domingo	3,0790	2,7002	4,4467	10,5190
Media	2,8896	2,2268	3,3404	8,4625

Tabla 6. Media y desviación típica para θ y k por día de la semana para la sección S3.

2. PERIÓDICO "Faro de Vigo"

2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	23/23	19/19	19/19	23/23	23/23	19/19	19/19
Total días	23 + 19 + 19 + 23 + 23 + 19 + 19 = 145 (80%)						

Tabla 7. Nº de días modelados con éxito/sin éxito por día de la semana.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,676	0,6569
k	< 0,05	0,6658	0,5347

Tabla 8. Test de normalidad, varianza y media para θ y k para todos los días de la semana.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	5,3344	6,1323	22,7321	46,3194
Martes	4,3351	6,0070	11,7826	18,2595
Miércoles	3,4273	2,1481	8,2898	8,5408
Jueves	3,9141	1,8780	12,0909	10,8670
Viernes	6,8709	11,9209	18,5630	35,4738
Sábado	4,3007	4,3308	16,1757	18,5166
Domingo	6,3634	10,2395	26,5300	50,5034
Media	4,9714	7,0705	16,6942	31,3946

Tabla 9. Media y desviación típica para θ y k por día de la semana.

2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0	0	18/18	X	1/1	8/9	X	7/7	X	9/9
Martes	3/3	1/1	16/17	X	0	11/11	X	11/12	X	10/10
Miércoles	0	1/1	19/20	X	3/3	8/8	X	13/13	X	12/12
Jueves	0	0	18/19	X	0	13/13	X	11/11	X	7/7
Viernes	0	1/1	16/18	X	1/1	11/11	X	7/7	X	9/9
Sábado	1/1	0	19/19	X	1/1	8/8	X	9/9	X	10/10
Domingo	2/2	0	19/19	X	2/2	7/7	X	8/8	X	7/7
Total días	6	3	125	X	8	66	X	66	X	64

Tabla 10. N° de días modelados con éxito/sin éxito por día de la semana y sección.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,5424	0,5457
k	< 0,05	0,5819	0,5626

Tabla 11. Test de normalidad, varianza y media para θ y k en la sección S3.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	4,7573	6,7370	4,33	8,0692
Martes	3,7922	2,9881	4,0544	7,5956
Miércoles	3,5368	2,8222	3,2114	6,4838
Jueves	2,7931	2,0602	1,1010	2,0642
Viernes	8,8682	17,7987	9,2756	23,0218
Sábado	3,3931	4,2996	2,9863	8,5988
Domingo	6,3222	16,3250	5,0958	14,2779
Media	4,7224	9,5971	4,2064	11,4768

Tabla 12. Media y desviación típica para θ y k por día de la semana para la sección S3.

3. PERIÓDICO "La Provincia"

3.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	22 / 25	24 / 24	22 / 25	21 / 25	23 / 24	25 / 25	24 / 25
Total días	22 + 24 + 22 + 21 + 23 + 25 + 25 = 162 (89%)						

Tabla 13. Nº de días modelados con éxito/sin éxito por día de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,9121	0,8978
k	< 0,05	0.819	0,8186

Tabla 14. Test de normalidad, varianza y media para θ y k.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	2,3450	1,4161	3,7863	9,9067
Martes	2,0825	0,5038	2,0701	3,1409
Miércoles	2,1440	1,0064	2,1532	4,4971
Jueves	2,2456	0,7160	1,7747	2,3388
Viernes	2,1002	0,4901	1,4867	1,8498
Sábado	2,5206	2,3668	4,7589	17,5408
Domingo	2,2713	0,6802	3,0180	5,0743
Media	2,2536	1,2076	2,7468	8,2800

Tabla 15. Media y desviación típica para θ y k por día de la semana.

3.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	2/2	0	21/22	0	0	8/8	1/1	13/15	0	7/7
Martes	0	0	23/23	0	0	7/7	3/3	11/12	0	2/2
Miércoles	0	0	18/21	0	0	7/7	3/3	16/17	1/1	2/3
Jueves	0	0	23/25	0	0	5/5	0	13/14	0	6/6
Viernes	2/2	0	24/25	0	0	6/6	3/3	14/16	1/1	8/8
Sábado	1/2	0	22/25	0	0	4/4	3/3	12/14	1/1	4/4
Domingo	2/2	0	22/24	0	0	5/5	1/1	14/16	0	4/4
Total días	7	0	153	0	0	42	14	93	3	33

Tabla 16. N° de días modelados con éxito/sin éxito por día de la semana y sección.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,6315	0,6738
k	< 0,05	0,5672	0,5651

Tabla 17. Test de normalidad, varianza y media para θ y k de la sección S3.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	2,6507	1,1958	3,4392	5,8945
Martes	3,5410	4,4806	8,1922	18,4694
Miércoles	2,5212	1,6892	2,0281	4,2925
Jueves	7,7088	23,2264	23,8414	80,9192
Viernes	4,7311	8,3670	8,5452	15,2650
Sábado	3,5881	4,0522	11,7306	33,5216
Domingo	4,1069	4,9977	18,9902	55,4358
Media	4,2002	10,0250	11,2839	40,8972

Tabla 18. Media y desviación típica para θ y k por día de la semana para la sección S3.

Tests estadísticos			
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,2071	0,1987
k	< 0,05	0,2396	0,2454

Tabla 19. Test de normalidad, varianza y media para θ y k de la sección S8.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	2,0574	0,7626	0.9864	1.0741
Martes	3,1737	2,1323	4.8143	9.1847
Miércoles	4,0053	4,1212	4.3693	9.1909
Jueves	14,4749	34,5357	22.6989	60.7731
Viernes	3,1531	2,4093	3.2653	4.1307
Sábado	3,1578	2,4903	3.7367	5.9888
Domingo	2,4886	1,6149	2.0735	4.2791
Media	4,6467	13,3083	5,9579	23,7047

Tabla 20. Media y desviación típica para θ y k por día de la semana de la sección S8.

4. PERIÓDICO "Levante-EMV"

4.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	25/25	25/25	25/25	24/24	24/24	24/24	25/25
Total días	25 + 25 + 25 + 24 + 24 + 24 + 25 = 172 (95%)						

Tabla 21. Nº de días modelados con éxito/sin éxito por día de la semana.

Todos los días de la semana					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,2396	< 0,05	< 0,05	< 0,05	< 0,05
Días: lunes a viernes					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,3938	0,6297	< 0,05	0,2585	0,2976
Días: sábado y domingo					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,4742	0,9599	< 0,05	0,2949	0,8616

Tabla 22. Test de normalidad, varianza y media para θ y k para todos los días de la semana, de lunes a viernes, y el par formado por el sábado y el domingo.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	3,1026	1,0046	6,1906	4,7024
Martes	2,8905	0,7643	5,1223	3,5710
Miércoles	3,2692	1,5903	7,5453	8,4944
Jueves	2,8670	0,6508	4,5282	3,5633
Viernes	3,0422	0,7712	5,4161	3,7262
Media	3,0356	1,0124	5,7733	5,2116

Tabla 23. Media y desviación típica para θ y k de lunes a viernes.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Sábado	2,5304	0,4785	3,4105	1,8913
Domingo	2,5379	0,5697	3,5179	2,4115
Media	2,5341	0,5207	3,4642	2,1455

Tabla 24. Media y desviación típica para θ y k del sábado y domingo.

4.2. Con secciones

Secciones										
Días	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	7/7	6/6	20/21	9/9	5/5	13/13	1/1	18/19	0	22/22
Martes	12/12	6/6	21/23	12/12	4/4	8/8	1/1	22/22	2/2	20/20
Miércoles	15/15	11/11	22/22	13/13	5/5	9/9	2/2	14/15	3/3	16/17
Jueves	12/12	3/3	20/21	11/11	2/2	8/8	4/4	16/16	5/5	18/20
Viernes	14/14	4/4	21/21	12/12	3/3	7/7	1/1	16/16	6/6	19/19
Sábado	9/9	2/2	18/20	12/12	1/1	6/6	4/4	14/14	0	22/22
Domingo	9/10	5/5	22/23	14/14	3/3	12/12	3/3	20/20	0	19/20
Total días	78	37	144	83	23	63	16	120	16	136

Tabla 25. N° de días modelados con éxito/sin éxito por día de la semana y sección.

Tests estadísticos			
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,4584	0,4380
k	< 0,05	0,441	0,3886

Tabla 26. Test de normalidad, varianza y media para θ y k de la sección S3.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	3,7030	3,5786	2,7597	6,4087
Martes	3,3775	2,5894	2,2740	4,3628
Miércoles	4,3284	3,3134	3,4054	4,7327
Jueves	2,7243	1,1251	1,0919	1,3057
Viernes	4,1568	3,0518	2,1080	3,5424
Sábado	6,3785	12,9055	8,2667	26,0147
Domingo	3,2175	1,7384	1,8806	2,8868
Media	3,9416	5,2044	3,0149	9,9744

Tabla 27. Media y desviación típica para θ y k por día de la semana para la sección S3.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,4413	0,4519
k	< 0,05	0,5536	0,571

Tabla 28. Test de normalidad, varianza y media para θ y k de la sección S8.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	4,5810	3,4949	3,6241	5,3926
Martes	5,3090	3,9266	4,4148	4,9843
Miércoles	4,7071	5,2796	3,7625	4,4827
Jueves	3,8287	3,5353	2,5012	3,7865
Viernes	12,5131	31,2294	12,6598	36,3314
Sábado	4,7162	3,6851	5,8211	10,5436
Domingo	4,4411	8,4181	4,1943	13,7356
Media	5,6789	12,3387	5,1916	15,1856

Tabla 29. Media y desviación típica para θ y k por día de la semana de la sección S8.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,5827	0,535
K	< 0,05	0,7136	0,6945

Tabla 30. Test de normalidad, varianza y media para θ y k de la sección S10.

Días	θ		k	
	Media	s.d.	Media	s.d.
Lunes	4,0055	6,1458	12,8445	33,0308
Martes	2,5382	1,0182	6,7998	13,1200
Miércoles	2,2980	1,0527	2,3174	3,2459
Jueves	3,0761	1,9612	9,6144	16,6010
Viernes	3,5718	3,2838	11,4104	17,2723
Sábado	2,5670	1,3914	7,5769	19,1849
Domingo	2,7007	1,6582	6,0616	11,4799
Media	2,9902	3,0131	8,2895	18,8347

Tabla 31. Media y desviación típica para θ y k por día de la semana de la sección S10.

