



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES

**UNIVERSIDAD DE VALLADOLID**

**ESCUELA DE INGENIERIAS INDUSTRIALES**

**GRADO EN INGENIERÍA DE ORGANIZACIÓN INDUSTRIAL**

# **Análisis de los flujos de tránsito en una calle peatonal**

**Autor:**

**Lobos Collantes, Fernando**

**Tutor:**

**Gordaliza Ramos, Alfonso**

**Fernández Temprano, Miguel Alejandro**

**Estadística e Investigación Operativa**

**Valladolid, septiembre de 2015.**



## ÍNDICE

1. Resumen y palabras clave .....	5
2. Introducción .....	7
2.1. Motivación y justificación .....	7
2.2. Objetivos .....	8
2.2.1. Subobjetivos .....	8
2.3. Localización .....	9
PARTE 1. FUNCIONAMIENTO DEL SISTEMA DE DETECCIÓN DE PEATONES.	
3. El Semáforo Inteligente .....	13
3.1. Introducción .....	13
3.2. Funcionamiento del Semáforo Inteligente .....	13
PARTE 2. MARCO TEÓRICO	
4. Metodologías estadísticas empleadas .....	19
4.1. Datos funcionales .....	19
4.2. ANOVA funcional .....	22
4.3. Análisis clúster funcional .....	24
4.4. Análisis de componentes principales (ACP) funcional .....	25
4.5. Test chi-cuadrado de homogeneidad .....	27
PARTE 3. DEPURACIÓN Y ANALISIS PRELIMINAR	
5. Depuración de datos y análisis preliminar .....	31
5.1. Depuración de datos .....	31
5.1.1. Introducción a la depuración de datos .....	31
5.1.1.1. Glosario de términos .....	31
5.1.2. Depuración de conteos de peatones .....	32
5.1.3. Depuración de los tiempos de espera .....	36
5.2. Análisis estadístico descriptivo preliminar .....	39
5.2.1. Análisis estadístico descriptivo de los conteos de peatones .....	39
5.2.2. Análisis estadístico descriptivo de los tiempos de espera .....	50
PARTE 4. ANÁLISIS ESTADÍSTICOS, CONCLUSIONES y LÍNEAS FUTURAS.	
6. Análisis mediante procedimientos funcionales .....	51

6.1. Análisis ANOVA funcional .....	51
6.1.1. ANOVA de los conteos totales .....	52
6.1.2. ANOVA de la espera total .....	58
6.2. Análisis clúster funcional .....	66
6.2.1. Análisis clúster de los conteos totales .....	66
6.2.2. Análisis clúster de la espera total .....	72
6.2.3. Relación entre los clústeres de los conteos totales y esperas totales.....	77
6.3. Análisis de ACP funcional .....	80
6.3.1. ACP de los conteos totales.....	80
6.3.2. ACP de la espera total .....	83
7. Análisis de los tiempos de espera .....	87
8. Conclusiones globales y Líneas Futuras.....	95
9. Apéndices: Implementación en R.....	99
9.1. ANOVA funcional.....	99
9.2. Análisis clúster funcional.....	101
9.3. ACP funcional .....	105
9.4. Test chi-cuadrado de homogeneidad para 2 muestras .....	106
10. Bibliografía.....	109

## 1. Resumen y palabras clave

**Resumen:** La calle Santiago es la calle más céntrica y una de las más comerciales de Valladolid. En este TFG se analiza el comportamiento del flujo de peatones en la calle Santiago a su paso por el cruce con la calle Claudio Moyano mediante técnicas estadísticas avanzadas basadas en datos funcionales como clustering, Análisis de Componentes Principales (ACP) o Análisis de la Varianza (ANOVA). También se estudiará la efectividad del semáforo inteligente ubicado en el cruce en la reducción del tiempo de esperas de peatones.

**Palabras clave:** flujo peatonal, datos funcionales, ACP, ANOVA, análisis clúster.

**Abstract.** Santiago Street is the most central Street and one of the most commercial areas in Valladolid. In this TFG, it is analysed the pedestrians flow behaviour at the crossing between Santiago Street and Claudio Moyano Street using advanced statistical techniques based on functional data like Clustering, Principal Component Analysis (PCA) or Analysis of Variance (ANOVA). It is also studied the effectivity of the smart traffic lights in reducing pedestrians waiting times at the crossing.

**Key words:** pedestrian flow, functional data, PCA, ANOVA, cluster analysis.



## **2. Introducción.**

### **2.1 Motivación y justificación**

La tendencia generalizada de los últimos años en los planes de urbanismo de peatonalizar el centro de las ciudades con grandes aceras obliga a realizar un análisis exhaustivo del comportamiento de los peatones en este tipo de calles. La calle Santiago en Valladolid es un claro ejemplo de esta peatonalización siendo la primera de este tipo en Valladolid.

En el cruce donde se estudiarán los flujos peatonales existe un carril de vehículos que circula en un único sentido. Debido a que en ciertos momentos el tráfico en esta zona es escaso, hay una gran cantidad de peatones cruzando el semáforo en rojo. Se ha llegado a reportar que aproximadamente dos tercios de los peatones que cruzan el paso lo hacían cuando el semáforo se encontraba en rojo. Este comportamiento generalizado de los peatones aumenta el riesgo de atropello en la zona por lo que se instaló un semáforo inteligente en el paso de peatones para reducir el tiempo de espera de los peatones y al mismo tiempo, disminuir el número de peatones que crucen en rojo.

En este contexto de la calle Santiago se analizará por franjas horarias y según el día:

- La afluencia de peatones.
- El número de personas que se detiene en el semáforo de la calle
- El tiempo de espera frente al semáforo.

Este Trabajo Fin de Grado ha sido posible gracias a la colaboración con CIDAUT (Fundación para la Investigación y Desarrollo en Transporte y Energía) y el Ayuntamiento de Valladolid debido al suministro de datos del semáforo inteligente que tienen instalado en la Calle Santiago de Valladolid.

También se agradece el tiempo dedicado y la ayuda prestada por CIDAUT para la correcta interpretación de los datos y del funcionamiento del semáforo inteligente. Se llevaron a cabo una reunión presencial y múltiples conversaciones a través de correo electrónico y teléfono. En todos esos contactos se resolvieron las dudas que fueron apareciendo según se iba avanzando en la depuración de los datos, así como se proporcionaron datos adicionales del semáforo para la realización de otros análisis que se consideraban de interés.

## **2.2. Objetivos**

Los objetivos del estudio son:

- Analizar el flujo de peatones en una calle peatonal como la calle Santiago identificando patrones que se repitan diariamente por franjas horarias así como clasificando los diversos días según la afluencia de público, tratando de averiguar las razones según las cuales se producen los distintos comportamientos del flujo peatonal.
- Evaluar la efectividad del semáforo inteligente para reducir los tiempos de espera de los peatones en el semáforo.

### **2.2.1 Subobjetivos**

Asimismo, las conclusiones de este estudio pueden ayudar a la toma de decisiones a muchos colectivos y/o organismos de la ciudad constituyendo los siguientes subobjetivos:

- Proporcionar información del flujo de peatones para ayudar a la coordinación de eventos públicos de gran afluencia en la calle Santiago, la plaza Zorrilla y sus alrededores. Ejemplos de esos eventos pueden ser manifestaciones, actuaciones de teatro en la calle, carreras populares, procesiones de Semana Santa, marchas solidarias, etc.
- Ayudar al Ayuntamiento a la planificación de la movilidad urbana.
- Proporcionar información comercial y estratégica de valor para comercios ya que la información sobre las franjas horarias y de los días con mayor afluencia puede resultar muy útil para adecuar su horario comercial y/o tasar correctamente el precio del alquiler en función de los clientes potenciales.
- Coordinar los servicios de limpieza y de entrega de paquetería para realizar sus trabajos en aquellos momentos en los que menos se pueda afectar a los peatones.

### 2.3. Localización

El control del paso y espera de los peatones se realiza con el semáforo inteligente que se encuentra en el cruce entre la calle Santiago y las calles Claudio Moyano y Doctrinos, junto a la plaza Zorrilla, en la zona centro del término municipal de Valladolid, de la comunidad autónoma de Castilla y León en España. La ubicación del semáforo inteligente en el mapa de Valladolid se muestra en la figura 2.1:



*Figura 2.1. Mapa de la ubicación del semáforo inteligente en Valladolid*

En las figuras 2.2 y 2.3 se adjuntan fotografías del paso desde ambos lados del paso de peatones.



*Figura 2.2. Fotografía del cruce entre la Calle Santiago y la Calle Claudio Moyano desde la acera más próxima a la Plaza Mayor*



*Figura 2.3. Fotografía del cruce entre la Calle Santiago y la Calle Claudio Moyano desde la acera más próxima a la Plaza de Zorrilla*

En las figuras 2.4 y 2.5 se puede observar el paso desde las calles Claudio Moyano y Doctrinos, respectivamente.



*Figura 2.4. Fotografía del paso desde la calle Claudio Moyano.*



*Figura 2.5. Fotografía del paso desde la calle Doctrinos.*

La figura 2.6 muestra una foto desde satélite del paso donde se encuentra el semáforo. En la zona izquierda de la fotografía se encuentra la calle Doctrinos, en la zona derecha está la calle Claudio Moyano y la calle Santiago se encuentra en zona central de la foto.



*Figura 2.6. Vista desde el satélite del cruce entre las calles Santiago, Claudio Moyano y Doctrinos. Fuente: Google Earth*

Las coordenadas exactas de la ubicación del cruce son: latitud  $41.6490495^{\circ}$  Norte y longitud  $4.7293505^{\circ}$  Oeste.

### **3. El Semáforo Inteligente**

#### **3.1. Introducción**

Como se ha comentado en el capítulo anterior los datos del flujo peatonal se obtienen a partir del sistema de detección de peatones instalado en el semáforo inteligente de la calle Santiago por la fundación CIDAUT, en colaboración con Telvent y el Ayuntamiento de Valladolid.

La implantación del semáforo inteligente se debe a la apuesta del Ayuntamiento de Valladolid por las nuevas tecnologías y el concepto de Smart City (ciudad inteligente). Fue instalado en pruebas en abril de 2014 y se ha reportado una reducción del 21% del tiempo de espera y un 23% de las aglomeraciones. Se trata de un proyecto pionero y ciudades como Londres han implantado tecnologías similares posteriormente en sus cruces. El alcalde de Valladolid en 2014, Francisco Javier León de la Riva, manifestó la intención de instalar más semáforos inteligentes en otros pasos con gran afluencia de peatones como en la calle Ferrari, la plaza Zorrilla, Colón, España o Fuente Dorada debido al éxito del sistema instalado en la calle Santiago.

El coste de su instalación fue de 13170 € más IVA.

#### **3.2. Funcionamiento del semáforo inteligente**

El sistema del semáforo inteligente está formado por 4 cámaras, 2 a cada lado del paso de peatones y un módulo de control. En las figuras 3.1 y 3.2, se puede apreciar la ubicación de las 4 cámaras en 4 farolas (2 a cada lado del paso).



*Figura 3.1. Ubicación de las 2 cámaras situadas en el lado más próximo a la Plaza Zorrilla*



*Figura 3.2. Ubicación de las 2 cámaras situadas en el lado más próximo a la Plaza Mayor*

Las figuras 3.3 y 3.4 muestran con detalle las cámaras que realizan el conteo de peatones. En la figura 3.4, además de la cámara se puede ver 1 de los 2 focos LED que entran en funcionamiento cuando el sistema detecta a algún peatón dentro

del horario del alumbrado nocturno con el fin de mejorar la visibilidad en el paso. Hay un foco LED a cada lado del cruce.



*Figura 3.3. Detalle de la cámara*

*Figura 3.4. Detalle de la cámara y el foco LED*

Las cámaras son capaces de detectar a los peatones y de diferenciar si el peatón se encuentra esperando o moviéndose. De hecho, las cámaras siguen la trayectoria que realiza el peatón en el paso. Estos datos se pueden registrar para realizar un posterior análisis como el que se realizará en el presente TFG.

A partir de la cantidad de peatones que se encuentran esperando, el módulo de control puede modificar el ciclo del semáforo. El ciclo del semáforo dura 100 segundos y se subdivide en las siguientes fases para peatones y vehículos de la siguiente manera:

- **Peatones:** 20 s en fase verde + 6 en verde intermitente + 74 s en fase roja
- **Vehículos:** 61 s en fase verde + 3 en fase ámbar + 36 s en fase roja

Hay que indicar que por motivos de seguridad que el semáforo de peatones no pasa de rojo a verde inmediatamente cuando el semáforo de vehículos se pone en rojo, sino que se espera 4 segundos. De la misma forma cuando el semáforo de peatones se pone en rojo el semáforo de vehículos se pasa a verde cuando transcurren 6 segundos. Entonces hay 10 segundos por ciclo (o 13 si se incluyen

los 3 segundos en ámbar) en los que ambos semáforos de peatones y vehículos están en rojo y por lo tanto ningún vehículo ni peatón debería cruzar.

Cuando se detecta que hay 17 o más peatones esperando para cruzar, el semáforo inteligente modifica la secuencia del ciclo del semáforo. La modificación de la secuencia ocurre si han transcurrido un mínimo de 39 segundos en fase verde de los vehículos. Esta condición se impone para garantizar siempre la fluidez del tráfico evitándose la formación de atascos.

De acuerdo con la condición impuesta para modificar la secuencia del semáforo se distinguen 2 posibles casos cuando se encuentran esperando 17 o más personas esperando para cruzar:

- **Caso 1:** la fase verde de vehículos ha estado durante menos de 39 segundos en el ciclo actual -> la transición a fase ámbar de vehículos tiene lugar una vez transcurran los 39 segundos en la fase verde de vehículos y 3 segundos después la fase roja se prolonga hasta el final del ciclo.
- **Caso 2:** la fase verde de vehículos se ha prolongado durante al menos 39 segundos en el ciclo presente -> la transición a fase ámbar para vehículos ocurre inmediatamente y 3 segundos después la fase roja se prolonga hasta el final del ciclo.