5. PERIÓDICO "La Nueva España"

5.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	25/26	24/26	26/26	24/26	26/26	25/26	25/26
Total días	25 + 24 + 26 + 24 + 26 + 25 + 25 = 175 (96%)						

Tabla 32. Nº de días modelados con éxito/sin éxito por día de la semana.

Todos los días de la semana					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,9789	0,4077	< 0,05	0,744	< 0,05
Días: lunes a viernes y domingo					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,9984	0,5085	< 0,05	0,7658	0,1581

Tabla 33. Test de normalidad, varianza y media para θ y k para todos los días de la semana, de lunes a viernes junto con el domingo.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	2,3361	0,2721	4,7656	3,1820
Martes	2,3526	0,3361	4,2088	3,3705
Miércoles	2,3358	0,3018	4,0546	2,2828
Jueves	2,3116	0,2993	4,0730	2,8895
Viernes	2,4672	0,2833	5,2357	2,7674
Domingo	2,3571	0,3013	6,0263	3,8599
Media	2,3610	0,2986	4,7341	3,1228

Tabla 34. Media y desviación típica para θ y k de lunes a viernes y el domingo.

Día	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Sábado	2,4603	0,3733	6,5764	3,9299

Tabla 35. Media y desviación típica para θ y k del sábado.

5.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	22/22	X	23/24	15/16	15/15	11/11	X	X	3/3	25/26
Martes	23/23	X	21/26	17/17	7/7	20/21	X	X	2/2	24/25
Miércoles	18/18	X	23/25	17/17	10/10	15/15	X	X	1/1	25/26
Jueves	21/21	X	24/26	18/18	14/14	17/17	X	X	4/4	25/25
Viernes	12/12	X	25/26	19/19	13/13	16/16	X	X	2/2	26/26
Sábado	19/19	X	25/26	14/14	11/12	12/13	X	X	2/2	25/26
Domingo	22/22	X	25/26	15/15	11/11	13/14	X	X	0	26/26
Total días	137	X	166	115	81	104	X	X	14	176

Tabla 36. Relación del nº de días modelados con éxito/sin éxito por día de la semana y sección.

Todos los días					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	< 0,05	< 0,05	< 0,05	< 0,05	< 0,05
Días: lunes a viernes					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,1423	0,09708	< 0,05	0,2971	0,2708
Días: sábado y domingo					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,4993	0,461	< 0,05	0,5816	0,5945

Tabla 37. Test de normalidad, varianza y media para θ y k para todos los días, de lunes a viernes, y el par formado por el sábado y el domingo para la sección S1.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	3,1016	1,4723	1,6002	1,8463
Martes	5,0040	3,7551	5,3675	7,1110
Miércoles	7,2648	7,0264	7,8585	10,4148
Jueves	6,6056	7,7764	7,1518	16,7630
Viernes	4,3059	3,3456	3,7929	5,1981
Media	5,2550	5,3831	5,1647	9,9798

Tabla 38. Media y desviación típica para θ y k de lunes a viernes para la sección S1.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Sábado	10,4301	11,9225	16,1280	21,9421
Domingo	14,8928	23,6320	21,4698	38,2515
Media	12,8247	19,0326	18,9944	31,4975

Tabla 39. Media y desviación típica para θ y k del sábado y el domingo para la sección S1.

Todos los días					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,2949	0,3539	< 0,05	0,4132	0,4035

Tabla 40. Test de normalidad, varianza y media para θ y k de lunes a viernes, y el par formado por el sábado y el domingo para la sección S3.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	2,3340	0,5694	3,5699	3,3997
Martes	2,3539	0,4016	2,5392	2,1621
Miércoles	4,3311	5,8132	17,6293	43,5896
Jueves	3,2020	2,2878	8,3044	14,5452
Viernes	2,5192	0,7292	3,8347	5,3394
Sábado	2,5695	1,6401	9,2213	31,1121
Domingo	4,0462	7,5479	17,7362	57,2599
Media	3,0599	3,8271	9,0965	30,6962

Tabla 41. Media y desviación típica para θ y k de lunes a viernes para la sección S3.

Días: lunes a viernes					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,5705	0,6536	< 0,05	0,5158	0,5727
Días: sábado y domingo					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	< 0,05	< 0,05	< 0,05	< 0,05	< 0,05

Tabla 42. Test de normalidad, varianza y media para θ y k de lunes a viernes, y el par formado por el sábado y el domingo para la sección S4.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	9,3416	8,8522	6,9112	8,6465
Martes	11,2823	12,0625	10,6537	16,6644
Miércoles	10,7984	14,9752	11,5591	20,3985
Jueves	6,2368	4,6567	4,5107	4,4643
Viernes	9,7540	9,7421	9,8519	14,7484
Media	9,4545	10,4970	8,7170	14,1398

Tabla 43. Media y desviación típica para θ y k de lunes a viernes para la sección S4.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Sábado	21,0956	23,7783	18,2762	25,4784
Domingo	6,4509	7,5525	3,9384	9,1313

Tabla 44. Media y desviación típica para θ y k del sábado y domingo para la sección S4.

Días: lunes a sábado					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,2893	0,3725	< 0,05	0,2755	0,2957

Tabla 45. Test de normalidad, varianza y media para θ y k de lunes a sábado, y el domingo en la sección S5.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	7,0375	5,8832	3,4697	4,6664
Martes	5,7157	5,5637	3,6789	5,6773
Miércoles	4,4229	3,4556	2,7904	4,1534
Jueves	5,1052	3,4939	2,4147	2,8609
Viernes	5,7148	6,5588	5,4577	9,0489
Sábado	11,8713	18,0509	11,6826	22,8691
Media	5,6593	8,6088	4,8424	10,5325

Tabla 46. Media y desviación típica para θ y k de lunes a sábado

para la sección S5.

Día	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Domingo	25,5612	39,8096	17,2864	32,2982

Tabla 47. Media y desviación típica para θ y k en la categoría del domingo para la sección S5.

Todos los días de la semana					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,5343	0,4909	< 0,05	0,6163	0,5699

Tabla 48. Test de normalidad, varianza y media para θ y k para todos los días de la semana para la sección S6.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	4,5006	4,8590	4,2343	10,2090
Martes	4,8365	5,0227	2,9763	5,6055
Miércoles	7,8177	9,3057	4,9415	7,1605
Jueves	4,4976	5,5927	2,4741	6,2660
Viernes	9,1579	15,7190	11,6717	33,7426
Sábado	4,1628	4,0993	2,9248	4,4869
Domingo	4,5360	2,4230	2,7414	2,5126
Media	5,7251	8,0951	4,6132	14,4048

Tabla 49. Media y desviación típica para θ y k de lunes a viernes para la sección S6.

Todos los días de la semana					
Test: Parámetro θ			Test: Parámetro k		
Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
< 0,05	0,3287	0,3595	< 0,05	0,3541	0,3469

Tabla 50. Test de normalidad, varianza y media para θ y k de lunes a viernes, y el par formado por el sábado y el domingo para la sección S10.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	2,3686	0,6883	6,6451	8,3945
Martes	2,6823	1,6887	9,4994	17,8137
Miércoles	2,6198	1,1199	8,2436	8,3297
Jueves	4,5939	9,0142	39,3762	138,6664
Viernes	2,5610	0,7168	4,7805	5,3329
Sábado	3,6702	3,5417	19,6982	50,3808
Domingo	3,1021	1,2162	16,2113	18,7923
Media	3,0848	3,7818	14,9026	56,7886

Tabla 51. Media y desviación típica para θ y k de lunes a viernes para la sección S10.

6. PERIÓDICO "Superdeporte"

6.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	21/26	20/26	20/26	18/26	18/26	17/26	19/26
Total días	21 + 20 + 20 + 18 + 18 + 17 + 19 = 133 (73%)						

Tabla 52. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Barlett/Levene (Varianza)	F-Anova (Media)
θ	< 0,05	0,9657	0,954
k	< 0,05	0,991	0,9929

Tabla 53. Test de normalidad, varianza y media para θ y k.

Días	Parámetro θ		Parámetro k	
	Media	s.d.	Media	s.d.
Lunes	2,4150	1,3505	1,8181	3,9545
Martes	2,4143	2,2202	2,6116	8,4931
Miércoles	2,4169	1,4504	2,0413	4,5352
Jueves	2,2451	0,7261	2,3322	6,6153
Viernes	2,2375	0,7723	1,4954	2,4774
Sábado	2,3556	0,9092	2,3534	3,7348
Domingo	2,0105	0,5466	1,6108	2,4391
Media	2,3028	1,2625	2,0357	4,9789

Tabla 54. Media y desviación típica para θ y k para todos los días.

ANEXO IV.

DATOS ESPECÍFICOS DE LA DISTRIBUCIÓN STRETCHED

1. PERIÓDICO "La Opinión A Coruña"

1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	26 / 26	23 / 26	22 / 26	21/ 26	24 / 26	25 / 26	24 / 26
Total días	26 + 23 + 22 + 21 + 24 + 25 + 24 = 165 (91%)						

Tabla 1. Nº de días modelados con éxito/sin éxito por días de la semana.

Parámetros	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,2733	0,3121
b	< 0,05	0,2486	0,2858
c	< 0,05	0,1652	0,2452

Tabla 2. Test de normalidad, varianza y media para a, b y c por días de la semana.

Días	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,1062	0,1638	1,2755	0,4321	0,0556	0,0812
Martes	-0,0628	0,0968	1,1633	0,2561	0,0363	0,0550
Miércoles	-0,0766	0,1493	1,1900	0,3664	0,0379	0,0685
Jueves	-0,1425	0,3457	1,3531	0,8769	0,0555	0,1122
Viernes	-0,1336	0,2059	1,3327	0,4987	0,0643	0,0877
Sábado	-0,1968	0,2821	1,5001	0,7376	0,0898	0,1132
Domingo	-0,2218	0,4310	1,5890	1,1382	0,0959	0,1399
Media	-0,1354	0,2635	1,3463	0,6823	0,0628	0,0986

Tabla 3. Media y desviación típica para a, b y c por días de la semana.