En la figura 3.5 se puede ver de manera esquemática como se distribuyen los tiempos en las distintas fases de vehículos y peatones tanto en el ciclo normal como en los 2 casos en los que la afluencia de público obliga a aumentar el tiempo de fase verde de peatones. Como se puede observar el tiempo adicional de fase verde de peatones se añade al final del ciclo. Las líneas azules representan el tiempo durante el cual se produce demanda (17 o más peatones esperando).



Figura 3.5. Esquema de distribución de las distintas fases para vehículos y peatones según el caso.

Este método de regular la circulación mediante el semáforo inteligente permite proporcionar un máximo de 48 segundos (se añaden como máximo 22 segundos) a los peatones para cruzar el paso.

El semáforo inteligente funciona de manera ininterrumpida las 24 horas al día y todos los días de la semana.



## **4. Metodologías estadísticas empleadas**

### **4.1. Datos funcionales**

Durante los últimos años han surgido numerosos instrumentos y sensores capaces de medir con precisión de manera casi continua una magnitud de un sistema. Habitualmente, dicha magnitud es definida respecto al tiempo, pero también puede ser definida respecto a la posición. Como consecuencia de esto se disponen de una gran cantidad de datos a lo largo de varias curvas o superficies. Estos datos se conocen como datos funcionales. De hecho, es más aconsejable que vectores de observaciones con muchas dimensiones se consideren como funciones ya que facilita su posterior análisis. Los datos funcionales se emplean en gran multitud de campos como medicina, meteorología, biometría, economía, geología, ingeniería, etc.

Como ejemplos de datos funcionales se pueden citar:

- Evolución temporal de la temperatura en una estación meteorológica, como en la figura 4.1.
- Medición del pulso cardíaco a lo largo del tiempo.
- Evolución temporal del peso de un bebé.
- Consumo de energía eléctrica en los hogares en función del tiempo.
- Velocidad de un corredor de maratón en función de la distancia recorrida.
- Evolución temporal de la cotización de un índice bursátil, como se puede observar en la figura 4.2.

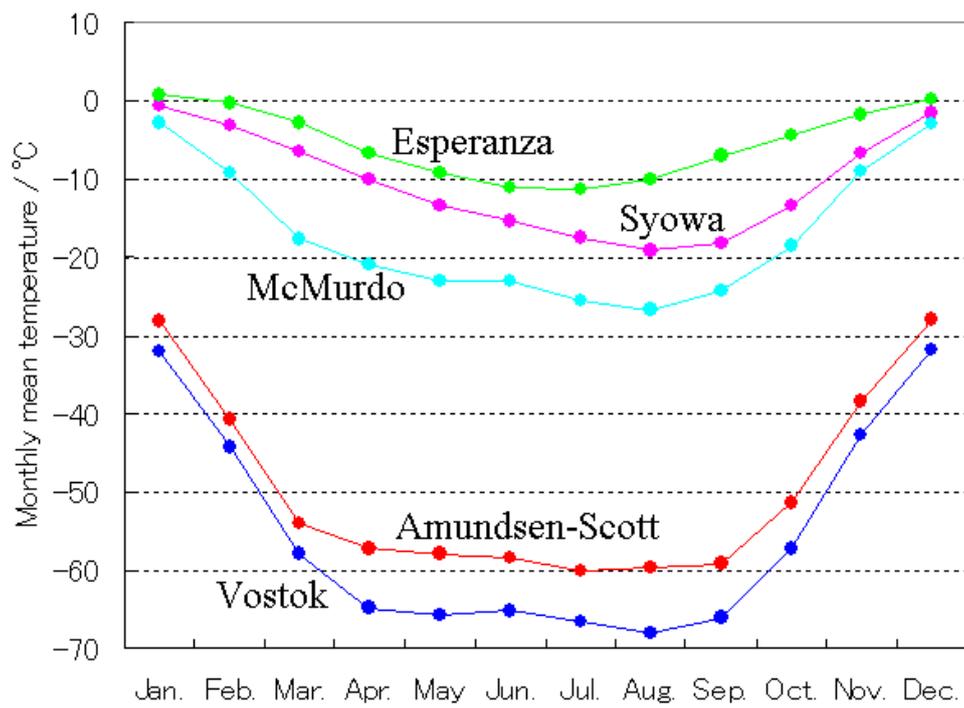


Figura 4.1. Evolución temporal de las temperaturas medias en estaciones meteorológicas de la Antártida. Fuente: [https://commons.wikimedia.org/wiki/File:Antarctic\\_monthly\\_mean\\_temperature.PNG](https://commons.wikimedia.org/wiki/File:Antarctic_monthly_mean_temperature.PNG)



*Figura 4.2. Evolución del índice bursátil Nasdaq entre 1994 y 2004. Fuente: [https://upload.wikimedia.org/wikipedia/commons/a/a6/NASDAQ\\_IXIC\\_-\\_dot-com\\_bubble.png](https://upload.wikimedia.org/wikipedia/commons/a/a6/NASDAQ_IXIC_-_dot-com_bubble.png)*

De la mano de los datos funcionales han ido surgiendo nuevos procedimientos estadísticos para analizar todos los datos surgidos a partir de estos sistemas conocido como Análisis de datos funcionales. En inglés se conoce con el nombre de Functional Data Analysis (FDA). El Análisis de datos funcionales considera cada curva u observación como una función. Con frecuencia se utilizan derivadas al analizar datos funcionales. En el análisis de datos funcionales se pueden encontrar adaptaciones para datos funcionales de procedimientos clásicos de la estadística como el ANOVA (Análisis de la varianza), el ACP (Análisis de Componentes Principales), el Análisis clúster, regresión lineal...

El primer paso antes de realizar esos análisis es convertir todo el conjunto de esos datos discretos en funciones que se puedan computar para cualquier valor de la variable independiente. Según [J. Ramsay y B.W. Silverman (2005)], la construcción de esas funciones se realiza en 2 etapas:

1. Se define un conjunto de bloques funcionales denominados funciones base.
2. Se crea un vector o matriz para definir la función como una combinación lineal de las bases de funciones.

Las funciones bases más empleadas son las series de Fourier y los splines. Las series de Fourier son empleadas en datos periódicos, que se repiten cada cierto periodo. Los splines son más complejos y suelen emplearse en datos no periódicos. A menudo estas dos bases necesitan ser complementadas con las bases constante y polinomio. Con esas 4 bases se puede tratar la mayor parte de los problemas en los que se aplica el análisis de datos funcionales

Una vez realizada la transformación de todos los datos en funciones, se pueden realizar algunas de las metodologías de la estadística clásica adaptadas al caso funcional. En los siguientes apartados se explicarán algunas de esas metodologías que se han aplicado en este estudio.

## 4.2. ANOVA funcional

### Introducción: caso univariante

El Análisis de la varianza para el caso univariante con un solo factor se basa en el siguiente modelo, según describe Douglas C. Montgomery (2004):

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, n$$

Siendo:

$y_{ij}$ : la observación  $ij$ -ésima

$\mu$ : la media general

$\tau_i$ : el efecto del tratamiento  $i$

$\varepsilon_{ij}$ : el error aleatorio que incluye las fuentes de variabilidad de factores no contemplados

$a$ : el número de niveles del factor

$n$ : el número de réplicas

Además, se deben cumplir las hipótesis de que los errores sean independientes y sigan una distribución normal con media cero y varianza  $\sigma^2$  constante.

En este caso se desea demostrar la igualdad de medias de los tratamientos:

$$\begin{cases} H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0 \\ H_1: \tau_i = 0 \text{ para al menos una } i \end{cases}$$

Se demuestra que se debe rechazar la hipótesis nula a un nivel de confianza  $1-\alpha$  si:

$$F_0 > F_{\alpha; a-1; N-a}$$

Donde  $F_0$  es:

$$F_0 = \frac{SS_{tratamientos}/(a-1)}{SS_{error}/(N-a)} = \frac{MS_{tratamientos}}{MS_E}$$

### Caso funcional

Sin embargo, el análisis de la varianza para datos funcionales con un solo factor se llevará a cabo de manera distinta. Se empleará el procedimiento descrito por [Antonio Cuevas, Manuel Febrero-Bande y Ricardo Fraiman(2004)], que utiliza un test inspirado en el F-test para la ANOVA univariante que hemos descrito previamente.

Su formulación matemática es la siguiente:

Sea  $X_{ij}(t)$  con  $j= 1, \dots, n_i$ ;  $t \in [a, b]$   $k$  muestras independientes de trayectorias tales que  $E(X_i(t)) = m_i(t)$

Las hipótesis que queremos comprobar son:

$$\begin{cases} H_0: m_1 = \dots = m_k \\ H_1: m_i \neq m_j \text{ para al menos un par } (i, j) \end{cases}$$

Se utiliza el estadístico  $V_n$  para medir la variabilidad entre grupos. La siguiente ecuación muestra cómo se calcula el estadístico  $V_n$ :

$$V_n = \sum_{i < j} n_i \| \bar{X}_i - \bar{X}_j \|^2$$

Se utiliza un procedimiento bootstrap que construye réplicas artificiales para estimar la función de distribución. Se puede demostrar que el estadístico  $V_n$  tiende a infinito cuando la hipótesis nula es falsa.

$$\text{si } p\text{-valor} \rightarrow 0 \Rightarrow V_n \rightarrow \infty$$

Además, este procedimiento tiene la ventaja de que no es necesario que se cumpla la hipótesis de homogeneidad de la varianza como ocurría con el análisis de la varianza en datos univariantes.

### 4.3. Análisis clúster funcional

El clustering o análisis clúster trata de agrupar las observaciones que sean muy similares de forma que observaciones de distintos grupos sean muy diferentes entre sí. Se pretende que las observaciones de un grupo sean homogéneas entre sí y heterogéneas con las de otros grupos.

Para la realización del análisis clúster de los datos se utilizará el algoritmo robusto de las k-medias recortadas [García-Escudero L.A, Gordaliza A (2005)].

Se parte de la metodología de las k-medias [McQueen 1967, Hartigan 1975]: Dada una muestra de  $p$  variables:  $X_1, \dots, X_n$ , se busca un conjunto de  $k$  puntos,  $m_1, \dots, m_k$ , que minimicen:

$$\min_{m_1, \dots, m_k} \frac{1}{n} \sum_{i=1}^n \inf_{1 \leq j \leq k} \|X_i - m_j\|^2$$

De esta manera, los clúster se generan asignando a cada observación al centro más cercano  $m_j$ . Hay algoritmos que implementan el método de las k-medias en un tiempo aceptable, incluso si el número de variables  $p$  es grande.

Un problema de esta metodología es su falta de robustez debido a que las observaciones outliers pueden modificar sustancialmente las observaciones asignadas a cada clúster. [Cuesta-Albertos, Gordaliza y Matrán (1997)] propusieron la robustificación de esta metodología mediante un procedimiento de recorte imparcial. En ella, dada un tamaño de recorte  $\alpha$  entre  $\gamma$  y 1, se busca un conjunto de  $k$  puntos,  $m_1, \dots, m_k$ , resolviendo el siguiente problema de doble minimización:

$$\widehat{W}_k(\alpha) := \min_{\mathbf{Y}} \min_{m_1, \dots, m_k} \frac{1}{[n(1-\alpha)]} \sum_{X_i \in \mathbf{Y}} \inf_{1 \leq j \leq k} \|X_i - m_j\|^2$$

Donde  $\mathbf{Y}$  varia en los subconjuntos de  $X_1, \dots, X_n$  que contienen  $[n(1-\alpha)]$  datos. Después cada observación no recortada es colocada en el clúster correspondiente al centro más cercano k-media recortada  $m_j$ . Hay que indicar que las 2 últimas expresiones son análogas, pero esta última tiene la ventaja de que una proporción  $\alpha$  de observaciones pueden ser recortadas.

Para llevar a cabo la metodología de las k-medias recortadas con datos funcionales es necesario transformar las curvas en vectores, es decir, pasar de infinitas dimensiones a finitas dimensiones. Para ello, [García-Escudero L.A y Gordaliza A. (2005)] proponen proyectar las curvas en una base funcional

adecuada, ya que las curvas se suavizan y se elimina el ruido molesto. Se considera una función de base de  $p$  dimensiones. Dada una única curva de datos  $\{(t_i^j, x_i(t_i^j))\}_{j=1}^{J_i}$  (resultado del registro de la curva  $x_i$  en los  $J_i$  instantes  $t_i^1 < t_i^2 < \dots < t_i^{J_i}$ ), se puede modelizar dicha curva mediante un ajuste lineal por mínimos cuadrados:

$$x_i(t_i^j) = \sum_{s=1}^p \beta_i^s B_s(t_i^j) + \epsilon_i, \quad j = 1, \dots, J_i$$

Entonces para  $n$  curvas, se obtendrán  $n$  vectores de coeficientes ajustados:

$$\beta_i = (\beta_i^1, \dots, \beta_i^p)', \quad i = 1, 2, \dots, n.$$

En nuestro caso se emplearán como base splines cúbicos. En un spline, hay que tener en cuenta que si tenemos  $S$  nodos interiores, el número de funciones de base será de  $p = S + 4$ :

$$\{\beta_1, \dots, \beta_p\}$$

#### 4.4. Análisis de Componentes Principales (ACP) funcional

El Análisis de Componentes Principales (ACP) se utiliza para conocer los principales modos de variación en los datos. Se empleará la explicación de [Ramsay J, Hooker G, Graves S (2009)].

En el primer paso del ACP funcional, se elige la función peso sondeo  $\xi$  para que las puntuaciones sondeo tengan la máxima variación:

$$\rho_\xi(x_i) = \int \xi(t)x_i(t)dt$$

Siendo:

$\rho_\xi(x_i)$  las puntuaciones sondeo

$\xi(t)$  la función peso sondeo

$x_i(t)$  la curva funcional  $i$

Para ello es necesario imponer la restricción de tamaño:

$$\int \xi^2(t)dt = 1$$

La curva media por definición es un modo de variación compartido por muchas curvas, por ello normalmente se elimina primero la media y después examina los residuales funcionales  $x_i - \bar{x}$ . La varianza sondeo de la puntuación  $Var[\int \xi(t)(x_i(t) - \bar{x}(t))^2 dt]$  asociada con el peso sondeo  $\xi$  es el valor de:

$$\mu = \max_{\xi} \left\{ \sum_i \rho_{\xi}^2(x_i) \right\} // \text{ sujeto a } \int \xi^2(t) dt = 1$$

En la terminología estándar,  $\mu$  y  $\xi$  se consideran los mayores autovalores y autofunciones, respectivamente, de la función estimada varianza-covarianza  $v$ :

$$v(s, t) = (N - 1)^{-1} \sum_i [x_i(s) - \bar{x}(s)][x_i(t) - \bar{x}(t)]$$

Siendo  $N$  el número de curvas.

Como en el caso del ACP multivariante, se puede construir paso a paso una secuencia no creciente de autovalores  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_k$  con la condición de que cada nueva autofunción computada en el paso  $l$  sea ortogonal a las computadas en paso previos:

$$\int \xi_j(t) \xi_l(t) dt = 0, j = 1, \dots, l - 1$$

$$\int \xi_l^2(t) dt = 1$$

En el análisis multivariante los pares de autovalores y autovectores se computan con el autoanálisis de la matriz covarianza  $\mathbf{V}$  resolviendo  $\mathbf{V}\xi_j = \mu_k \xi_j$ . Para datos funcionales es similar: se calculan autofunciones  $\xi_j$  de la función de covarianza bivalente  $v(s,t)$  como soluciones de:

$$\int v(s, t) \xi_j(t) dt = \mu_j \xi_j(s) [x_i(t) - \bar{x}(t)] dt$$

Los vectores de coeficientes  $\mathbf{c}_i$ ,  $i = 1, \dots, N$  contienen los coeficientes  $c_{ij}$  que definen el ajuste óptimo para cada función  $x_i$ , y son conocidos como puntuaciones de la componente principal:

$$c_{ij} = \rho_{\xi_j}(x_i - \bar{x}) = \int \xi_j(t) dt$$

Las puntuaciones de la componente principal son muy útiles para conocer la naturaleza de la variación identificada por el análisis de componentes principales.

#### 4.5. Test Chi-cuadrado de homogeneidad

El test Chi-cuadrado se utiliza para determinar si las frecuencias de las clases de varias poblaciones se encuentran igualmente distribuidas en todas las poblaciones.

Matemáticamente el problema se expresa de la siguiente manera:

Se consideran  $m$  poblaciones independientes:  $r_1, r_2, \dots, r_m$ ; divididas en  $n$  clases:  $C_1, C_2, C_3, \dots, C_n$ .

Si se llama  $p_{ij}$  a la probabilidad de la clase  $i$  en la población  $j$ , las hipótesis que se quieren comprobar son las siguientes:

$$\left\{ \begin{array}{l} p_{11} = \dots = p_{1m} \\ H_0: \dots \\ p_{n1} = \dots = p_{nm} \\ H_1: \text{Al menos una de las igualdades anteriores es falsa} \end{array} \right.$$

Para comprobar si se cumplen las hipótesis se utiliza el estadístico  $\chi^2$  que se calcula de la siguiente manera:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

Siendo:

$O_{i,j}$  el número de observaciones de la población  $i$  en la clase  $j$

$E_{i,j}$  el valor esperado para la clase  $i$  en la población  $j$ . Se calcula según:

$$E_{i,j} = \frac{n_j * S_i}{N}$$

Siendo:

$n_j$  el número de observaciones de la clase  $j$  en todas las poblaciones. Se calcula:

$$n_j = \sum_{i=1}^m O_{i,j}$$

$S_j$  el número de observaciones de la población  $j$ . Se calcula de la siguiente manera:

$$s_i = \sum_{j=1}^n O_{i,j}$$

N es el número de observaciones en todas las poblaciones o el número de observaciones en todas las clases juntas. Se calcula así:

$$N = \sum_{j=1}^n n_j = \sum_{i=1}^m s_i$$

En la tabla 4.1, se puede ver la tabla de relación entre las poblaciones y la clase utilizando toda la nomenclatura anterior.

<b>Población/clase</b>	<b>c<sub>1</sub></b>	<b>c<sub>2</sub></b>	<b>..</b>	<b>c<sub>n</sub></b>	
<b>r<sub>1</sub></b>	<i>O<sub>1,1</sub></i>	<i>O<sub>1,2</sub></i>		<i>O<sub>1,n</sub></i>	<i>S<sub>1</sub></i>
<b>r<sub>2</sub></b>	<i>O<sub>2,1</sub></i>	<i>O<sub>2,2</sub></i>		<i>O<sub>2,n</sub></i>	<i>S<sub>2</sub></i>
<b>..</b>					
<b>r<sub>m</sub></b>	<i>O<sub>m,1</sub></i>	<i>O<sub>m,2</sub></i>		<i>O<sub>m,n</sub></i>	<i>S<sub>m</sub></i>
	n <sub>1</sub>	n <sub>2</sub>		n <sub>3</sub>	N

*Tabla 4.1. Tabla de relación entre las poblaciones y las clases*

El número de grados de libertad en un test chi-cuadrado de homogeneidad con n clases y m poblaciones es de (n-1)(m-1).

Con el valor del estadístico chi-cuadrado y el número de grados de libertad se obtiene el p-valor a partir de la función de distribución que determinará si se debe rechazar la hipótesis nula o aceptarla.

Esta distribución chi-cuadrado es la distribución de la suma de cuadrados de variables aleatorias normales. Sean  $Z_i$  variables independientes aleatorias normales (0,1), entonces:

$$\chi^2 = \sum_{i=1}^n Z_i^2$$

Entonces  $\chi^2$  es la distribución chi-cuadrado para n grados de libertad. La forma de la distribución depende del número de grados de libertad como se puede observar en la figura 4.1.

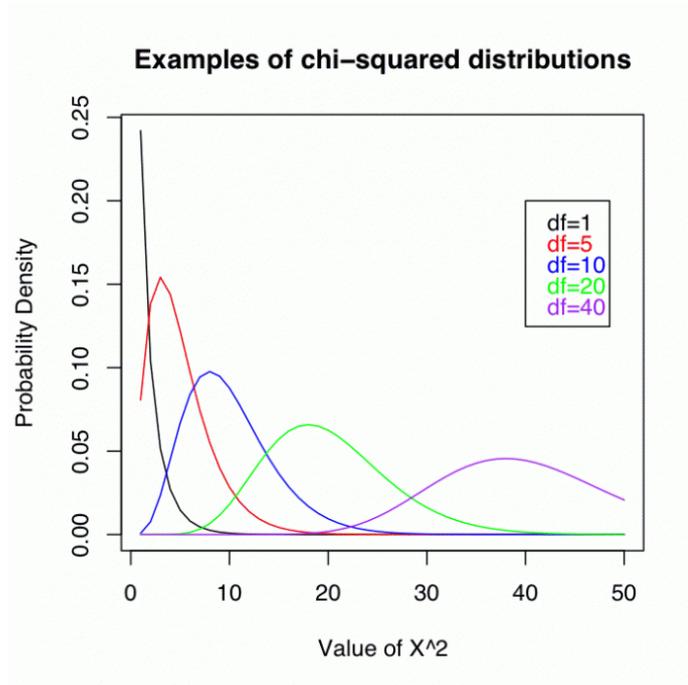


Figura 4.3. Forma de la distribución chi-cuadrado según el número de grados de libertad



## 5. Depuración de los datos y análisis preliminar

### 5.1. Depuración de datos

#### 5.1.1. Introducción a la depuración de datos

Los datos proporcionados por el sistema que realiza el conteo de peatones y de los tiempos de espera no siempre son correctos, y deben ser comprobados antes de comenzar a trabajar con ellos. De esta manera conseguiremos realizar un buen análisis estadístico. Además, también estos sistemas pueden proporcionar los datos de manera compleja, y por tanto en esta etapa también se deberá realizar las modificaciones necesarias para que el análisis posterior con el software estadístico R sea más sencillo de realizar.

Por último, en la depuración se desecharán los datos proporcionados por el sistema que no sean necesarios ya que esto agilizará el tiempo de ejecución con los programas estadísticos y se evitarán posibles confusiones en la interpretación de los datos.

#### 5.1.1.1 Glosario de términos

A continuación se adjuntan una serie de términos que se emplearán bastante a lo largo del TFG y que es necesario conocer el sentido que se quiere indicar y así evitar malas interpretaciones:

- **Conteos de peatones:**
  - **Conteo total:** se trata del número de personas total que pasan por el área de las cámaras del semáforo. Incluye a las personas que pasan en paralelo por la calle Claudio Moyano, las que cruzan, las que esperan y toda otra trayectoria que pase por delante de la cámara.
  - **Espera total:** se trata del número de peatones que se detienen completamente en el semáforo. En este número se encuentra tanto los que se esperan a que el semáforo se ponga verde para cruzar y los que se detienen por cualquier otra causa.
  
- **Tiempos de espera:** es el tiempo que cada peatón espera el semáforo.

### 5.1.2. Depuración de conteos de peatones

Los datos de cada una de las 4 cámaras se proporcionaban en un formato en 7 columnas de la siguiente manera:

- **Fecha\_HoraFinDelBloque:** la fecha en formato americano unido por un guión bajo de la hora de la medición (Año-Mes-Día\_HHMMSS).
- **ConteoHaciaElPaso:** Es el número de personas que se dirigen al paso con intención de cruzarlo.
- **ConteoSentidoContrario:** Número de personas que cruzan el paso hacia la cámara.
- **Conteolzqda:** Personas que pasan por el área de la cámara pero no cruzan el paso desplazándose lateralmente hacia la izquierda de la cámara por la calle Claudio Moyano.
- **ConteoDrcha:** Igual que en el Conteolzqda, pero yendo hacia la derecha de la cámara.
- **Total.** Número de personas detectadas. En esta categoría se encuentra la suma de todos los conteos anteriores y de todas las trayectorias no contempladas: peatones que cruzan la calle en diagonal o que se encuentran parados.
- **Espera.** Número de personas que esperan a cruzar el paso.

En la tabla 5.1 se muestra la apariencia de la toma de datos tal y como el sistema los recoge.

Fecha_HoraFinDelBloque	ConteoHaciaElPaso	ConteoSentidoContrario	Conteolzqda	ConteoDrcha	Total	Espera
2015-01-07_083000	12	2	14	3	31	1
2015-01-07_084500	9	2	3	5	19	1
2015-01-07_090000	12	6	13	5	37	1
2015-01-07_090934	11	8	9	0	30	2
2015-01-07_091500	2	0	6	7	15	1

*Tabla 5.1. Extracto de la apariencia de la toma de datos del sistema (sin depurar).*

Para facilitar el análisis en función de la hora, se separa la fecha de la hora. Además, se cambia el formato de la fecha al estándar en España (Día-Mes-Año). En la tabla 5.2 se puede ver un ejemplo del extracto de los datos con este nuevo formato.

07/01/2015	111500	97	54	11	23	195	17
07/01/2015	113000	113	54	14	16	210	30
07/01/2015	114500	130	85	21	15	254	29
07/01/2015	120000	142	116	25	18	335	48
07/01/2015	121500	165	104	21	15	326	47

*Tabla 5.2. Extracto de la apariencia de los datos tras separar la fecha de la hora y estandarizar la fecha.*

En los datos del flujo de peatones del cruce de la calle Santiago con Claudio Moyano junto a la Plaza de Zorrilla (Valladolid), se encontraron una serie de datos que no tenían mucho sentido:

- Registro de datos con fechas del año 2017. En la tabla 5.3, se ve un ejemplo de este caso con 2 datos con fecha del 28 de septiembre de 2017.
- Datos registrados cada pocos segundos.
- Falta de datos durante un intervalo en el que las cámaras funcionaban. Esto se puede observar también en la tabla 5.3 ya que se pasa del 28 de octubre al 10 de diciembre (ignorando el error con fecha del año 2017)

28/10/2014	140000	0	0	0	0	0	0
28/10/2014	141500	0	0	0	0	0	0
28/09/2017	164500	22	16	10	11	66	4
28/09/2017	170000	47	42	24	8	139	11
10/12/2014	133000	75	59	12	9	166	4
10/12/2014	134500	47	25	9	10	95	5
10/12/2014	135824	27	23	6	11	71	2

*Tabla 5.3 Ejemplo de datos con errores*

Estas incidencias se reportaron a la entidad que proporciona los datos (CIDAUT) y se llegó a la conclusión de que se debían a apagones o cortes de luz y resets del sistema que regula el semáforo en su mayor parte. Tras estas conclusiones se procede a su eliminación inmediata del registro de datos para poder realizar un correcto análisis estadístico.

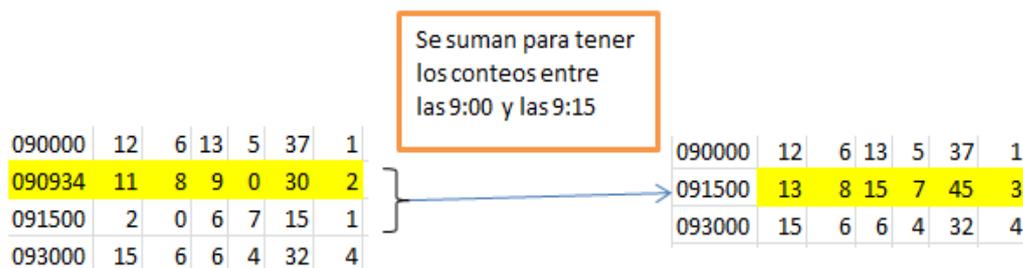
## Realización de los conteos

Los conteos siempre se producen por defecto cada 15 minutos, coincidiendo con las horas en punto, y cuarto, y media, y menos cuarto. Sin embargo, en algunas ocasiones hay conteos intermedios. Estos se producen en un intervalo de tiempo inferior, aunque las mediciones a en punto, a y cuarto, a y media y a menos cuarto siempre se mantendrán a pesar de esas tomas de datos esporádicas.