1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/4	0	17/26	0	0/1	0	3/9	1/5	0/2	1/3
Martes	0/4	0	15/26	0	0	0	0/7	1/2	0	1/6
Miércoles	0/4	0	17/26	0/1	0/3	0/2	1/7	2/7	0	1/6
Jueves	1/4	0/1	9/26	0/1	0/1	0	0/9	2/6	0	0/3
Viernes	0/1	0	15/26	0	0/1	0	1/7	0/4	0/1	2/6
Sábado	0/3	0	16/26	0	0/2	0/1	1/6	0/7	0	0/4
Domingo	1/2	0	18/26	0	0	0	3/11	0/8	0/1	0/5
Total días	2	0	107	0	0	0	9	6	0	5

Tabla 4. N° de días modelados con éxito por días de la semana y sección.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,9136	0,9105
b	< 0,05	0,9046	0,9039
c	< 0,05	0,7614	0,7963

Tabla 5. Test de normalidad, varianza y media para a, b y c en la sección S3, todas los días menos el sábado.

Días	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,1369	0,1726	1,3346	0,4302	0,0662	0,0801
Martes	-0,1225	0,1569	1,2911	0,3738	0,0582	0,0715
Miércoles	-0,1209	0,1579	1,2884	0,3829	0,0577	0,0690
Jueves	-0,0469	0,0934	1,1086	0,2162	0,0226	0,0449
Viernes	-0,2339	0,3075	1,5357	0,6708	0,0932	0,1069
Domingo	-0,2129	0,3863	1,5430	1,0172	0,0879	0,1353
Media	-0,1537	0,2461	1,3708	0,6085	0,0677	0,0929

Tabla 6. Media y desviación típica para a, b y c por días de la semana para la sección S3 excepto el sábado.

Día	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Sábado	-0,5216	0,6974	2,2287	1,6346	0,1669	0,1719

Tabla 7. Media y desviación típica para a, b y c del sábado en la sección S3.

2, PERIÓDICO "Faro de Vigo"

2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	17/26	19/26	17/26	18/26	17/26	17/26	20/26
Total días	$17 + 19 + 17 + 18 + 17 + 17 + 20 = 125$ (69%)						

Tabla 8. Nº de días modelados con éxito/sin éxito por días de la semana.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	< 0,05	< 0,05
b	< 0,05	< 0,05	< 0,05
c	< 0,05	< 0,05	< 0,05

Tabla 9. Test de normalidad, varianza y media para a, b y c para todos los días de la semana.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,2589	0,261
b	< 0,05	0,2274	0,23
c	< 0,05	0,1802	0,1839

Tabla 10. Test de normalidad, varianza y media para a, b y c para todos los días excepto el viernes.

Días	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,0047	0,0161	1,0100	0,0342	0,0023	0,0077
Martes	-0,0676	0,1381	1,1523	0,3045	0,0296	0,0544
Miércoles	-0,0217	0,0552	1,0474	0,1166	0,0103	0,0240
Jueves	-0,0356	0,0639	1,0857	0,1580	0,0192	0,0353
Sábado	-0,0301	0,0673	1,0787	0,1766	0,0179	0,0402
Domingo	-0,0526	0,0917	1,1434	0,2451	0,0332	0,0518
Media	-0,0364	0,0833	1,1991	0,15	0,0194	0,0403

Tabla 11. Media y desviación típica para a, b y c por días de la semana.

Día	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Viernes	-0,1094	0,1486	1,2716	0,3794	0,0558	0,0707

Tabla 12. Media y desviación típica para a, b y c del viernes.

2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/1	0/2	3/24	X	1/3	0/18	X	1/16	X	3/23
Martes	0/4	2/3	2/25	X	1/4	2/19	X	0/16	X	4/21
Miércoles	0/1	0/1	3/26	X	0/5	2/18	X	4/20	X	2/22
Jueves	0/1	0/2	1/26	X	0/4	2/19	X	2/20	X	1/23
Viernes	0/2	0/4	4/25	X	0/2	2/17	X	2/18	X	1/21
Sábado	0/2	1/1	3/26	X	0/3	1/18	X	0/18	X	0/21
Domingo	1/3	0/1	3/24	X	0/4	1/16	X	2/19	X	7/22
Total días	1	3	19	X	2	10	X	11	X	18

Tabla 13. N° de días modelados con éxito/sin éxito para los días de la semana por sección.

3, PERIÓDICO "La Provincia"

3.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	17 / 25	16 / 25	17 / 25	15 / 25	16 / 25	16 / 25	20 / 25
Total días	17 + 16 + 17 + 21 + 16 + 16 + 20 = 123 (68%)						

Tabla 14. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,9268	0,9268
b	< 0,05	0,8842	0,8842
c	< 0,05	0,8665	0,8665

Tabla 15. Test de normalidad, varianza y media para a, b y c para todos los días de la semana.

Días	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,0498	0,1114	1,1437	0,3267	0,0328	0,0766
Martes	-0,0354	0,1169	1,0902	0,29916	0,0176	0,0559
Miércoles	-0,0398	0,1325	1,0991	0,3419	0,0184	0,0616
Jueves	-0,0121	0,0251	1,0268	0,0556	0,0059	0,0124
Viernes	-0,0393	0,0813	1,0976	0,2024	0,0215	0,0438
Sábado	-0,0637	0,1861	1,1780	0,5413	0,0332	0,0951
Domingo	-0,0296	0,0768	1,0728	0,1892	0,0155	0,0398
Media	-0,0385	0,1117	1,1012	0,3063	0,0207	0,0592

Tabla 16. Media y desviación típica para a y b por días de la semana.

3.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	0/7	0	10/25	0	0	3/11	2/7	2/23	1/2	0/12
Martes	1/6	0/1	8/24	0	0	1/13	0/7	0/19	0/1	1/12
Miércoles	0/4	0	8/25	0	0	0/14	0/5	2/22	0/2	1/9
Jueves	0/4	0/1	9/25	0	0/1	1/10	0/8	1/20	0/1	0/12
Viernes	0/6	0	12/25	0	0	0/12	0/6	2/23	0/1	1/15
Sábado	1/5	0	10/25	0	0	1/12	2/8	2/21	0/1	0/13
Domingo	0/9	0/1	11/25	0	0	2/11	3/6	1/24	0/2	0/12
Total días	2	0	68	0	0	8	7	10	1	3

Tabla 17. N° de días modelados con éxito/sin éxito para los días de la semana por sección.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,8032	0,5846
b	< 0,05	0,7553	0,5553
c	< 0,05	0,829	0,5596

Tabla 18. Test de normalidad, varianza y media para a, b y c en la sección S3, todos los días de la semana.

Días	Parámetros					
	a		b		c	
	Media	s,d,	Media	s,d,	Media	s,d,
Lunes	-0,1856	0,2444	1,5051	0,6953	0,1004	0,1391
Martes	-0,3455	0,3081	1,9979	0,9442	0,1943	0,1706
Miércoles	-0,1965	0,2911	1,5705	0,8323	0,1094	0,1558
Jueves	-0,3119	0,4061	1,8565	1,1869	0,1458	0,2025
Viernes	-0,1622	0,2590	1,4389	0,6866	0,0859	0,1301
Sábado	-0,1264	0,2517	1,3436	0,7096	0,0637	0,1209
Domingo	-0,1483	0,2341	1,3936	0,6566	0,0767	0,1276
Media	-0,2035	0,2827	1,5638	0,8108	0,1067	0,1479

Tabla 19. Media y desviación típica para a, b y c por días de la semana para la sección S3.

4. PERIÓDICO "Levante-EMV"

4.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	24/25	25/25	25/25	24/25	25/24	25/25	25/25
Total días	24 + 25 + 25 + 24 + 25 + 25 + 25 = 173 (95%)						

Tabla 20. Nº de días modelados con éxito/sin éxito por días de la semana.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,458	0,3502
b	< 0,05	0,394	0,3235
c	< 0,05	0,2338	0,1556

Tabla 21. Test de normalidad, varianza y media para a, b y c todos los días menos el domingo.

Días	Parámetro a		Parámetro b		Parámetro c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,1641	0,1976	1,4247	0,5638	0,0851	0,0917
Martes	-0,2218	0,1935	1,5897	0,5451	0,1186	0,1007
Miércoles	-0,2539	0,2966	1,6860	0,8416	0,1269	0,1317
Jueves	-0,2273	0,2905	1,5998	0,8101	0,1104	0,1209
Viernes	-0,1927	0,2175	1,4901	0,5659	0,0952	0,0893
Sábado	-0,1127	0,2198	1,2857	0,5684	0,0527	0,0903
Media final	-0,1954	0,2399	1,5127	0,6626	0,0982	0,1065

Tabla 22. Media y desviación típica para a, b y c de lunes a sábado.

Día	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Domingo	-0,0631	0,0914	1,1569	0,2272	0,0346	0,0489

Tabla 23. Media y desviación típica para a, b y c del domingo.

4.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	3/19	1/9	2/24	3/22	0/11	2/19	1/6	0/23	5/8	4/24
Martes	3/20	3/14	1/25	5/21	3/10	3/18	0/4	1/24	5/14	3/25
Miércoles	4/20	2/14	3/25	3/22	3/9	3/17	0/5	3/24	5/13	2/25
Jueves	4/17	3/11	3/23	8/22	2/5	3/15	0/4	1/23	2/16	2/24
Viernes	8/23	2/12	3/25	8/23	1/5	2/14	1/6	4/21	3/11	6/25
Sábado	2/14	1/11	4/24	5/24	1/6	1/13	0/6	4/21	0/4	5/25
Domingo	1/21	0/10	3/25	2/20	0/6	2/19	2/8	2/23	1/2	7/25
Total días	25	12	19	34	10	16	4	15	21	29

Tabla 24. N° de días modelados con éxito/sin éxito de los días de la semana por sección.

5. PERIÓDICO "La Nueva España"

5.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
N° días	26/26	26/26	26/26	26/26	26/26	26/26	26/26
Total días	26 + 26 + 26 + 26 + 26 + 26 + 26 = 182 (100%)						

Tabla 25. N° de días modelados con éxito/sin éxito para todos los días de la semana.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,1312	0,1126
b	< 0,05	0,0894	0,08854
c	< 0,05	0,05471	0,05881

Tabla 26. Test de normalidad, varianza y media para a, b y c para todos los días de la semana menos el sábado.

Días	Parámetro a		Parámetro b		Parámetro c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,0469	0,0639	1,1281	0,1786	0,0319	0,0438
Martes	-0,0456	0,0799	1,1228	0,2168	0,0282	0,0489
Miércoles	-0,0406	0,0547	1,1070	0,1464	0,0264	0,0356
Jueves	-0,0435	0,0641	1,1154	0,1749	0,0281	0,0419
Viernes	-0,0787	0,0893	1,2114	0,2488	0,0506	0,0548
Domingo	-0,0945	0,1349	1,2778	0,4307	0,0688	0,1028
Media	-0,0583	0,0865	1,1604	0,2549	0,0390	0,0601

Tabla 27. Media y desviación típica para a, b y c de lunes a viernes junto con el domingo.