Dichos conteos intermedios provocan un problema. Los intervalos a comparar dejan de ser semejantes: no son siempre comparables dos datos a la misma hora de diferentes días.

Por ejemplo, un recuento a las 9:15:00 en un día en el que no hay conteos intermedios entre las 9 y las 9:15 corresponde a la medición del número de personas que ha pasado durante esos 15 minutos. Sin embargo, si hay un conteo intermedio a las 9:05:00, el recuento a las 9:15:00 corresponde a la gente pasa entre esos 10 minutos.

Por tanto, para solucionar este problema se asignan todos los conteos al siguiente cuarto de hora, salvo que el dato sea un cuarto de hora, y se suman todos los conteos. De esta manera, los recuentos de personas que se disponen serán los comprendidos durante los 15 minutos previos que nos indicará la hora. En la figura 5.1 se puede ver de una manera más intuitiva lo explicado anteriormente.



*Figura 5.1. Gráfico explicativo del procedimiento para tener los conteos cada 15 minutos*

A continuación, se opta por quedarnos con los conteos totales y de espera, suprimiendo el resto de conteos que consideramos no serán de interés en nuestro estudio.

Finalmente, se generan una serie de datos como mes, día de la semana, festivo, día del mes, etc, a partir de los datos de la fecha para facilitar el análisis del flujo de personas en función del tiempo en distintos períodos.

Los archivos de datos generados para cada cámara tienen una disposición de 176 filas y 101 columnas. Las 176 filas se deben a que se dispone de mediciones en 176 días. Las 5 primeras columnas representan la Fecha (en el formato Día-Mes-Año), día de la semana (1=Lunes,... 7=Domingo), mes (1=Enero,... 12=Diciembre), semana (número de la semana del año) y Festivo (adquiere los Valores SI/NO dependiendo de si el día es festivo). Las 96 columnas restantes representan los 96 cuartos de hora de un día en los que se van tomando las mediciones de peatones. Más tarde se decidió incluir 3 columnas indicando el tipo de día (Laborable, Laborable no escolar, Sábado y Domingo-festivo), ubicación del día en el mes (Principio, Medios, Finales) y precipitación en el día (SI/NO).

Posteriormente, como en todas las cámaras se miden los mismos flujos desde diferentes ángulos se decide coger para realizar el análisis estadístico el máximo de cada observación de las 4 cámaras, tanto para el conteo total como la espera total. De esta manera se evitan los problemas que surgirían al realizar el análisis estadístico de aquellas observaciones en las que no se registraba correctamente el paso de los peatones en alguna cámara.

En la tabla 5.4 se puede apreciar un fragmento del archivo de datos resultante para el conteo total de peatones. El formato usado para la espera total es el mismo.

Fecha	Día.Semana	Mes	Semana	Festivo	X0	X0.25	X0.5	X0.75	X1	X1.25	X1.5	
12/09/2014		5	9	37	NO	441	384	227	362	664	297	172
13/09/2014		6	9	37	NO	447	308	298	290	185	185	260
14/09/2014		7	9	37	SI	520	510	532	427	342	274	253
15/09/2014		1	9	38	NO	50	52	25	15	19	5	11
16/09/2014		2	9	38	NO	11	11	11	11	5	8	4
17/09/2014		3	9	38	NO	27	27	6	10	6	4	2
18/09/2014		4	9	38	NO	14	17	10	17	11	8	2
19/09/2014		5	9	38	NO	21	22	17	22	19	10	4
20/09/2014		6	9	38	NO	42	27	29	45	33	25	23
21/09/2014		7	9	38	SI	140	125	96	91	61	68	78
22/09/2014		1	9	39	NO	37	79	92	100	99	57	71
23/09/2014		2	9	39	NO	33	17	12	8	6	12	7
24/09/2014		3	9	39	NO	24	14	11	14	3	11	12

Tabla 5.4. Fragmento del archivo de datos definitivo para el conteo total de peatones

Se aprecia, que a pesar de coger el máximo de las 4 cámaras, hay 5 días en los que hay franjas horarias sin ningún dato debido a que en alguna medición del día no funcionaba ninguna de las 4 cámaras:

- El primer día (11/9) y el último día (6/3) porque las mediciones no empezaban o terminaban a las 0:00.
- El 28/10, el 27/11 y el 10/12 por fallo de las cámaras probablemente por apagón.

Esos días son eliminados y finalmente el análisis estadístico se realizará con las mediciones de 171 días entre el 12 de septiembre de 2014 y 5 de marzo (ambos incluidos).

### 5.1.3. Depuración de los tiempos de espera

En estos datos no se encontraron ningún tipo de fallo en el sistema. El registro de los tiempos de espera de cada peatón se realiza en 3 columnas:

- En la primera la fecha en formato americano.
- En la segunda la hora en formato HH:MM:SS.
- En la tercera el tiempo de espera en segundos (número entero redondeado sin decimales).

Un ejemplo del registro de los tiempos de espera se puede ver en la tabla 5.5.

2014-04-26	15:30:47	1
2014-04-26	15:30:47	1
2014-04-26	15:28:00	1
2014-04-26	15:28:01	2
2014-04-26	15:32:20	0
2014-04-26	15:32:36	0
2014-04-26	15:42:42	1

*Tabla 5.5. Ejemplo del registro de los tiempos de espera*

En primer lugar, se verificó que los tiempos de espera tuvieran sentido. Se apreció que no había, como cabía esperar cuando el semáforo está en verde en una hora de gran tránsito en la calle Santiago, muchas mediciones seguidas de tiempo de espera de 0 segundos. Posteriormente, se confirmó que el sistema mide los tiempos de espera de cada peatón solamente si se detiene completamente durante un muy pequeño tiempo umbral. Es decir, todos los tiempos de espera de 0 segundos se debían a que el peatón se había detenido un

pequeño instante, y que la gente que no se detiene no se encuentra en este registro.

Como se comentó en el capítulo 3, el ciclo del semáforo para los peatones es de 20 segundos en verde, 6 en verde intermitente y 74 en rojo (el ciclo es de 100 segundos), en caso de que haya 17 peatones o más esperando el sistema aumenta el tiempo en verde. Sin embargo, se detectan tiempos de espera muy superiores a esos 74 segundos de la fase roja. Esos datos se deben a que simplemente el peatón se detiene sin intención de cruzar o por una falsa detección, así que se eliminan para poder realizar una correcta interpretación. Se eliminan todos aquellos datos con espera superior a 80 segundos, ya que se considera que un peatón puede llegar al paso y al observar que se encuentra en verde intermitente espere al siguiente ciclo esperando los 6 segundos de la fase verde intermitente y los 74 de rojo.

El número de datos eliminados es razonable: 277 de 23706, un 1,16%. No parece que haya alguna razón como una manifestación o marcha que distorsione los tiempos de espera. Los datos eliminados se deberán probablemente a personas que se detienen a hablar con otras personas que se encuentran en el cruce.

De todos los datos disponibles, 2 días corresponden al período en el que el semáforo inteligente no estaba activado y por tanto, el ciclo era fijo, y el resto son días en los que el semáforo inteligente funcionaba correctamente.

Los 2 días con el semáforo inteligente desactivado son el lunes 5 de mayo de 2014 y el martes 6 de mayo de 2014.

A continuación, para dejar preparado los datos para realizar el análisis de la efectividad del semáforo inteligente, se decide comparar individualmente los 2 días con el sistema activado con los 2 mismos días de la semana anterior, es decir, se compara el lunes 5 de mayo de 2014 con el lunes 28 de abril de 2014 y el martes 6 de mayo de 2014 con el martes 6 de mayo de 2014. Esa comparación se realiza en franjas horarias debido a la gran cantidad de datos disponibles.

De esta manera, se crean 6 ficheros de datos para cada día con las franjas horarias 0-14 (mañana), 14-19.5 (tarde) y 19.5-24 (noche). Se eligen estas franjas para que salgan aproximadamente el mismo número de observaciones de tiempo de espera (aproximadamente unas 300) en cada franja horaria. Los ficheros de datos contienen una única columna con el tiempo de espera de cada peatón. Un ejemplo de un extracto de un fichero de datos se puede observar en la tabla 5.6.

1  
15  
2  
0  
1  
10  
1  
0  
1  
7  
10  
61  
43  
22

*Tabla 5.6. Extracto de un fichero de datos de los tiempos de espera*

## 5.2. Análisis descriptivo de los datos

### 5.2.1 Análisis descriptivo de los conteos de peatones

Tras depurar los datos, se realiza un análisis descriptivo previo en el que se apreciará el comportamiento del flujo de peatones en función de la hora mediante una serie de gráficas.

#### **Media del conteo total de peatones a lo largo de los días de la semana**

En la figura 5.2, se puede ver la representación del número de personas que pasa de media cada 15 minutos en función del día de la semana (1= Lunes, 2= Martes,..., 7=Domingo)

Se observa que hay 2 picos de mayor afluencia, a las 12:30 y a las 20:00 en casi todos los días. El sábado y el domingo estos picos están ligeramente desplazados a la derecha produciéndose más tarde. Además, el pico del sábado a las 20:30 es considerablemente mayor al resto, siendo el momento en que pasan mayor número de personas de toda la semana. El flujo de peatones a primeras horas de la madrugada solo es apreciable el viernes, el sábado y el domingo, siendo el resto de días prácticamente nulos. Sin embargo, de 7 a 12:30 pasan menos personas de media los sábados y domingos que el resto de la semana.

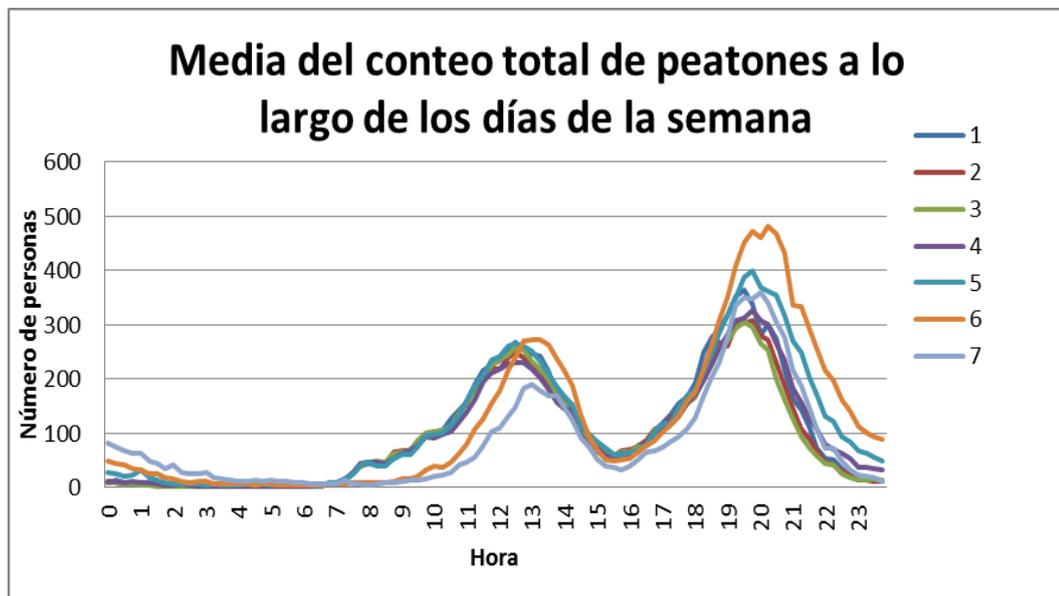


Figura 5.2. Media del conteo total de peatones a lo largo de los días de la semana

### **Media del número de peatones que esperan en función del día de la semana**

En la figura 5.3 se observa la representación del número de peatones que esperan para cruzar en función del día de la semana. Como es esperable, en los días que se tenía mayor flujo total de peatones son los que tienen un mayor número de personas que esperan para cruzar. Además, se observa que la relación entre espera y conteo total no es lineal, ya que se aprecia que la espera aumenta mucho más rápidamente cuando el conteo total es mayor.

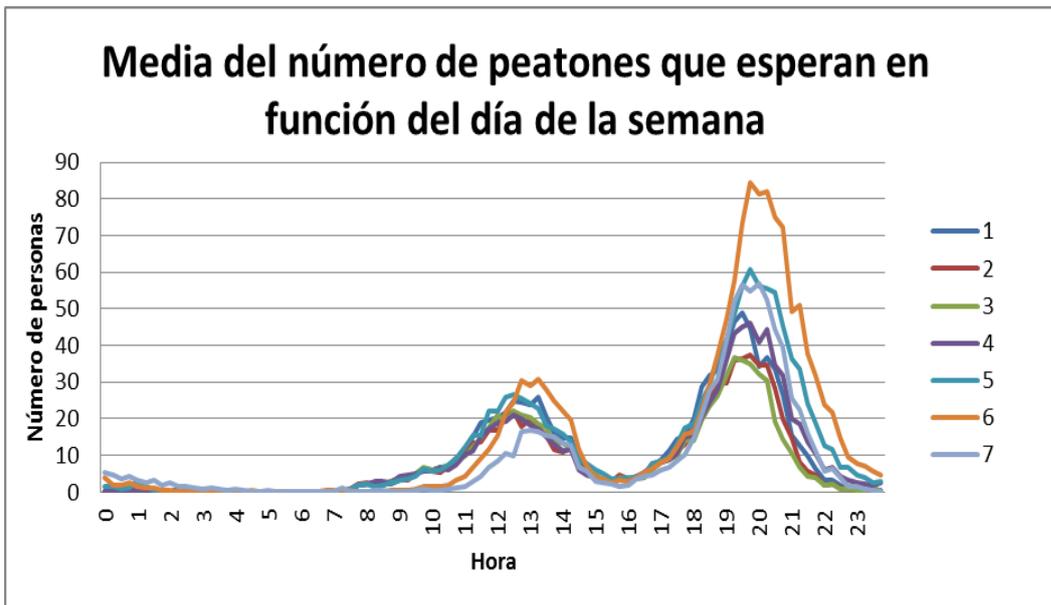


Figura 5.3. Media del número de peatones que esperan en función del día de la semana.