Día	Parámetros					
	a		b		c	
	Media	s,d,	Media	s,d,	Media	s,d,
Sábado	-0,1124	0,1112	1,3188	0,3265	0,0782	0,0779

Tabla 28. Media y desviación típica para a, b y c del sábado.

5.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	3/26	x	22/26	5/25	5/21	0/21	x	x	3/14	23/26
Martes	3/25	x	17/24	4/25	2/20	3/25	x	x	4/12	23/26
Miércoles	7/26	x	20/26	5/24	2/21	6/23	x	x	4/11	21/26
Jueves	3/26	x	23/26	6/25	6/23	3/23	x	x	2/12	24/26
Viernes	5/26	x	20/26	3/23	3/19	3/22	x	x	2/9	24/26
Sábado	5/26	x	18/26	3/24	4/20	4/19	x	x	0/4	26/26
Domingo	6/26	x	23/26	2/26	3/18	6/22	x	x	0/2	25/26
Total días	32	x	143	28	25	25	x	x	15	166

Tabla 29. N° de días modelados con éxito/sin éxito para los días de la semana por sección.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,956	0,956
b	< 0,05	0,9503	0,9503
c	< 0,05	0,937	0,937

Tabla 30. Test de normalidad, varianza y media para a, b y c en la sección S3 para todos los días de la semana.

Días	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,1609	0,3143	1,4653	0,8991	0,0838	0,1398
Martes	-0,1464	0,3385	1,3995	0,9083	0,0680	0,1261
Miércoles	-0,1206	0,2734	1,3656	0,8429	0,0672	0,1398
Jueves	-0,1223	0,2436	1,3504	0,7008	0,0672	0,1194
Viernes	-0,0708	0,1266	1,1933	0,3533	0,0418	0,0732
Sábado	-0,1016	0,2872	1,2959	0,8307	0,0540	0,1206
Domingo	-0,1389	0,2459	1,4099	0,7385	0,0818	0,1274
Media	-0,1237	0,2624	1,3568	0,7578	0,0669	0,1211

Tabla 31. Media y desviación típica para a, b y c por días de la semana en la sección S3.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,2348	0,2974
b	< 0,05	0,239	0,2955
c	< 0,05	0,1238	0,1995

Tabla 32. Test de normalidad, varianza y media para a, b y c en la sección S10 para todos los días de la semana.

Días	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,0908	0,1828	1,2833	0,5509	0,0649	0,1121
Martes	-0,0690	0,1596	1,2121	0,4998	0,0457	0,1034
Miércoles	-0,1627	0,2861	1,4964	0,8958	0,1007	0,1530
Jueves	-0,0998	0,1376	1,3242	0,4517	0,0817	0,1077
Viernes	-0,0294	0,0602	1,0775	0,1595	0,0187	0,0367
Sábado	-0,1253	0,2220	1,3639	0,6624	0,0747	0,1251
Domingo	-0,1214	0,1854	1,3761	0,5905	0,0900	0,1208
Media	-0,0993	0,1869	1,3032	0,5788	0,0678	0,1136

Tabla 33. Media y desviación típica para a, b y c todos los días de la semana en la sección S10.

6. PERIÓDICO "Superdeporte"

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	12/24	9/25	12/25	11/25	10/25	7/25	6/25
Total días	12 + 9 + 12 + 11 + 10 + 7 + 6 = 67 (37%)						

Tabla 34. Nº de días modelados con éxito/sin éxito para todos los días de la semana.

Parámetros	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,8792	0,8792
b	< 0,05	0,868	0,868
c	< 0,05	0,8368	0,8368

Tabla 35. Test de normalidad, varianza y media para a, b y c en la sección S3 para todos los días de la semana.

Días	Parámetros					
	a		b		c	
	Media	s.d.	Media	s.d.	Media	s.d.
Lunes	-0,0296	0,0939	1,0611	0,192	0,0117	0,036
Martes	-0,0037	0,0111	1,0075	0,0226	0,0017	0,0052
Miércoles	-0,0353	0,0999	1,0721	0,196	0,0143	0,0362
Jueves	-0,0303	0,0968	1,064	0,203	0,0121	0,0375
Viernes	-0,026	0,0551	1,0629	0,136	0,0148	0,0326
Sábado	0	0	1	0	0	0
Domingo	0	0	1	0	0	0
Media	-0,0209	0,0719	1,0448	0,1491	0,0091	0,0288

Tabla 36. Media y desviación típica para a, b y c para todos los días.

ANEXO V.

DATOS ESPECÍFICOS DE LA DISTRIBUCIÓN BOX-COX

1. PERIÓDICO "La Opinión A Coruña"

1.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	19/26	15/26	13/26	13/26	14/26	18/26	14/26
Total días	19 + 15 + 13 + 13 + 14 + 18 + 14 = 106 (58%)						

Tabla 1. Nº de días modelados con éxito/sin éxito en todos los días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,7953	0,2227
b	< 0,05	0,7528	0,1559
λ	< 0,05	0,5145	0,1233
R	< 0,05	0,5369	0,2339

Tabla 2. Test de normalidad, varianza y media para a, b, c y R.

Días	a		b		λ		
	Media	σ	Media	σ	Media	Media li	Media ls
Lunes	-1,5464	0,0585	3,7051	0,1499	-0,0526	-0,0631	-0,0474
Martes	-1,5877	0,0702	3,3788	0,1710	-0,1333	-0,1667	-0,1267
Miércoles	-1,4470	0,0685	3,2408	0,1719	-0,2154	-0,2462	-0,1923
Jueves	-1,4454	0,0655	3,1702	0,1687	-0,1923	--0,23	-0,1846
Viernes	-1,5419	0,0767	3,6556	1,1866	-0,0786	-0,1214	-0,0643
Sábado	-1,6767	0,075	3,9612	1,1806	-0,0333	-0,05	-0,0167
Domingo	-1,29	0,0541	3,2344	0,1354	-0,1286	-0,15	-0,1071
Media	-1,5154	0,0669	3,5112	0,1673	-0,1113	-0,1377	-0,0981

Tabla 3. Media y desviación típica para a, b y c para los días de la semana.

Días	R	
	Media	σ
Lunes	0,9587	0,0182
Martes	0,9468	0,0315
Miércoles	0,9351	0,0415
Jueves	0,9431	0,0184
Viernes	0,9421	0,0194
Sábado	0,9517	0,0226
Domingo	0,9511	0,0253
Media	0,9478	0,0262

Tabla 4. Media y desviación típica para R para los días de la semana.

1.2. Con secciones

Categorías	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	5/5	0	18/26	0	1/1	0	9/9	6/6	2/2	4/4
Martes	3/4	0	19/26	2/2	3/3	1/1	7/7	3/3	0	8/8
Miércoles	5/5	0	15/26	2/2	3/3	2/2	9/9	7/7	0	7/7
Jueves	6/6	1/1	17/26	1/1	2/2	0	10/10	6/6	0	3/3
Viernes	1/1	0	17/26	0	1/1	0	8/8	5/5	1/1	6/6
Sábado	3/3	0	21/26	0	2/2	1/1	6/6	7/7	0	5/5
Domingo	2/2	0	18/26	1/1	0	0	12/12	8/8	1/1	7/7
Total días	25	1	125	6	12	4	61	39	4	40

Tabla 5. N° días modelados éxito/sin éxito todos los días de semana por sección.

	Tests estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,1155	0,8297
b	< 0,05	0,08837	0,5041
λ	< 0,05	0,6478	0,3178
R	< 0,05	0,3034	0,08379

Tabla 6. Test de normalidad, varianza y media para a, b, λ y R para la sección S3 para todos los días de la semana.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,7193	0,4553	3,8428	1,0332	-0,0786	-0,1000	-0,0643
Martes	-1,6665	0,4077	3,4267	0,9422	-0,1765	-0,2059	-0,1500
Miércoles	-1,5772	0,4714	3,3189	1,0676	-0,2667	-0,3454	-0,2454
Jueves	-1,5214	0,5354	3,0500	1,0374	-0,3267	-0,4286	-0,2294
Viernes	-1,6004	0,6639	3,4778	1,3659	-0,2000	-0,1928	-0,1600
Sábado	-1,8028	0,9406	3,9578	2,1435	-0,1316	-0,1470	-0,0889
Domingo	-1,5549	0,7159	3,5982	1,5867	-0,1625	-0,2428	-0,0937
Media	-1,6415	0,6274	3,5433	1,3979	-0,1879	-0,2281	-0,1440

Tabla 7. Media y desviación típica para a, b, λ y R por días de la semana para la sección S3.

Días	R	
	Media	σ
Lunes	0,9558	0,0173
Martes	0,9381	0,0314
Miércoles	0,9279	0,0417
Jueves	0,9146	0,0456
Viernes	0,9259	0,0526
Sábado	0,9330	0,0381
Domingo	0,9376	0,0364
Media	0,9336	0,0395

Tabla 8. Media y desviación típica para R por días de la semana para la sección S3.

2. PERIÓDICO "Faro de Vigo"

2.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	24 / 26	23 / 26	22 / 26	21 / 26	25 / 26	25 / 26	25 / 26
Total días	24 + 23 + 22 + 21 + 25 + 25 + 25 = 165 (91%)						

Tabla 9. Nº de días modelados con éxito/sin éxito por días de la semana.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,3228	0,0679	2,8578	0,1565	-0,2125	0,2417	-0,1917
Martes	-1,3796	0,0713	3,0149	0,1678	-0,1826	-0,1869	-0,1522
Miércoles	-1,4277	0,0705	2,9177	0,1665	-0,1954	-0,2091	-0,1818
Jueves	-1,3825	0,0673	3,0012	0,1583	-0,1619	-0,1714	-0,1428
Viernes	-1,6281	0,0795	3,3616	0,1774	-0,124	-0,14	-0,108
General	-1,4315	0,0715	3,0364	0,1655	-0,1748	-0,1896	-0,1548

Tabla 10. Media y desviación típica para a, b, λ y R de lunes a viernes.

Días	R	
	Media	σ
Lunes	0,9395484	0,0369
Martes	0,9406633	0,0272
Miércoles	0,9414736	0,0307
Jueves	0,9479048	0,0190
Viernes	0,9478554	0,0356
General	0,9435	0,0306

Tabla 11. Media y desviación típica para R de lunes a viernes.

Test estadísticos - Todas las categorías			
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,3176	< 0,05
Test estadísticos - Lunes a Viernes			
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,8638	0,2297
b	< 0,05	0,1096	0,2166
λ	< 0,05	0,05	0,4686
R	< 0,05	0,817	0,8084
Test estadísticos - Sábado y Domingo			
	Shapiro-Wilk (Normal)	Barlett/Levene (Varianza)	F-Anova (Media)
a	0,9832	0,959	0,3831
b	0,1828	0,7917	0,1847
λ	< 0,05	0,761	0,2087
R	< 0,05	0,3596	0,1143

Tabla 12. Test de normalidad, varianza y media para a, b, λ y R.

Días	a		b		λ		
	Media	σ	Media	σ	Media	li	ls
Sábado	-1,1596	0,0651	2,8705	0,1467	-0,184	-0,212	-0,156
Domingo	-1,2475	0,0613	3,1527	0,1396	-0,12	-0,132	-0,096
General	-1,2036	0,0632	3,0116	0,1432	-0,152	-0,172	-0,126

Tabla 13. Media y desviación típica para a, b, λ para el sábado y el domingo.

Días	R	
	Media	σ
Sábado	0,9349	0,0311
Domingo	0,9482	0,0272
General	0,9415	0,0297

Tabla 14. Media y desviación típica para R para el sábado y el domingo.

2.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	1/1	2/2	24/24	X	3/3	18/18	X	14/14	X	23/23
Martes	7/7	3/3	25/25	X	4/4	6/6	X	18/18	X	21/21
Miércoles	1/1	1/1	26/26	X	5/5	18/18	X	20/21	X	22/22
Jueves	1/1	2/2	26/26	X	4/4	19/19	X	20/21	X	23/23
Viernes	2/2	5/5	26/26	X	2/2	17/17	X	18/18	X	21/21
Sábado	2/2	1/1	26/26	X	3/3	18/18	X	18/18	X	21/21
Domingo	4/4	1/1	24/24	X	4/4	16/16	X	14/14	X	22/22
Total días	18	15	177	X	25	112	X	122	X	153

Tabla 15. N° de días modelados con éxito/sin éxito para todos los días de la semana por sección.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,9299	0,929
b	< 0,05	0,2068	0,4080
λ	< 0,05	0,3889	0,6079
R	< 0,05	0,09437	0,0558

Tabla 16. Test de normalidad, varianza y media para a, b, λ y R en la sección S3.

Días	a		b		λ		
	Media	σ	Media	σ	Media	li.	ls.
Lunes	-1,2784	1,6842	2,5149	2,7385	-0,8300	-1,07	-0,515
Martes	-1,1966	2,3405	1,5606	2,2083	-1,2762	-1,6	-0,8857
Miércoles	-1,1697	1,2824	1,9032	1,2945	-0,8800	-1,048	-0,644
Jueves	-1,9104	3,4176	2,4131	2,2137	-0,8048	-0,9524	-0,6428
Viernes	-1,4724	3,1824	1,5409	1,1313	-0,9500	-1,1667	-0,6958
Sábado	-1,2718	1,9569	2,9414	3,7238	-0,8680	-1,14	-0,592
Domingo	-1,7727	2,8220	2,6932	3,8393	-0,8833	-1,15	-0,625
Media	-1,4358	2,4523	2,2280	2,6577	-0,925	-1,1581	-0,6562

Tabla 17. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S3.

Días	R	
	Media	σ
Lunes	0,8682	0,1038
Martes	0,8769	0,0913
Miércoles	0,9131	0,0708
Jueves	0,9176	0,0588
Viernes	0,9246	0,0690
Sábado	0,9216	0,0622
Domingo	0,9199	0,0563
Media final	0,9074	0,0755

Tabla 18. Media y desviación típica para R para todos los días de la semana en la sección S3.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,5462	0,5462
b	< 0,05	0,8382	0,8382
λ	< 0,05	0,7629	0,8834
R	< 0,05	0,05	0,1687

Tabla 19. Test de normalidad, varianza y media para a, b, λ y R en la sección S6.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-623,83	2552,34	76,38	280,98	-0,42	-0,83	-0,22
Martes	-141,36	346,15	117,45	222,23	-0,5	-0,7	-0,83
Miércoles	-114,80	258,79	53,01	126,61	-0,44	-0,72	-0,52
Jueves	-53,73	158,37	79,47	234,75	-0,57	-0,83	-0,33
Viernes	-1336,10	4043,69	183,09	437,17	-0,35	-0,53	-0,31
Sábado	-163,44	470,99	105,92	249,39	-0,32	-0,72	-0,37
Domingo	-635,98	2201,24	155,75	313,86	-0,09	-0,29	-0,27
Media	-455,32	2063,21	107,63	279,43	-0,35	-0,66	-0,29

Tabla 20. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S6.

Días	R	
	Media	σ
Lunes	0,8750	0,1129
Martes	0,7960	0,2385
Miércoles	0,9286	0,0722
Jueves	0,8968	0,1344
Viernes	0,9363	0,0723
Sábado	0,8596	0,1736
Domingo	0,9165	0,0926
Media final	0,8958	0,1268

Tabla 21. Media y desviación típica para R para todos los días de la semana en la sección S6.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,36	0,36
b	< 0,05	0,2373	0,2387
λ	< 0,05	0,4878	0,5348
R	< 0,05	0,8481	0,3901

Tabla 22. Test de normalidad, varianza y media para a, b, λ y R en la sección S8.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-50,61	169,41	146,14	295,06	-0,43	-0,78	-0,29
Martes	-4,83	10,49	71,64	176,99	-0,45	-0,67	-0,4
Miércoles	-4,46	15,90	5,38	13,66	-1,09	-1,27	-0,73
Jueves	-36,84	113,56	51,92	143,93	-0,78	-0,93	-0,62
Viernes	-876,66	3036,24	109,32	330,26	-0,5	-0,87	-0,52
Sábado	-919,92	3057,29	217,27	448,14	-0,68	-0,99	-0,63
Domingo	-184,93	655,83	55,22	144,01	-0,16	-0,62	0,18
Media	-299,53	1677,51	91,26	259,32	-0,62	-0,89	-0,46

Tabla 23. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S8.

Días	R	
	Media	σ
Lunes	0,7876	0,1771
Martes	0,8438	0,1908
Miércoles	0,8743	0,1378
Jueves	0,8996	0,1369
Viernes	0,8997	0,1168
Sábado	0,8579	0,1250
Domingo	0,8789	0,1610
Media final	0,8659	0,1499

Tabla 24. Media y desviación típica para R para todos los días de la semana en la sección S8.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,5556	0,5555
b	< 0,05	0,6857	0,6883
λ	< 0,05	0,56	0,9945
R	< 0,05	0,05	< 0,05

Tabla 25. Test de normalidad, varianza y media para a, b, λ y R en la sección S10.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-942,12	2974,09	152,71	351,34	0	-0,34	0,06
Martes	-1209,22	3541,44	155,18	375,47	-0,06	-0,33	-0,1
Miércoles	-1474,86	3752,88	229,08	422,03	-0,14	-0,37	-0,54
Jueves	-179,61	595,74	67,44	185,61	-0,27	-0,66	-0,32
Viernes	-220,20	492,74	101,85	184,51	-0,12	-0,4	-0,19
Sábado	-517,38	1547,13	96,55	237,22	-0,25	-0,76	-0,4
Domingo	-755,11	2464,49	121,34	312,69	-0,15	-0,38	-0,13
Media	-756,49	2513,94	132,01	306,35	-0,14	-0,46	-0,23

Tabla 26. Media y desviación típica para a, b, λ y R para todos los días de la semana en la sección S10.

3. PERIÓDICO "La Provincia"

3.1. Sin secciones

Categorías	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	17/25	19/25	20/25	18/25	19/25	17/25	17/25
Total días	17 + 19 + 20 + 18 + 19 + 17 + 17 = 127 (73%)						

Tabla 27. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Bartlett/Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,2321	0,4541
b	0,1132	0,5139	0,5116
λ	< 0,05	0,733	0,6723
R	< 0,05	0,6743	0,5316

Tabla 28. Test de normalidad, varianza y media para a, b, c y R.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,4088	0,066	2,4904	0,1769	-0,3824	-0,4176	-0,3471
Martes	-1,1793	0,0525	2,3075	0,1467	-0,4053	-0,4421	-0,3632
Miércoles	-1,4741	0,0688	2,5841	0,1863	-0,345	-0,375	-0,315
Jueves	-1,5374	0,073	2,7285	0,1893	-0,3556	-0,4222	-0,3056
Viernes	-1,4844	0,0632	2,7636	0,172	-0,3053	-0,3211	-0,2789
Sábado	-1,2799	0,06	2,347	0,1625	-0,4235	-0,4529	-0,3941
Domingo	-1,4014	0,06	2,6501	0,1599	-0,2824	-0,3	-0,2706
Media final	-1,396	0,0634	2,5546	0,1707	-0,3567	-0,3898	-0,3244

Tabla 29. Media y desviación típica para a, b y λ.

Días	R	
	Media	σ
Lunes	0,9290	0,0352
Martes	0,9179	0,0447
Miércoles	0,9287	0,0346
Jueves	0,9247	0,0490
Viernes	0,9370	0,0316
Sábado	0,9262	0,0422
Domingo	0,9434	0,0249
Media final	0,9294	0,0381

Tabla 30. Media y desviación típica para R.