### **Media del conteo total de peatones en función de si el día es festivo o no**

En la figura 5.4 tenemos la representación de la variación del flujo de personas medio en función de si el día es festivo o no festivo. Las conclusiones son muy similares a las del gráfico anterior: hay más actividad a primeras horas de la madrugada, por la mañana pasa menos gente y el pico matinal se produce más tarde en los días festivos.



Figura 5.4. Media del conteo total de peatones en función de si el día es festivo.

#### **Media de la espera total en función de si el día es festivo o no**

La figura 5.5 muestra la media de las personas que esperan en función de si el día es festivo o no. Se observa claramente que el pico matinal es más reducido y más tardío los días festivos. El pico vespertino es, sin embargo, superior los días festivos y se produce a la misma hora que los no festivos.



Figura 5.5. Media de la espera total en función de si el día es festivo o no

### Media del conteo total de peatones en función del mes

La figura 5.6 muestra la media del flujo de personas en función del mes (1=Enero, 2=Febrero... 12=Diciembre). Recordemos que los datos están recogidos entre el 11/9/14 y el 5/3/15, por lo que se tienen 7 meses y septiembre y marzo se encuentran incompletos. Todos los meses tienen un comportamiento muy similar, solo sobresalen 2 meses. En diciembre hay un mayor flujo en el pico de la mañana y especialmente en el de la tarde, probablemente se deba a la actividad comercial por las compras de Navidad. Septiembre destaca especialmente por el mayor flujo por la noche (21-2 horas), probablemente motivado por la actividad nocturna de los días de fiestas de Valladolid.

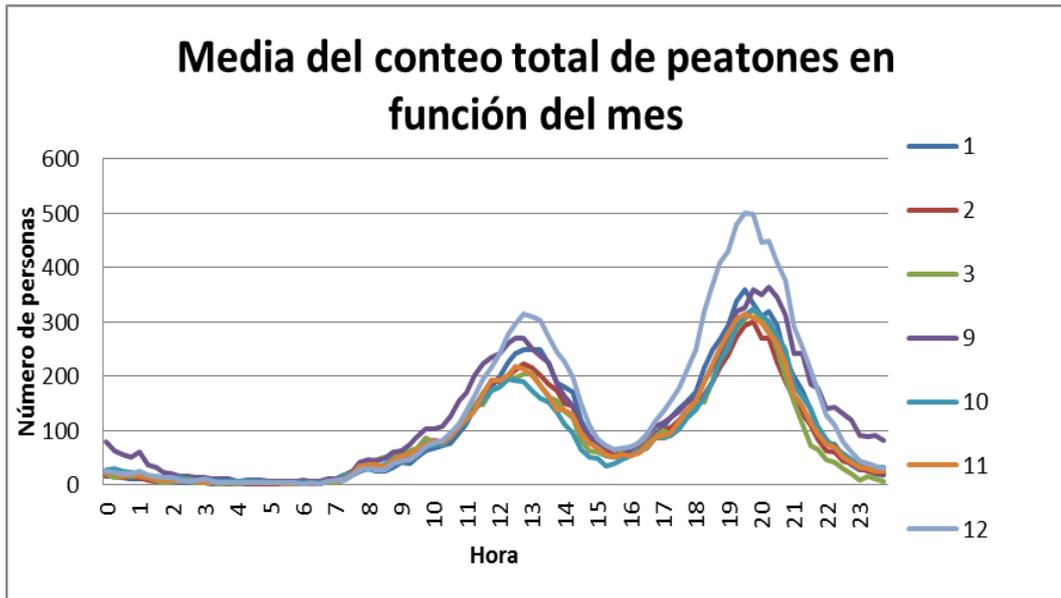


Figura 5.6. Media del conteo total de peatones en función del mes

**Media de la espera total en función del mes**

En la figura 5.7 se observa la media de la espera total en los meses observados. Destaca la gran espera total en el mes diciembre tanto por la mañana como por la tarde como ocurría en el conteos totales, llegándose a casi doblar respecto al resto de meses en el pico que se produce por la tarde. En menor medida se observa algo más de espera por la tarde y por la noche en septiembre.

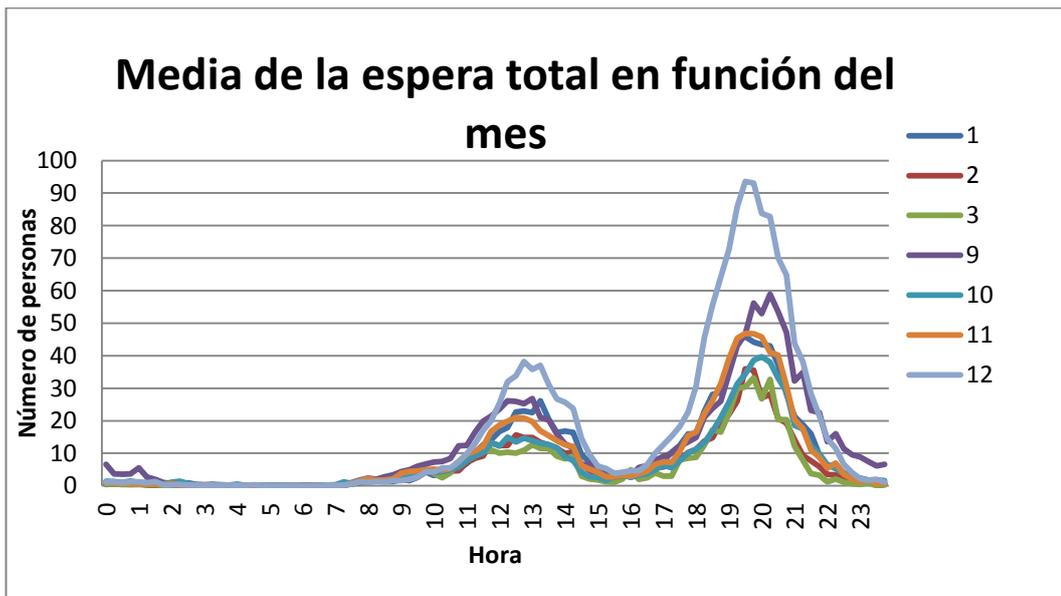


Figura 5.7. Media de la espera total en función del mes

### Media del conteo total de peatones en los fines de semana de los diferentes meses

La figura 5.8 nos ayuda a comprender el comportamiento durante los fines de semana (sábados y domingos) durante los diferentes meses. Se observa una gráfica muy parecida a la de la figura 5.6 en la que se observaban todos los días de los meses, acentuándose más el pico de la tarde en diciembre, respecto a la gráfica de todos los días del mes.

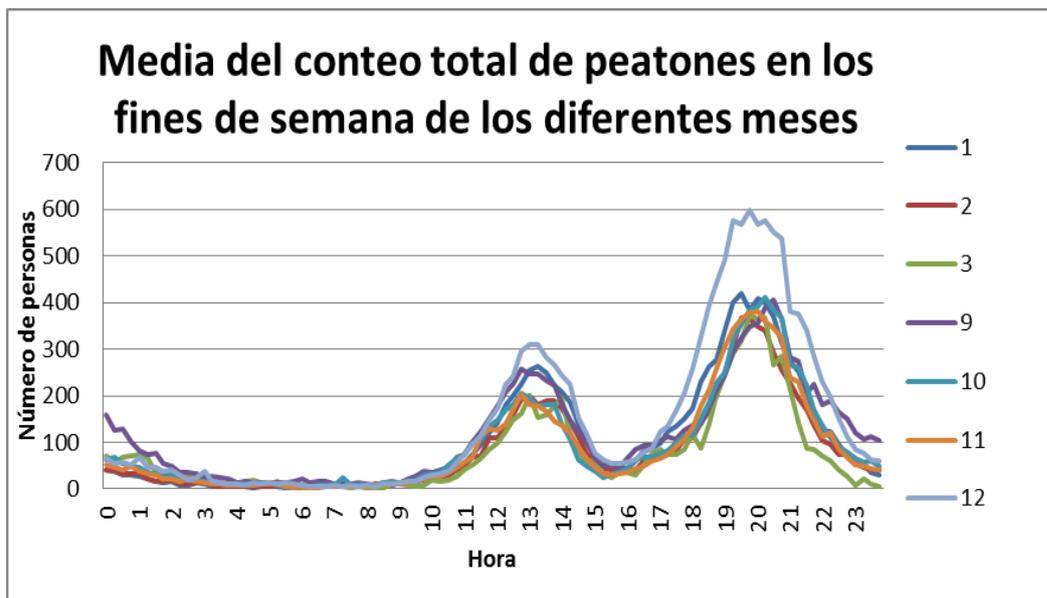


Figura 5.8. Media del conteo total de peatones en los fines de semana de los diferentes meses.

### Media del conteo total de peatones en función del día del mes

También es interesante conocer el comportamiento del flujo de peatones en función del día del mes. En la figura 5.9 se observa la media del conteo de personas en los primeros y en los últimos días del mes. No se aprecian grandes diferencias entre días, por lo que es previsible que un día sea de principios o finales de un mes no influya en el número de personas que pasan por la calle Santiago. Lo más destacable, es que el flujo en los días 1 en el pico de las 12.30 es ligeramente inferior al de resto de días, aunque esto puede que sea motivado porque muchos de esos días 1 eran festivos (4 de los 7 datos).

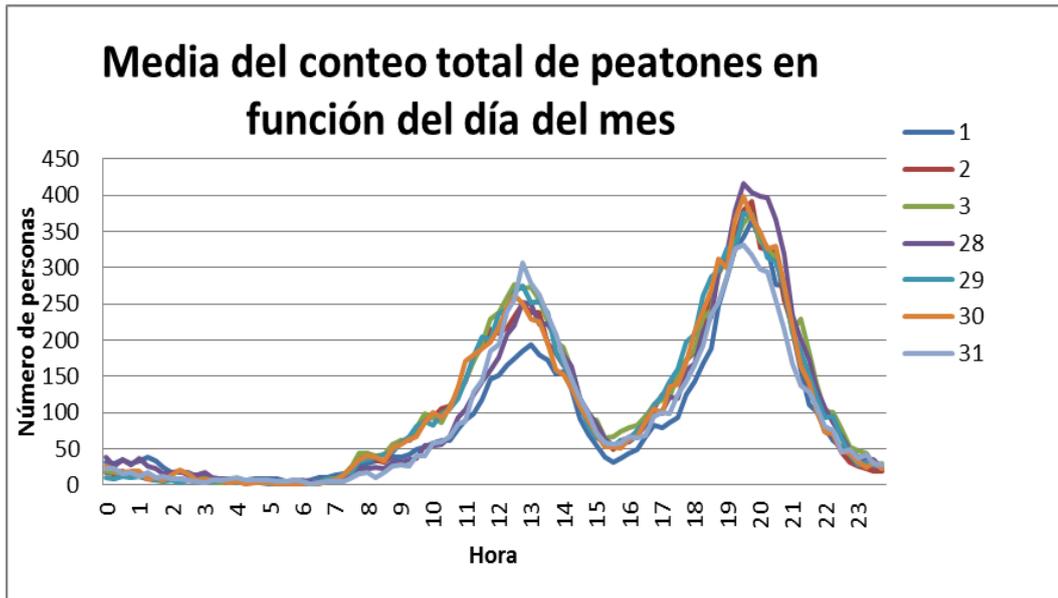


Figura 5.9. Media del conteo total de peatones en función del día del mes.

### **Conteo total de peatones de varios días de la época navideña**

La época de Navidad también puede aportar datos relevantes al estudio. En la figura 5.10 se han graficado los días más representativos del periodo navideño: Nochebuena, Navidad, Nochevieja, Año Nuevo y Reyes, y algunos de los días previos en los que es previsible que se realicen muchas compras. También, se ha añadido el día de 7 enero por el comienzo de las rebajas y por ser un día en el que se realizan muchas devoluciones en los comercios.

Se aprecia claramente que el día 5 de enero por la tarde tiene un flujo mucho mayor que el resto de días, previsiblemente se debe a la cabalgata de los Reyes Magos y a las compras de última hora.

Es notable la escasa actividad el día de Reyes, siendo sus picos de la mañana y de la tarde los más bajos comparados con el resto de días analizados. También, destaca el flujo entre la 1 y las 3 de la madrugada del día de Año Nuevo respecto al nulo flujo de peatones el resto de días, supuestamente debido a las fiestas nocturnas que se celebran en Año Nuevo.