3.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	7/7	0/0	21/25	0/0	0/0	12/12	8/8	23/23	2/2	14/14
Martes	8/8	0/0	18/24	0/0	0/0	14/14	8/8	22/22	1/1	13/13
Miércoles	5/5	0/0	22/25	0/0	0/0	15/15	6/6	24/25	2/2	10/10
Jueves	2/2	1/1	22/25	0/0	1/1	10/10	8/8	21/21	1/1	14/14
Viernes	7/7	0/0	23/25	0/0	0/0	13/14	6/6	25/25	1/1	18/18
Sábado	6/6	0/0	21/25	0/0	0/0	12/12	9/9	21/21	1/1	13/13
Domingo	10/10	1/1	22/25	0/0	0/0	11/11	6/6	24/24	2/2	13/13
Total días	45	2	149	0	1	87	51	160	10	95

Tabla 31. Relación del nº de días modelados con éxito/sin éxito por día de la semana y sección.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,662	0,5133
b	< 0,05	0,5661	0,644
λ	< 0,05	0,9954	0,8242
R	< 0,05	0,8342	0,8551

Tabla 32. Test de normalidad, varianza y media para a, b, λ y R en la sección S3.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,1445	0,5279	2,4035	1,2555	-0,3684	-0,4059	-0,3842
Martes	-1,1928	0,5338	2,2159	1,0205	-0,4722	-0,4412	-0,3937
Miércoles	-1,4449	0,7365	2,4847	1,2742	-0,4273	-0,47	-0,3428
Jueves	-1,2601	0,7770	2,4331	1,4491	-0,3895	-0,485	-0,4286
Viernes	-1,5428	0,8628	2,9194	1,7511	-0,3095	-0,3952	-0,2684
Sábado	-1,2424	0,5548	2,1608	0,9838	-0,4750	-0,5421	-0,4158
Domingo	-1,4354	1,0055	2,8141	2,5107	-0,3454	-0,285	-0,2869
Media	-1,3307	0,7427	2,5053	1,5617	-0,3964	-0,4313	-0,3579

Tabla 33. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S3.

Días	R	
	Media	σ
Lunes	0,8962	0,0963
Martes	0,9091	0,0648
Miércoles	0,9032	0,0747
Jueves	0,8918	0,0868
Viernes	0,9206	0,0573
Sábado	0,9018	0,0763
Domingo	0,9163	0,0596
Media final	0,9057	0,0739

Tabla 34. Media y desviación típica para R para todos los días de la semana en la sección S3.

Test estadísticos - Sábado y Domingo			
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,6819	0,6764
b	< 0,05	0,7367	0,7278
λ	< 0,05	0,4937	0,8144
R	< 0,05	0,9922	0,9542

Tabla 35. Test de normalidad, varianza y media para a, b, λ y R en la sección S8.

Días	a		b		λ		
	Media	σ	Media	σ	Media	li.	ls.
Lunes	-1,6597	2,8725	3,3715	7,1182	-0,0273	-0,6	-0,4952
Martes	-1,1881	1,2089	55,2669	162,2228	-0,02722	-0,48	-0,2048
Miércoles	-29,0880	134,5929	36,7872	155,8217	-0,0244	-0,2647	-0,1818
Jueves	-34,2661	143,6124	44,2185	165,9423	0	-0,1187	0,23
Viernes	-12,0867	40,3339	21,4248	66,4210	-0,0125	-0,4105	-0,2409
Sábado	-9,2469	34,2172	21,4688	80,7747	-0,0412	-0,2385	-0,4
Domingo	-5,1927	17,5357	11,7762	43,0844	-0,029	-0,6470	-0,2167
Media	-13,1437	76,1659	27,3374	111,8151	-0,2264	-0,3917	-0,2161

Tabla 36. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S8.

Días	R	
	Media	σ
Lunes	0,8199	0,1288
Martes	0,7968	0,1778
Miércoles	0,8424	0,1419
Jueves	0,8436	0,1454
Viernes	0,8122	0,1679
Sábado	0,8213	0,1623
Domingo	0,8180	0,1423
Media final	0,8219	0,1509

Tabla 37. Media y desviación típica para R para todos los días de la semana en la sección S8.

4. SERVICIO "Levante-EMV"

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	6 / 25	6 / 25	8 / 25	5 / 25	5 / 25	4 / 25	5 / 25
Total días	6 + 6 + 8 + 5 + 5 + 4 + 5 = 39 (22%)						

Tabla 38. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Bartlett/Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,3524	0,3479
b	< 0,05	0,358	0,3608
λ	< 0,05	0,2837	0,2226
R	0,4595	0,3494	0,2525

Tabla 39. Test de normalidad, varianza y media para a, b, λ y R.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,8926	0,1537	4,8768	0,6248	0,15	0,13	0,17
Martes	-1,9184	0,2985	4,1901	0,5238	0,017	0,017	0,017
Miércoles	-1,8633	0,2329	4,5613	0,6988	0,09	0,09	0,09
Jueves	-19,7574	40,0079	25,8139	48,3851	0,24	0,24	0,24
Viernes	-1,8740	0,0771	4,6289	0,5712	0,1	0,1	0,1
Sábado	-1,7339	0,3374	4,1494	0,8764	0	0	0
Domingo	-1,6001	0,3041	3,6237	0,6777	-0,1	-0,1	-0,1
Media	-4,1632	13,4019	7,1666	16,1811	0,07	0,07	0,08
Media - sin jueves	-1,8259	0,2503	4,3750	0,7288	0,05	0,0470	0,0529

Tabla 40. Media y desviación típica para a, b y λ.

Días	R	
	Media	σ
Lunes	0,9726	0,0087
Martes	0,9761	0,0058
Miércoles	0,9680	0,0114
Jueves	0,9760	0,0181
Viernes	0,9731	0,0088
Sábado	0,9648	0,0103
Domingo	0,9601	0,0151
Media final	0,9703	0,0112

Tabla 41. Media y desviación típica para R.

4.1.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	20/20	10/10	24/24	22/22	11/11	20/20	6/6	23/23	8/8	24/24
Martes	20/20	14/14	25/25	21/21	10/10	19/19	4/4	24/24	14/14	25/25
Miércoles	20/20	14/14	25/25	22/22	9/9	18/18	5/5	24/24	13/13	25/25
Jueves	17/17	11/11	23/23	22/22	5/5	15/15	4/4	24/24	16/16	24/24
Viernes	23/23	12/12	25/25	23/23	5/5	15/15	6/6	21/21	11/11	25/25
Sábado	15/16	12/12	24/24	24/24	6/6	14/14	13/13	21/21	4/4	25/25
Domingo	21/21	10/10	25/25	20/20	6/6	20/20	9/9	23/23	2/2	24/25
Total días	136	83	171	154	52	121	47	160	68	172

Tabla 42. Nº de días modelados con éxito/sin éxito por días y sección.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,4311	0,4313
b	< 0,05	0,6388	0,6398
λ	< 0,05	0,1831	0,8064
R	< 0,05	< 0,05	< 0,05

Tabla 43. Test de normalidad, varianza y media para a, b, λ y R en la sección S1.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1269,87	3950,72	138,30	409,27	-0,16	-0,36	-0,19
Martes	-1233,22	3787,94	131,36	398,87	-0,45	-0,65	-0,52
Miércoles	-10088	414,88	36,72	126,46	-0,57	-0,70	-0,55
Jueves	-7,6647	14,0639	10,36	26,57	-0,4	-0,56	-0,04
Viernes	-31,9028	131,72	36,15	154,95	-0,59	-0,76	-0,39
Sábado	-942,62	3756,72	136,06	393,37	-0,45	-0,61	-0,57
Domingo	-405,75	1384,19	66,57	213,99	-0,71	-1,07	-0,68
Media	-558,73	2515,61	78,18	279,49	-0,48	-0,68	-0,43

Tabla 44. Media y desviación típica a, b y λ todos los días en la sección S1.

Días	R	
	Media	σ
Lunes	0,9302	0,0896
Martes	0,9092	0,0892
Miércoles	0,9319	0,0714
Jueves	0,9045	0,1055
Viernes	0,9527	0,0775
Sábado	0,8313	0,1860
Domingo	0,8627	0,1401

Tabla 45. Media y desviación típica R todos los días en la sección S1.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,4449	0,4454
b	< 0,05	0,4632	0,4656
λ	< 0,05	0,7183	0,6967
R	< 0,05	0,5408	0,75

Tabla 46. Test de normalidad, varianza y media para a, b, λ y R en la sección S3.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,3252	1,4214	1,9527	1,7893	-0,95	-1,02	-0,83
Martes	-1,4697	2,1619	2,1009	3,5648	-0,93	-1,07	-0,79
Miércoles	-473,80	2353,61	69,4225	336,6052	-0,81	-0,88	-0,81
Jueves	-1,7530	2,1141	2,11	1,7892	-0,89	-0,98	-0,78
Viernes	-3,0985	3,7781	3,6628	6,0148	-0,58	-0,71	-0,42
Sábado	-7,5601	27,3145	5,8477	17,0601	-0,7	-0,86	-0,50
Domingo	-2,7687	3,9714	2,9284	4,5917	-0,73	-0,84	-0,58
Media	-71,8252	899,99	12,7989	128,84	-0,79	-0,91	-0,67

Tabla 47. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S3.

Días	R	
	Media	σ
Lunes	0,9343	0,0404
Martes	0,9337	0,0404
Miércoles	0,9405	0,0443
Jueves	0,9155	0,0776
Viernes	0,9209	0,0778
Sábado	0,9233	0,0679
Domingo	0,9237	0,0403
Media	0,9275	0,0573

Tabla 48. Media y desviación típica para R para todos los días de la semana en la sección S3.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,7086	0,7083
b	< 0,05	0,4007	0,4009
λ	< 0,05	0,4275	0,2127
R	< 0,05	0,08	0,2403

Tabla 49. Test de normalidad, varianza y media para a, b, λ y R en la sección S4.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-646,59	2197,62	113,94	270,74	0	-0,32	0,004
Martes	-320,45	1441,85	46,05	198,04	-0,04	-0,64	-0,31
Miércoles	-34,59	77,20	42,09	114,99	-0,007	-0,32	0,06
Jueves	-10,14	20,05	7,76	13,35	-0,03	-0,52	-0,17
Viernes	-623,22	2933,41	70,23	295,15	-0,04	-0,49	-0,38
Sábado	-737,38	3126,60	145,83	376,49	-0,003	-0,27	-0,22
Domingo	-13,26	51,44	22,43	90,52	-0,08	-0,94	-0,66
Media	-352,17	77,20	65,81	232,56	-0,28	-0,49	-0,23

Tabla 50. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S4.