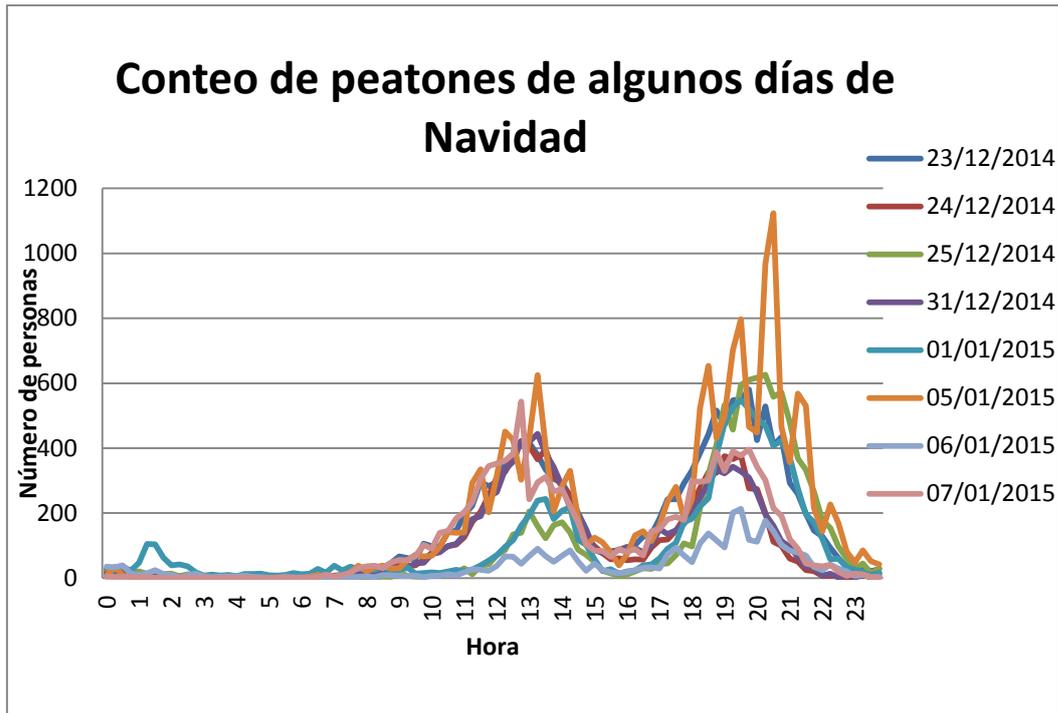


Figura 5.10. Conteo de peatones de algunos días de la época navideña

**Comparación entre la media del Conteo Total de peatones y la media de peatones que esperan**

En la figura 5.11, se compara el número de personas que pasan por la calle Santiago respecto a los que esperan para cruzar el paso. Se observa que la espera es considerable solo en las horas donde hay un pico de personas totales pasando por la calle (en torno a las 12:30 y 19:30)

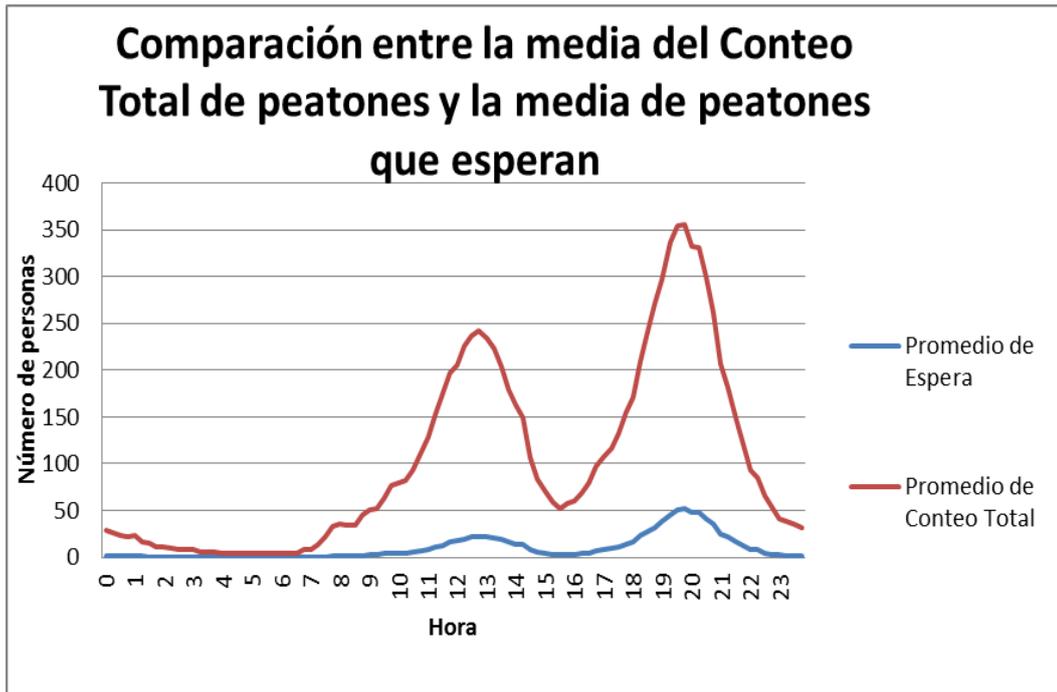


Figura 5.11. Comparación entre la media del conteo total y la espera total

**Conteo total de peatones en función de la precipitación**

En la figura 5.12 se puede ver el conteo total de peatones en función de la precipitación. Se observa que los días sin lluvia hay mayor actividad peatonal que los días de lluvias como cabe esperar tanto en el pico matinal como en el que se produce por la tarde.

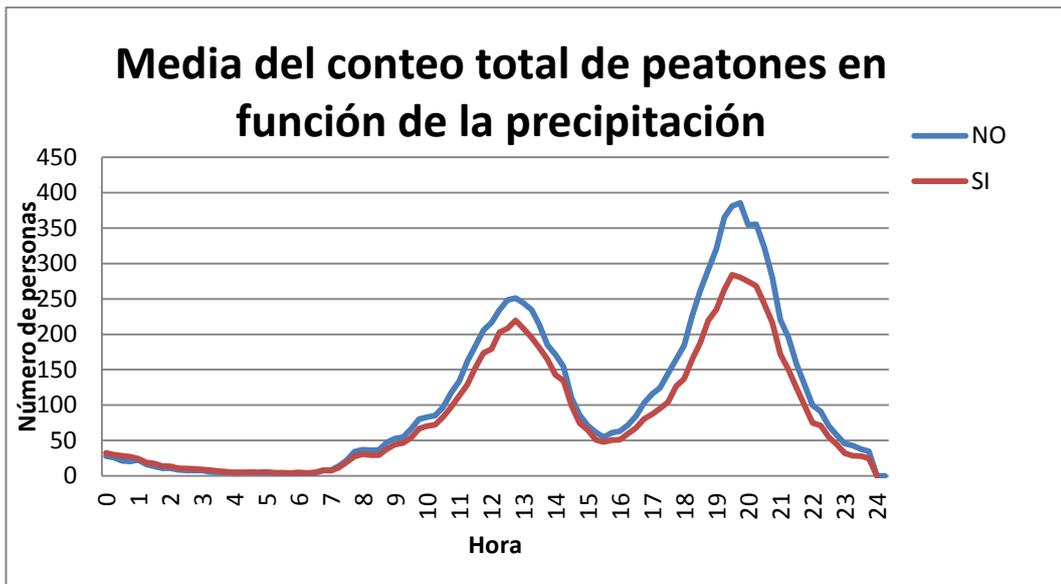
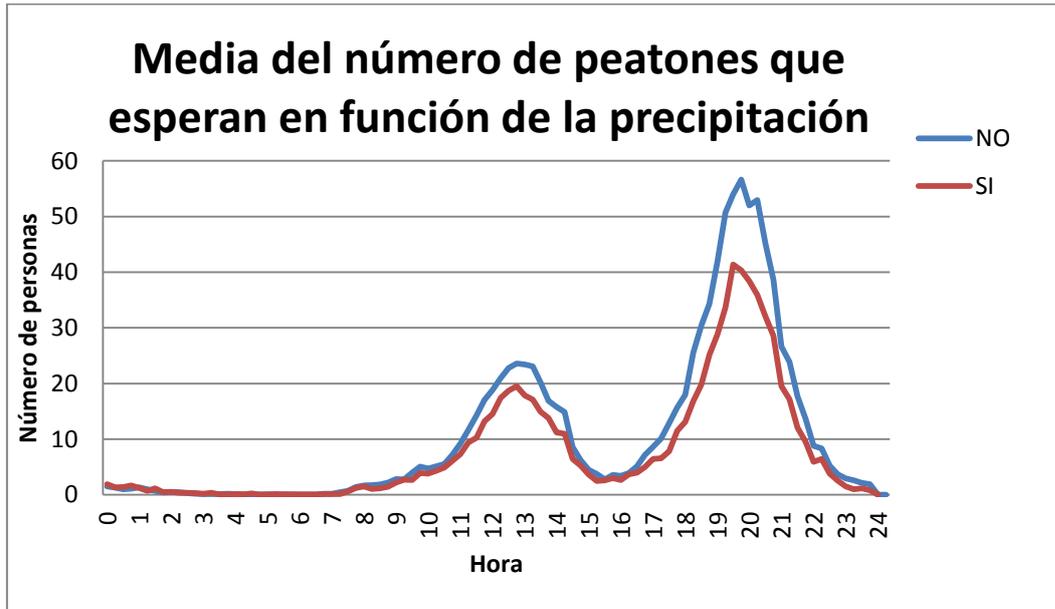


Figura 5.12. Media del conteo total de peatones en función de la precipitación

### Espera total de peatones en función de la precipitación

La figura 5.13 nos muestra la media de peatones que esperan a cruzar en función de si ha llovido durante el día más de 1 mm. Se aprecia que los días lluviosos los viandantes se detienen menos en el semáforo como es esperable para evitar mojarse, especialmente por las tardes.



*Figura 5.13. Media del número de peatones que esperan en función de la precipitación*

### Conteo total de peatones en función del tipo de día

En la figura 5.14 se encuentra la gráfica del conteo total de peatones en función del tipo de día. Se observan diferencias claras entre los 4 tipos de días: los laborables no escolares (días principalmente en Navidad) son los que tienen mayor flujo de peatones tanto por la mañana como por la tarde. El sábado también tiene una afluencia mayor de peatones que los otros 2 tipos de días especialmente por la tarde. También destaca que el pico matinal se produce más tarde sábados y domingos y festivos.

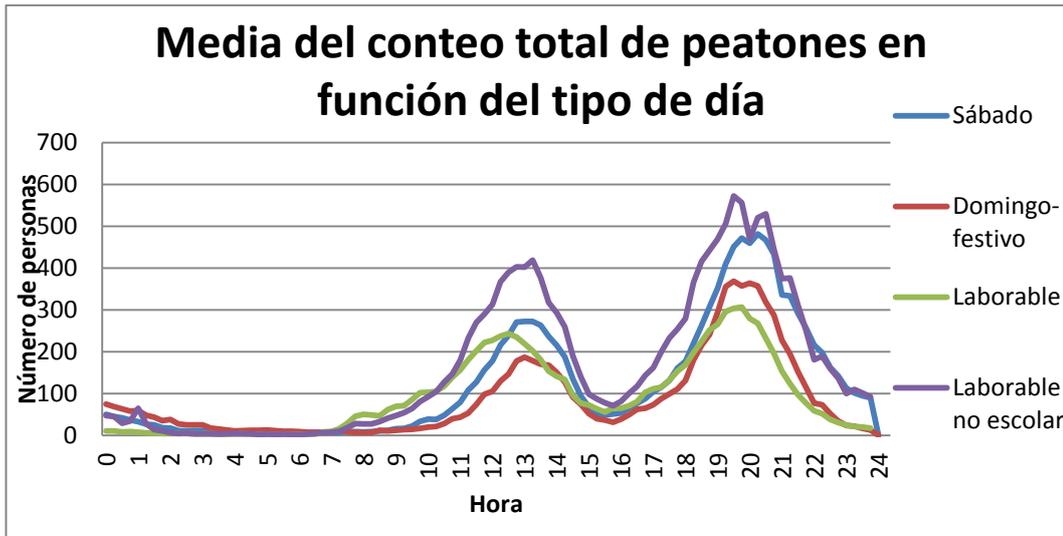


Figura 5.14. Media del conteo total de peatones en función del tipo de día

**Espera total de peatones en función del tipo de día**

En la figura 5.15 se puede apreciar la espera total en función del tipo de día. Se aprecian importantes diferencias entre los tipos de días. En los días laborables los peatones esperan muy poco por la mañana y por la tarde. Los domingos la espera es muy baja por la mañana repunta bastante por la tarde. Los días laborables no escolares son los días que más se espera en ambos picos. Los sábados la espera es ligeramente superior a laborables y domingos-festivos pero por la tarde se encuentra a niveles similares a los de los laborables no escolares.

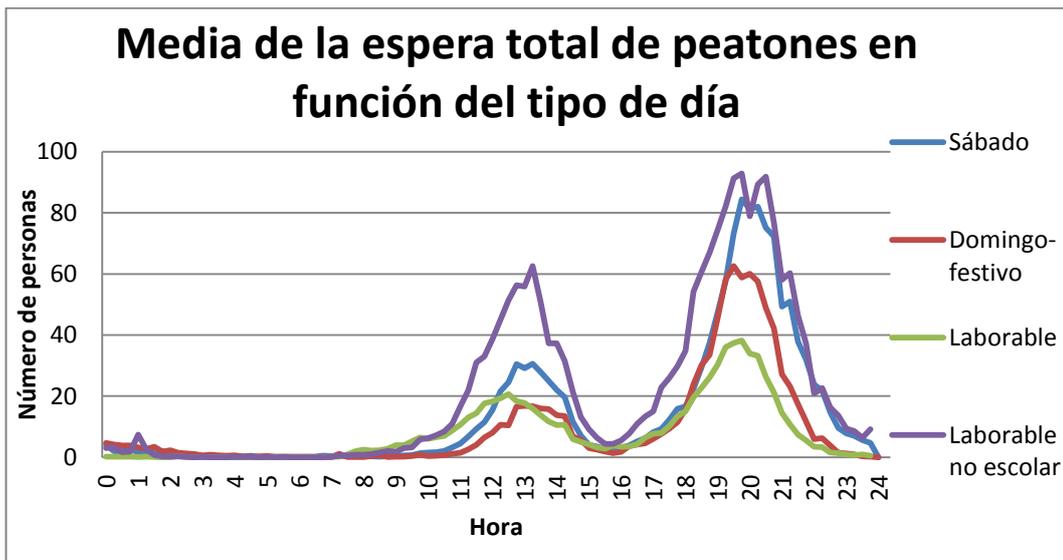
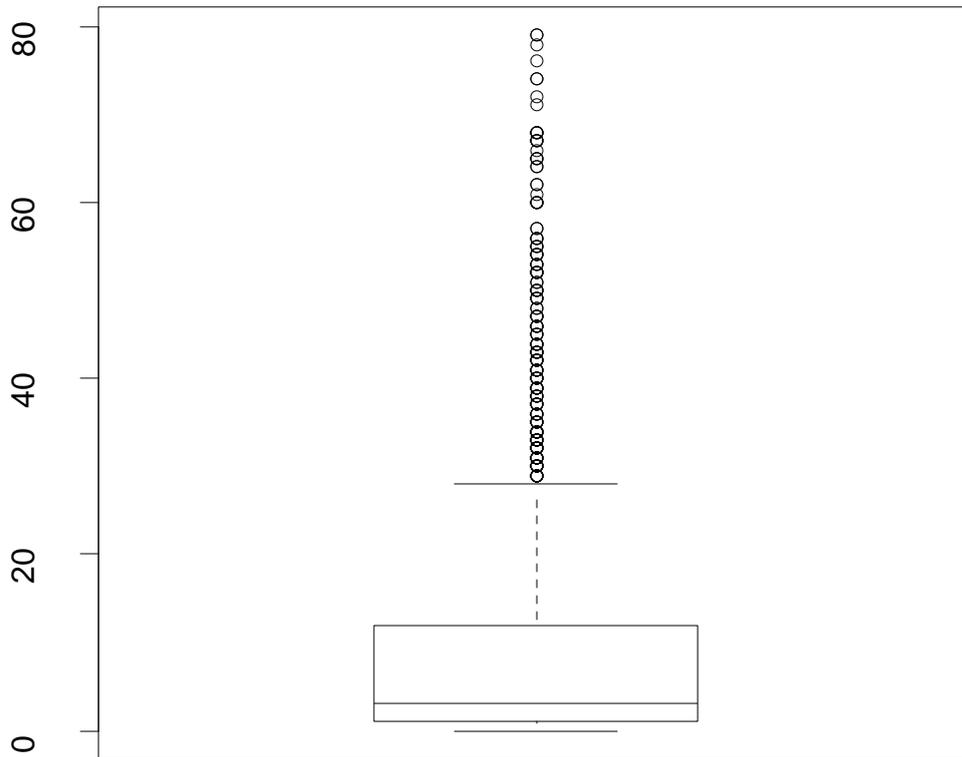


Figura 5.15. Media de la espera total de peatones en función del tipo de día

### 5.2.2 Análisis descriptivo de los tiempos de espera

En los tiempos de espera, en la figura 5.16 se ha realizado un diagrama de cajas y bigotes para conocer la naturaleza de los datos. Se observa que el primer cuartil es 1, la mediana es 3 y el tercer cuartil es 12, resultando el Rango Intercuartílico de 11. El “bigote” inferior se corresponde con el mínimo de los datos que es 0. El “bigote” superior es el tercer cuartil + 1,5 veces el rango intercuartílico, que es 28,5. Por tanto, se aprecia que hay bastantes datos atípicos por encima del “bigote” superior.



*Figura 5.16. Diagrama de cajas y bigotes para los tiempos de espera del 28 y 29 de abril, y del 5 y 6 de mayo de 2014*

Posteriormente, en el capítulo 7 se incluirán más análisis descriptivos de los tiempos de espera como histogramas, que se consideran más adecuados incluir en dicho capítulo para una mejor comprensión de la metodología estadística empleada para analizar la efectividad del semáforo inteligente en la reducción de los tiempos de espera.

## 6. Análisis mediante procedimientos funcionales

En este capítulo se expondrán los análisis realizados mediante los procedimientos funcionales explicados en el capítulo 4 a los datos del flujo de peatones en la calle Santiago.

### 6.1. Análisis ANOVA funcional

El análisis de la varianza funcional se realizará respecto los siguientes factores tanto para los conteos totales y las esperas totales de peatones:

- **Día de la semana.** Indica el día de la semana de la medición. Hay 7 niveles:
  - **Lunes.**
  - **Martes.**
  - **Miércoles.**
  - **Jueves.**
  - **Viernes.**
  - **Sábado.**
  - **Domingo.**
- **Mes.** Muestra el mes de la medición. Se dispone de 7 niveles:
  - **Enero.**
  - **Febrero.**
  - **Marzo.**
  - **Septiembre.**
  - **Octubre.**
  - **Noviembre.**
  - **Diciembre.**
- **Festivo.** Indica si el día es festivo (domingo o fiesta laboral) o no festivo. Sus niveles son:
  - **Festivo**
  - **No festivo**
- **Tipo de día.** Clasifica el día en uno de los siguientes niveles:
  - **Laborable:** incluye todos los días laborables salvo los sábados y aquellos días en los que según el calendario escolar no hay clase.
  - **Laborable no escolar:** en este nivel se encuentran aquellos días laborables en los que los escolares no tienen clase excepto los sábados.
  - **Sábado:** incluye todos los sábados salvo aquellos que sean festivos.
  - **Domingo-festivo:** incorpora todos los domingos y festivos.

- **Ubicación del día en el mes.** Variable categórica con 3 niveles: 1, 2 y 3.
  - **1:** se corresponde con los días comprendidos entre el 1 y el 10 de cada mes (incluidos).
  - **2:** se corresponde con los días comprendidos entre el 11 y el 20 de cada mes (incluidos).
  - **3:** se corresponde con los días comprendidos entre el 21 y el 31 de cada mes (incluidos).
- **Precipitación.** Variable categórica con 2 niveles:
  - **SÍ:** en el día la precipitación acumulada es mayor o igual a 1 mm.
  - **NO:** en el día no ha habido precipitaciones o son inferiores a 1 mm.

El número de réplicas artificiales en el procedimiento bootstrap será de  $N= 10000$ .

Todos los análisis de la varianza se realizarán a un nivel de significación de  $\alpha= 0.01$ .

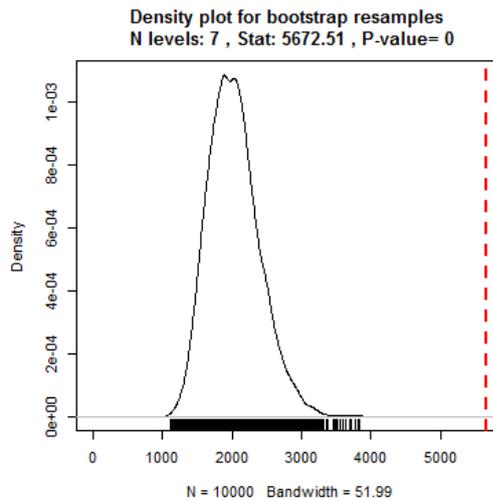
### 6.1.1 ANOVA de los conteos totales

#### ANOVA para el factor día de la semana

Las hipótesis a considerar son:

$$\begin{cases} H_0: m_{Lunes} = m_{Martes} = m_{Miércoles} = m_{Jueves} = m_{Viernes} = m_{Sábado} = m_{Domingo} \\ H_1: \text{Al menos un par de los anteriores es distinto} \end{cases}$$

El valor del estadístico  $V_n$  es 5672.51 y el p-valor es 0, entonces se rechaza la hipótesis nula y por tanto existen diferencias significativas entre los días de la semana. Se confirman las sospechas de que existían diferencias significativas entre días de la semana al comparar las curvas de las medias de peatones según el día de la semana como se vio en la Figura 5.2, probablemente debido al mayor flujo peatonal por la tarde los sábados y a que el pico matinal los sábados y domingos se producía más tarde. En la figura 6.1 se puede observar la gráfica de densidad a partir de las réplicas del procedimiento bootstrap, el valor del estadístico se encuentra claramente en la cola de la función de distribución y por ello su p-valor es prácticamente cero.



*Figura 6.1. Gráfica de la función de densidad de probabilidad de los conteos totales para el factor día de la semana*

### **ANOVA para el factor mes**

Las hipótesis para el análisis de la varianza para el factor mes son:

$$\begin{cases} H_0: m_{Enero} = m_{Febrero} = m_{Marzo} = m_{Septiembre} = \dots = m_{Diciembre} \\ H_1: \text{Al menos un par de los anteriores es distinto} \end{cases}$$

El estadístico  $V_n$  es 4998.96 y el p-valor es 0: rechazamos la hipótesis nula y existen diferencias significativas entre los meses. En la figura 6.2, se puede apreciar la función de densidad para el factor mes. Como vimos en la figura 5.6, las diferencias significativas probablemente se deban al gran flujo de peatones en diciembre tanto por la mañana y por la tarde, y a la gran cantidad de personas en horario nocturno respecto al resto de meses que tienen un comportamiento bastante similar en todas las franjas horarias.

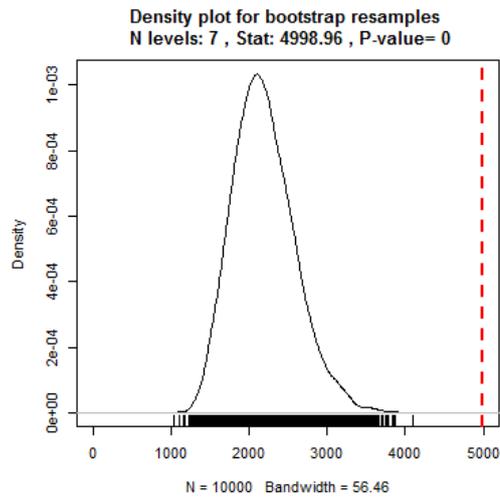


Figura 6.2. Gráfica de la función de densidad de probabilidad de los conteos totales para el factor mes

### ANOVA para el factor festivo

Para el factor festivo se plantean las siguientes hipótesis:

$$\begin{cases} H_0: m_{Festivo} = m_{No festivo} \\ H_1: m_{Festivo} \neq m_{No festivo} \end{cases}$$

En este caso el estadístico  $V_n$  es 274.71, obteniéndose un p-valor a partir de la función de distribución de 0.0002. Se rechaza la hipótesis nula y existen diferencias significativas entre los dos niveles del factor festivo. Las diferencias se deben, como se señaló en la figura 5.4, al pico matinal más tardío y el mayor flujo peatonal de madrugada en los días festivos. En la figura 6.3 se grafica la función de distribución en función del valor del estadístico  $V_n$ .

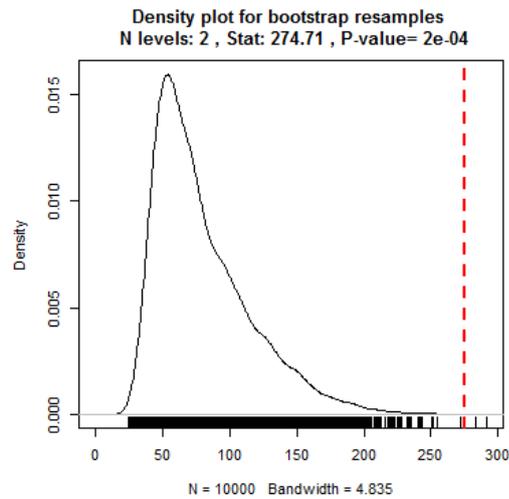


Figura 6.3. Gráfica de la función de densidad de probabilidad de los conteos totales para el factor festivo

### ANOVA para el factor tipo de día

Las hipótesis para el factor tipo de día son:

$$\begin{cases} H_0: m_{Laborable} = m_{Laborable\ no\ escolar} = m_{Sábado} = m_{Domingo-Festivo} \\ H_1: Al\ menos\ un\ par\ de\ los\ anteriores\ es\ distinto \end{cases}$$

El estadístico  $V_n$  resulta ser 2505.749. A partir de la función de densidad obtenida por bootstrap se halla que el p-valor es 0: se rechaza la hipótesis nula y existen diferencias significativas al menos entre un par de los niveles del factor tipo de día. Las diferencias se deben al mayor flujo los laborables no escolares y los sábados comparado con los laborables y los domingos-festivos como se comentó en la figura 5.14.

La gráfica de la función densidad obtenida mediante bootstrap se encuentra en la figura 6.4.

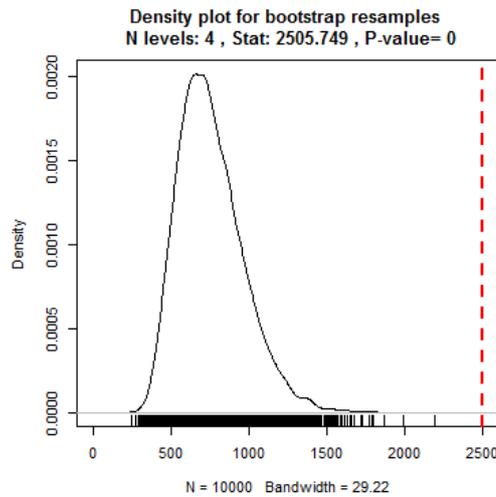


Figura 6.4. Gráfica de la función de densidad de probabilidad de los conteos totales para el factor tipo de día

#### ANOVA para el factor ubicación del día en el mes

Se consideran las siguientes hipótesis para el factor ubicación del día en el mes:

$$\begin{cases} H_0: m_1 = m_2 = m_3 \\ H_1: \text{Al menos un par de los anteriores es distinto} \end{cases}$$

Al realizar el análisis ANOVA para el factor ubicación del día en el mes se tiene que el valor del estadístico  $V_n$  es 250.997. A partir de la función de densidad se obtiene que el p-valor es 0.2688. Entonces el factor ubicación del día en el mes no es significativo. Con este análisis confirmamos las sospechas de que no existían diferencias significativas cuando se comparó las medias de algunos días de principio de mes con otros de final mes que se apreció con la figura 5.9. La gráfica de densidad obtenida a partir del procedimiento bootstrap se encuentra en la figura 6.5.

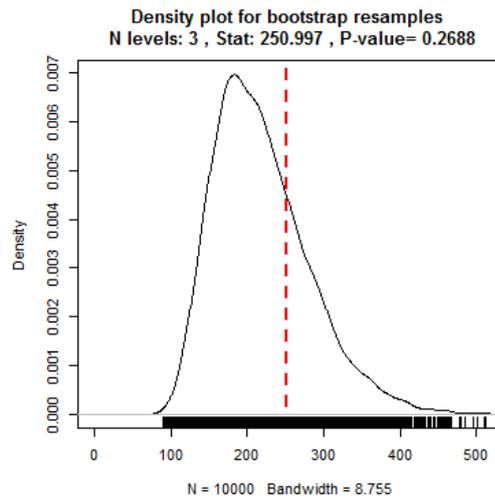


Figura 6.5. Gráfica de la función de densidad obtenida mediante bootstrap para el factor ubicación del día en el mes

### ANOVA para el factor precipitación

Las hipótesis consideradas para el factor precipitación son:

$$\begin{cases} H_0: m_{SI} = m_{NO} \\ H_1: m_{SI} \neq m_{NO} \end{cases}$$

En el caso del factor precipitación, el estadístico  $V_n$  es 221.413 resultando ser su p-valor 0 a partir de la función de densidad. Así se confirman que las diferencias entre los dos niveles son significativas. Como se comentó en la figura 5.12 las diferencias se deben al menor flujo peatonal durante los días lluviosos. En la figura 6.6 se puede apreciar la función de densidad.