Días	R	
	Media	σ
Lunes	0,9188	0,0979
Martes	0,9356	0,0672
Miércoles	0,9540	0,0335
Jueves	0,9326	0,1007
Viernes	0,9688	0,0317
Sábado	0,9511	0,0627
Domingo	0,9528	0,0439
Media	0,9450	0,0683

Tabla 51. Media y desviación típica para R para todos los días de la semana en la sección S4.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,4528	0,4529
b	< 0,05	0,5061	0,5062
λ	< 0,05	0,8652	0,5717
R	< 0,05	0,7697	0,6405

Tabla 52. Test de normalidad, varianza y media para a, b, λ y R en la sección S6.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-5,6973	20,15	7,1365	17,1290	-0,75	-1,73	-0,29
Martes	-391,5877	1381,18	87,9154	232,27	-0,38	-1,33	-0,28
Miércoles	-9,8551	33,11	37,8308	130,07	-0,5	-0,5	-0,5
Jueves	-9,9662	23,39	2,4108	2,9538	-0,94	-1,71	-0,59
Viernes	-994,2207	3846,13	135,9090	391,64	-0,53	-0,57	-0,49
Sábado	-179,18	520,87	80,2488	184,41	-0,11	-0,67	-0,09
Domingo	-60,0826	173,35	79,1583	234,11	-0,57	-1,55	-0,4
Media	-219,04	1466,37	60,13	206,64	-0,55	-1,29	-0,38

Tabla 53. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S6.

Días	R	
	Media	σ
Lunes	0,8482	0,1296
Martes	0,8238	0,1723
Miércoles	0,9048	0,1489
Jueves	0,8907	0,1272
Viernes	0,8692	0,1598
Sábado	0,8343	0,2062
Domingo	0,8889	0,1322
Media	0,8658	0,1528

Tabla 54. Media y desviación típica para R para todos los días de la semana en la sección S6.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,5922	0,5914
b	< 0,05	0,2375	0,2321
λ	< 0,05	0,4841	0,1345
R	< 0,05	0,8629	0,7267

Tabla 55. Test de normalidad, varianza y media para a, b, λ y R en la sección S8.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-2,4458	4,28	2,7096	4,94	-0,96	-1,17	-0,79
Martes	-5,3445	12,55	4,5522	10,01	-0,79	-0,9	-0,62
Miércoles	-33,4589	85,45	20,7305	68,79	-0,74	-0,93	-0,63
Jueves	-539,8537	2592,51	94,9245	287,34	-0,15	-0,43	-0,21
Viernes	-26,6524	75,44	11,4156	24,14	-0,44	-0,63	-0,18
Sábado	-22,6695	80,71	8,8402	26,04	-0,72	-0,63	-0,6
Domingo	-295,5682	1414,91	41,8399	194,64	-0,91	-1,15	-0,78
Media	-136,76	1139,39	27,1337	138,23	-0,67	-0,87	-0,55

Tabla 56. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S8.

Días	R	
	Media	σ
Lunes	0,9094	0,0887
Martes	0,9105	0,0914
Miércoles	0,8962	0,1115
Jueves	0,8877	0,1218
Viernes	0,9304	0,0749
Sábado	0,8936	0,1169
Domingo	0,8772	0,1256
Media	0,9004	0,1054

Tabla 57. Media y desviación típica para R para todos los días de la semana en la sección S8.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,4253	0,4271
b	< 0,05	0,5001	0,5041
λ	< 0,05	0,8075	0,8145
R	< 0,05	0,4116	0,0943

Tabla 58. Test de normalidad, varianza y media para a, b, λ y R en la sección S10.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,1763	4,75	1,8334	4,47	-1,34	-1,49	-1,19
Martes	-0,9117	1,64	2,2815	4,79	-1,12	-1,24	-1,02
Miércoles	-15,7627	72,71	7,2557	28,27	-1,36	-1,57	-1,21
Jueves	-1,6647	4,92	4,9413	15,17	-1,04	-1,3	-0,82
Viernes	-1,1610	4,3676	1,0909	0,91	-1,39	-1,48	-1,18
Sábado	-0,5659	0,72	1,1635	0,84	-1,24	-1,38	-1,03
Domingo	-0,5725	0,9248	1,3309	1,30	-1,16	-1,35	-1
Media	-3,1509	27,91	2,8449	12,42	-1,24	-1,4	-1,06

Tabla 59. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S10.

Días	R	
	Media	σ
Lunes	0,9216	0,0293
Martes	0,9098	0,0596
Miércoles	0,8745	0,0724
Jueves	0,9017	0,0471
Viernes	0,8992	0,0727
Sábado	0,8985	0,0624
Domingo	0,9195	0,0482
Media	0,9034	0,0589

Tabla 60. Media y desviación típica para R para todos los días de la semana en la sección S10.

5. PERIÓDICO "La Nueva España"

5.1. Sin secciones

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	4/26	10/26	7/26	10/26	12/26	10/26	7/26
Total días	4 + 10 + 7 + 10 + 12 + 10 + 7 = 60 (33%)						

Tabla 61. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Bartlett/Levene (Varianza)	F-Anova (Media)
a	0,8496	0,5416	0,1553
b	0,877	0,6152	0,4703
λ	< 0,05	0,3329	0,2005
R	0,2394	0,0756	< 0,05

Tabla 62. Test de normalidad, varianza y media para a, b, λ y R todos los días menos el sábado.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,4751	0,0220	3,9536	0,0801	0,075	0,075	0,075
Martes	-1,5956	0,0266	3,7629	0,1005	-0,04	-0,04	-0,02
Miércoles	-1,5937	0,0281	3,7825	0,1022	-0,01	-0,01	-0,01
Jueves	-1,5945	0,0236	3,8308	0,0868	-0,01	-0,01	-0,01
Viernes	-1,5661	0,0224	3,9601	0,0828	-0,05	-0,05	-0,05
Domingo	-1,4389	0,0223	3,7149	0,0846	0,01	0,01	0,01
General	-1,5525	0,0242	3,8341	0,2699	0,02	0,02	0,02

Tabla 63. Media y desviación típica para a, b y λ.

Días	R	
	Media	σ
Lunes	0,9808	0,0073
Martes	0,9717	0,0076
Miércoles	0,9731	0,0067
Jueves	0,9798	0,0033
Viernes	0,9806	0,0063
Domingo	0,9755	0,0028
Media	0,9769	0,0068

Tabla 64. Media y desviación típica para R.

Día	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Sábado	-1,5326	0,0240	4,0959	0,0884	0,02	0,02	0,02

Tabla 65. Media y desviación típica para a, b y λ para el sábado.

R		
Día	Media	σ
Sábado	0,9775	0,0053

Tabla 66. Media y desviación típica para R para el sábado.

5.2. Con secciones

Días	Secciones									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lunes	26/26	X	24/25	26/26	21/21	21/21	14/14	X	14/14	26/26
Martes	26/26	X	25/26	25/25	20/21	26/26	26/26	X	12/12	25/26
Miércoles	26/26	X	25/26	23/23	21/21	23/23	26/26	X	12/12	26/26
Jueves	26/26	X	25/26	25/25	20/20	23/23	26/26	X	12/12	25/26
Viernes	26/26	X	25/26	23/23	19/19	22/22	23/26	X	9/9	26/26
Sábado	26/26	X	26/26	23/23	20/20	18/19	25/26	X	4/4	22/22
Domingo	26/26	X	24/25	26/26	18/18	22/22	25/26	X	2/2	26/26
Total días	182	X	174	171	139	155	165	X	65	176

Tabla 67. Relación del nº de días modelados con éxito/sin éxito por categoría y sección.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,6554	0,6573
b	< 0,05	0,6836	0,6896
λ	< 0,05	0,617	0,3032
R	< 0,05	0,4062	0,3598

Tabla 68. Test de normalidad, varianza y media para a, b, λ y R en la sección S1.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-2,21	5,45	2,1854	1,66	-0,79	-0,92	-0,64
Martes	-20,7398	73,95	22,8731	66,82	-0,36	-0,6	-0,19
Miércoles	-84,6854	410,55	20,8389	82,54	-0,43	-0,56	-0,37
Jueves	-152,98	757,24	31,0773	122,16	-0,54	-0,72	-0,46
Viernes	-36,8388	115,70	9,7637	20,76	-0,19	-0,36	-0,05
Sábado	-11,9103	36,99	21,0824	72,44	-0,33	-0,5	-0,21
Domingo	-20,1458	44,01	35,5832	72,34	-0,11	-0,29	-0,01
Media	-47,0728	328,70	20,49	72,39	-0,39	-0,56	-0,28

Tabla 69. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S1.

Días	R	
	Media	σ
Lunes	0,9351	0,0309
Martes	0,9335	0,0464
Miércoles	0,9481	0,0490
Jueves	0,9441	0,0329
Viernes	0,9497	0,0478
Sábado	0,9493	0,0296
Domingo	0,9548	0,0242
Media	0,9449	0,0385

Tabla 70. Media y desviación típica para R para todos los días de la semana en la sección S1.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,1681	0,3812
b	< 0,05	0,6123	0,7263
λ	< 0,05	0,9358	0,7679
R	< 0,05	0,1926	0,06

Tabla 71. Test de normalidad, varianza y media para a, b, λ y R en la sección S3.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,0392	0,4851	2,2954	1,0053	-0,49	-0,54	-0,45
Martes	-1,2005	0,8630	2,1256	0,9357	-0,58	-0,63	-0,56
Miércoles	-1,2401	0,9909	2,0244	1,0383	-0,72	-0,78	-0,69
Jueves	-1,5171	1,2061	2,4895	0,9691	-0,47	-0,5	-0,44
Viernes	-1,4970	1,1103	2,4407	1,1024	-0,49	-0,54	-0,47
Sábado	-1,1521	0,7563	2,3974	1,5771	-0,57	-0,66	-0,5
Domingo	-1,0982	0,6720	2,3549	0,8602	-0,44	-0,49	-0,4
Media	-1,2507	0,90	2,3043	1,0896	-0,54	-0,59	-0,5

Tabla 72. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S3.

Días	R	
	Media	σ
Lunes	0,9434	0,0284
Martes	0,9401	0,0249
Miércoles	0,9480	0,0176
Jueves	0,9527	0,0219
Viernes	0,9410	0,0302
Sábado	0,9299	0,0344
Domingo	0,9359	0,0240
Media	0,9415	0,0269

Tabla 73. Media y desviación típica para R para todos los días de la semana en la sección S3.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,2631	0,2643
b	< 0,05	0,258	0,2648
λ	< 0,05	< 0,05	0,9553
R	< 0,05	0,5094	0,6592

Tabla 74. Test de normalidad, varianza y media para a, b, λ y R en la sección S4.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-8,8649	13,6261	4,8151	4,5687	-0,22	-0,41	-0,06
Martes	-7,8610	15,6599	13,1709	29,5853	-0,16	-0,29	-0,02
Miércoles	-95,4322	253,7066	58,2448	133,1917	-0,1	-0,26	-0,11
Jueves	-8,8649	13,6261	4,8151	4,5687	-0,22	-0,41	-0,06
Viernes	-889,0155	3622,2365	100,5351	351,8665	-0,17	-0,27	-0,1
Sábado	-614,2332	1766,5157	130,8363	373,2611	-0,15	-0,38	-0,08
Domingo	-157,1702	733,9061	49,1224	213,9503	-0,41	-0,51	-0,27
Media	-246,2578	1520,532	50,5218	214,7231	-0,21	-0,36	-0,1

Tabla 75. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S4.

Días	R	
	Media	σ
Lunes	0,9333	0,0535
Martes	0,9431	0,0631
Miércoles	0,9579	0,0318
Jueves	0,9333	0,0535
Viernes	0,9566	0,0336
Sábado	0,9447	0,0546
Domingo	0,9385	0,0955
Media	0,9436	0,0589

Tabla 76. Media y desviación típica para R para todos los días de la semana en la sección S4.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,1644	0,1645
b	< 0,05	0,2308	0,2313
λ	< 0,05	0,1210	0,6757
R	< 0,05	0,6006	0,6917

Tabla 77. Test de normalidad, varianza y media para a, b, λ y R en la sección S5.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-21,5248	78,7091	27,0478	111,9723	-0,46	-0,71	-0,25
Martes	-1128,8375	3421,3138	180,1553	457,0156	-0,125	-0,3	-0,33
Miércoles	-11,8022	25,8313	11,8911	33,6929	-0,63	-0,72	-0,51
Jueves	-7,7376	19,3199	3,5387	4,4758	-0,7	-0,95	-0,43
Viernes	-56,4056	140,8976	70,8947	159,5387	-0,36	-0,64	-0,18
Sábado	-1188,2762	3447,4661	135,6686	374,8022	-0,325	-0,43	-0,28
Domingo	-1749,3825	5103,2537	182,2983	498,5490	-0,19	-0,37	-0,23
Media	-573,79	2630,059	85,1319	298,7526	-0,4	-0,59	-0,32

Tabla 78. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S5.

Días	R	
	Media	σ
Lunes	0,9344	0,0786
Martes	0,9311	0,1103
Miércoles	0,9609	0,0414
Jueves	0,9105	0,1109
Viernes	0,95	0,662
Sábado	0,9219	0,1267
Domingo	0,9271	0,1049
Media	0,9339	0,0944

Tabla 79. Media y desviación típica para R para todos los días de la semana en la sección S5.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,4615	0,4617
b	< 0,05	0,5459	0,5481
λ	< 0,05	0,483	0,6366
R	< 0,05	0,1532	0,1773

Tabla 80. Test de normalidad, varianza y media para a, b, λ y R en la sección S6.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-534,0967	2433,0873	62,3248	269,8804	-0,48	-0,74	-0,12
Martes	-30,2822	89,2319	47,3327	136,6162	-0,66	-0,77	-0,63
Miércoles	-4,8371	7,5334	2,2830	2,9954	-0,97	-1,1	-0,72
Jueves	-9,3106	16,1934	7,7723	13,7818	-0,55	-0,75	-0,2
Viernes	-10,3788	24,9554	3,5135	4,6028	-0,62	-0,75	-0,49
Sábado	-114,7464	479,8459	24,9561	98,7722	-0,74	-0,97	-0,48
Domingo	-44,6041	121,3833	18,1913	44,2775	-0,44	-0,62	-0,22
Media	-100,6698	910,1529	23,8546	119,714	-0,64	-0,81	-0,41

Tabla 81. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S6.

Días	R	
	Media	σ
Lunes	0,8767	0,1039
Martes	0,9052	0,1289
Miércoles	0,9002	0,1180
Jueves	0,8831	0,1190
Viernes	0,9569	0,0361
Sábado	0,8828	0,1124
Domingo	0,9206	0,0885
Media	0,9042	0,1068

Tabla 82. Media y desviación típica para R para todos los días de la semana en la sección S6.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,7654	0,7121
b	< 0,05	0,6567	0,8512
λ	< 0,05	0,2832	0,5999
R	< 0,05	0,05	< 0,05

Tabla 83. Test de normalidad, varianza y media para a, b, λ y R en la sección S7 todos los días menos el lunes.

Día	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1499,61	3851,815	177,2605	416,3	-0,16	-0,37	-0,28

Tabla 84. Media y desviación típica para a, b y λ del lunes de la sección S7.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Martes	-1,3289	1,5836	2,8819	3,2987	-0,61	-0,7	-0,55
Miércoles	-1,6745	2,8623	3,8747	9,1166	-0,61	-0,7	-0,53
Jueves	-1,5242	1,7936	2,5237	1,3674	-0,46	-0,53	-0,41
Viernes	-1,8895	1,2153	3,2177	1,5954	-0,28	-0,29	-0,26
Sábado	-1,1436	0,9597	2,4916	1,8955	-0,64	-0,67	-0,62
Domingo	-1,3478	1,1543	2,6896	1,7801	-0,48	-0,52	-0,46
Media	-1,4799	1,7190	2,9459	4,2036	-0,52	-0,58	-0,48

Tabla 85. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S7.

Días	R	
	Media	σ
Martes	0,9569	0,0315
Miércoles	0,9557	0,0317
Jueves	0,9505	0,0304
Viernes	0,9701	0,0151
Sábado	0,9742	0,0113
Domingo	0,9682	0,0160

Tabla 86. Media y desviación típica de R para todos los días de la semana en la sección S7.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk	Levene	F-Anova
	(Normal)	(Varianza)	(Media)
a	< 0,05	0,9483	0,9485
b	< 0,05	0,9403	0,9408
λ	< 0,05	0,6359	0,9712
R	< 0,05	0,9638	0,8738

Tabla 87. Test de normalidad, varianza y media para a, b, λ y R en la sección S9.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1499,6168	3851,8155	177,2605	416,3001	-0,16	-0,37	-0,28
Martes	-830,2901	2012,2992	126,8336	271,5921	-0,18	-0,29	0,02
Miércoles	-830,2901	2012,2992	126,8336	271,5921	-0,18	-0,29	0,02
Jueves	-1465,0774	3234,8372	176,4250	363,6555	-0,06	-0,11	0
Viernes	-1743,6817	4870,5962	229,0818	508,1675	-0,08	-0,23	-0,45
Sábado	-1,7362	2,0264	2,0417	1,7658	-0,72	-1,1	-0,42
Domingo	-3,4073	0,5399	1,4583	0,5182	-0,65	-0,8	-0,5
Media	-1141,683	3070,205	149,4704	344,9627	-0,17	-0,33	-0,16

Tabla 88. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S9.

Días	R	
	Media	σ
Lunes	0,9213	0,1486
Martes	0,9476	0,1016
Miércoles	0,9476	0,1016
Jueves	0,9342	0,1094
Viernes	0,9420	0,0899
Sábado	0,8692	0,1099
Domingo	0,8668	0,1883
Media	0,9314	0,1128

Tabla 89. Media y desviación típica de R para todos los días de la semana en la sección S9.

	Test estadísticos - Sábado y Domingo		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,4113	0,4115
b	< 0,05	0,4353	0,4352
λ	< 0,05	0,4881	0,4881
R	< 0,05	< 0,05	0,06

Tabla 90. Test de normalidad, varianza y media para a, b, λ y R en la sección S10.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-0,0856	0,2873	0,6634	0,7271	-1,9545	-1,92	-1,65
Martes	-0,9650	3,4506	2,7372	8,0666	-1,875	-1,83	-1,08
Miércoles	-0,0932	0,2351	0,7965	0,8027	-1,8696	-2	-1,46
Jueves	-9,5869	46,1389	19,1174	87,7832	-1,6667	-2	-1,24
Viernes	-0,2947	0,6798	4,6098	15,9336	-1,875	-1,9	-1,07
Sábado	-0,1539	0,3229	0,9363	1,1138	-2	-1,94	-1,37
Domingo	-0,15	0,3006	0,8914	0,9371	-2	-1,95	-1,41
Media	-1,6102	17,4462	4,2497	33,7802	-1,8864	-1,94	-1,33

Tabla 91. Media y desviación típica para a, b y λ para todos los días de la semana en la sección S10.

Días	R	
	Media	σ
Lunes	0,8784	0,0510
Martes	0,9107	0,0346
Miércoles	0,8709	0,0688
Jueves	0,8786	0,0936
Viernes	0,8823	0,0813
Sábado	0,8961	0,0453
Domingo	0,9174	0,0365
Media	0,8905	0,063

Tabla 92. Media y desviación típica para a, b, λ y R para todos los días de la semana en la sección S10.

6. PERIÓDICO "Superdeporte"

Días	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Nº días	18/24	23/25	21/25	17/25	18/25	22/25	21/25
Total días	18 + 23 + 21 + 17 + 18 + 22 + 21 = 140 (86%)						

Tabla 93. Nº de días modelados con éxito/sin éxito por días de la semana.

	Test estadísticos		
	Shapiro-Wilk (Normal)	Levene (Varianza)	F-Anova (Media)
a	< 0,05	0,2285	0,2201
b	< 0,05	0,3645	0,2949
λ	< 0,05	0,8252	0,2544
R	< 0,05	0,3643	0,2032

Tabla 94. Test de normalidad, varianza y media para a, b, λ y R.

Días	a		b		λ		
	Media	σ	Media	σ	Media	l.i.	l.s.
Lunes	-1,6492	0,0872	2,7865	0,2041	0	0	0
Martes	-1,6849	0,0981	2,5552	0,2310	0	-2	0
Miércoles	-1,6770	0,1024	2,7047	0,2364	0	0	-0,5
Jueves	-1,7268	0,1038	2,4286	0,2333	-1,5	-2	0
Viernes	-3,8065	0,4274	1,3996	0,9190	0	-0,6	0
Sábado	-1,9030	0,1396	2,2744	0,3197	0	-1	0
Domingo	-1,3844	0,0909	2,1609	0,2188	-1	-1,5	0
General	-1,9462	0,1458	2,3401	0,3292	-1	-0,17	1

Tabla 95. Media y desviación típica para a, b y λ por día.

Categorías	R	
	Media	σ
Lunes	0,9356	0,0283
Martes	0,9175	0,05156
Miércoles	0,9197	0,0327
Jueves	0,9102	0,05778
Viernes	0,8986	0,0556
Sábado	0,9062	0,0459
Domingo	0,8962	0,0647
Media final	0,9120	0,0481

Tabla 96. Media y desviación típica de R por día.