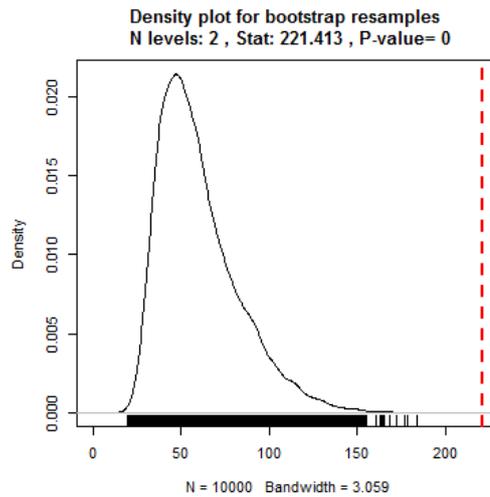


Figura 6.6. Gráfica de la función densidad para el factor precipitación.

## 7.2 ANOVA de la espera total

### ANOVA para el factor día de la semana

Las hipótesis a considerar son:

$$\begin{cases} H_0: m_{Lunes} = m_{Martes} = m_{Miércoles} = m_{Jueves} = m_{Viernes} = m_{Sábado} = m_{Domingo} \\ H_1: \text{Al menos un par de los anteriores es distinto} \end{cases}$$

El estadístico  $V_n$  es 1241.829 y el p-valor obtenido a partir de la función de distribución obtenida por bootstrapping que corresponde a ese valor de  $V_n$  es 0. Las diferencias significativas se deben principalmente a la mayor espera los sábados como se vio en la figura 5.3. En la figura 6.7, se encuentra la gráfica de la función de distribución para el factor día de la semana.

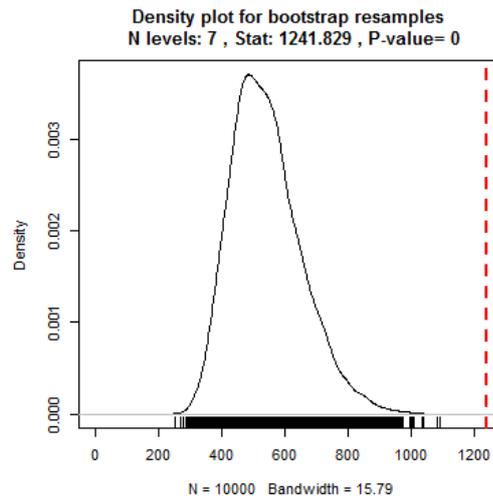


Figura 6.7. Gráfica de la función distribución de para el factor día de la semana

### ANOVA para el factor mes

Para el análisis de la varianza de la espera total para el factor se plantean las siguientes hipótesis:

$$\begin{cases} H_0: m_{Enero} = m_{Febrero} = m_{Marzo} = m_{Septiembre} = m_{Octubre} = m_{Noviembre} = m_{Diciembre} \\ H_1: \text{Al menos un par de los anteriores es distinto} \end{cases}$$

En este caso el estadístico  $V_n$  es 1417.907. El cálculo del p-valor a partir de la función de densidad obtenida mediante bootstrap es 0. Por tanto, existen diferencias significativas al menos entre 2 meses. Seguramente esas diferencias se deban a la mayor espera en el mes de diciembre respecto al resto de meses como se analizó en la figura 5.8. En la figura 6.8, se puede ver la gráfica de la función de densidad.

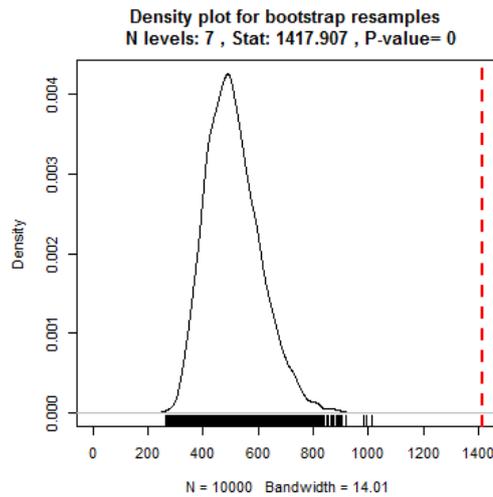


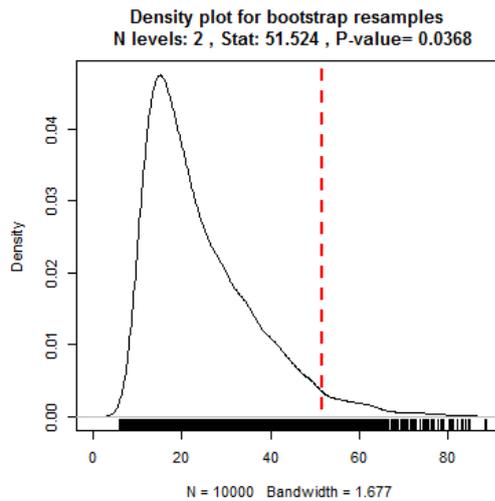
Figura 6.8. Gráfica de la función de distribución de la espera total para el factor mes

### ANOVA para el factor festivo

Para el factor festivo, se eligen estas hipótesis:

$$\begin{cases} H_0: m_{Festivo} = m_{No festivo} \\ H_1: m_{Festivo} \neq m_{No festivo} \end{cases}$$

El estadístico  $V_n$  es 51.524. Empleando la función de distribución calculada por bootstrapping se obtiene que el p-valor es 0.0368: no existen diferencias significativas entre los días festivos y no festivos a un nivel de significación 0.01. Sin embargo, a un nivel de significación 0.05 sí que habría diferencias significativas. En la figura 6.9, se muestra la gráfica de la función de distribución según el valor del estadístico  $V_n$ .



*Figura 6.9. Gráfica de la función de densidad de la espera total para el factor festivo*

### **ANOVA para el factor tipo de día**

Para el factor tipo de día se plantean estas hipótesis:

$$\begin{cases} H_0: m_{Laborable} = m_{Laborable\ no\ escolar} = m_{Sábado} = m_{Domingo-Festivo} \\ H_1: Al\ menos\ un\ par\ de\ los\ anteriores\ es\ distinto \end{cases}$$

El estadístico  $V_n$  tiene un valor de 559.94. El p-valor a partir de la función de distribución es 0: por tanto, existen diferencias significativas entre al menos un par de los tipos de día. Esto se observó claramente en la figura 5.15, siendo los sábados y los laborables no escolares días con mayor espera que los otros dos tipos de días. En la figura 6.10, se puede apreciar la función de densidad según el valor de  $V_n$ .

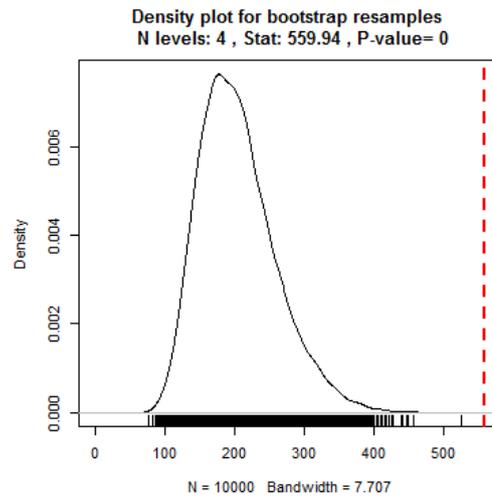


Figura 6.10. Gráfica de la función de densidad de la espera total según el tipo de día

### ANOVA para el factor ubicación del día en el mes

Se enuncian las siguientes hipótesis para el factor ubicación del día en el mes:

$$\begin{cases} H_0: m_1 = m_2 = m_3 \\ H_1: \text{Al menos un par de los anteriores es distinto} \end{cases}$$

El estadístico  $V_n$  es 54.376. El p-valor calculado mediante la función de densidad hallada mediante el procedimiento bootstrap es 0.4788. No existen diferencias significativas en la espera total entre los 3 niveles del factor. La función de distribución se muestra graficada en la figura 6.11.

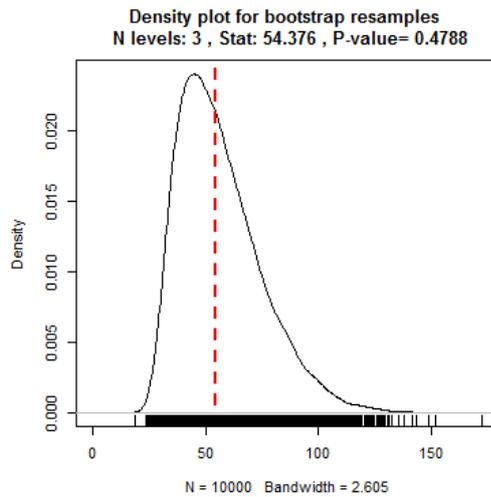


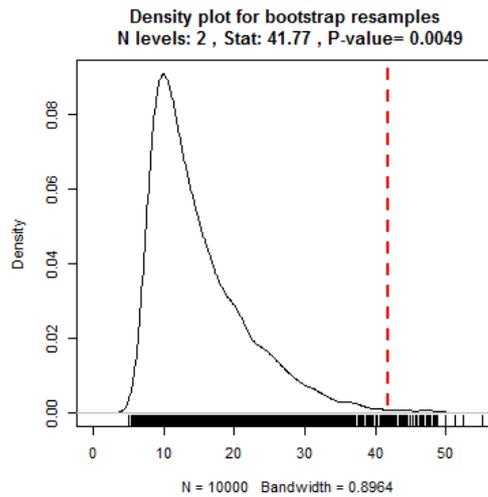
Figura 6.11. Gráfica de la función de densidad de la espera total según la ubicación del día en el mes

### ANOVA para el factor precipitación

Se plantean estas 2 hipótesis:

$$\begin{cases} H_0: m_{SI} = m_{NO} \\ H_1: m_{SI} \neq m_{NO} \end{cases}$$

El estadístico  $V_n$  es 41,77. Con la ayuda de la función de densidad obtenida por bootstrapping, se obtiene que el p-valor es 0.0049: existen diferencias significativas entre los días con precipitación y los que no llueve. Se vio en la figura 5.13 que los días de lluvia los viandantes se detienen menos en el paso. En la figura 6.12, se adjunta la gráfica de la función de distribución según el valor  $V_n$



*Figura 6.12. Gráfica de la función densidad para la espera total según la precipitación*

A continuación, en la tabla 6.1 se adjunta una tabla resumen con los p-valores de todos los análisis de la varianza y su significación estadística para los distintos factores.

Factor	Conteos totales		Espera total	
	p-valor	¿Significativo con $\alpha= 0.01$ ?	p-valor	¿Significativo con $\alpha= 0.01$ ?
Día de la semana	0	Sí	0	Sí
Mes	0	Sí	0	Sí
Festivo	0.0002	Sí	0.0368	No, pero a un nivel de significación $\alpha= 0.05$ sí.
Tipo de día	0	Sí	0	Sí
Ubicación del día en el mes	0.2688	No	0.4788	No
Precipitación	0	Sí	0.0049	Sí

*Tabla 6.1. Tabla resumen de p-valores y significación de factores para conteos totales y espera total.*

Se aprecia que los factores que son significativos para los conteos totales, también lo son para la espera total excepto en el caso del factor festivo. Esto puede deberse a que aunque parezca claro que existen diferencias significativas en el conteo total de días festivos y no festivos, dichas diferencias no son tan grandes como para producir una modificación del comportamiento de los peatones al esperar en el semáforo a un nivel de significación  $\alpha= 0.01$ .

## 6.2. Análisis clúster funcional

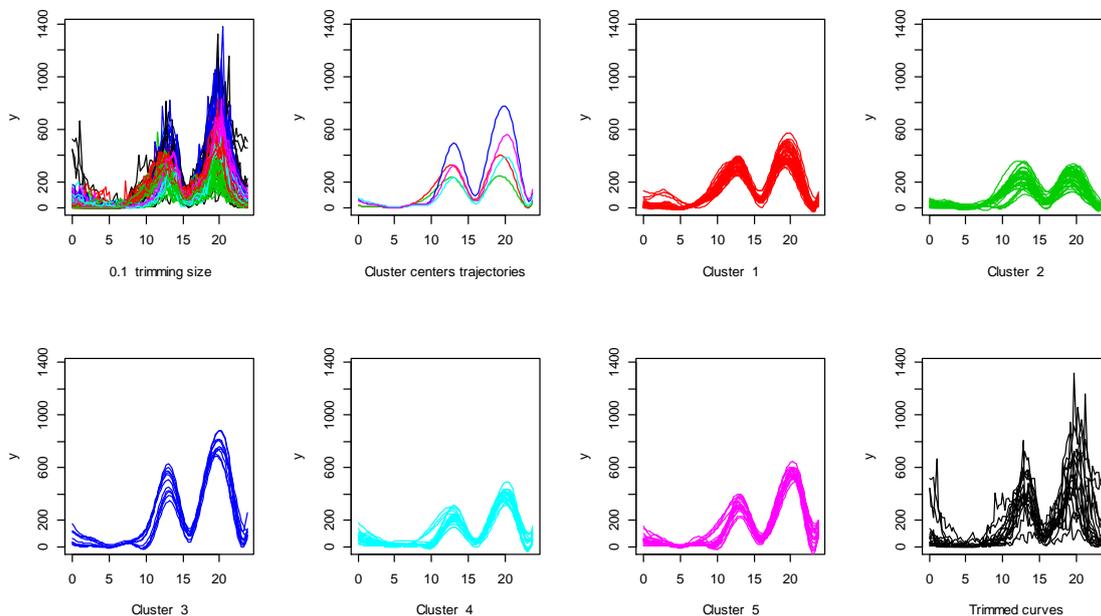
El análisis clúster mediante el procedimiento de las k-medias recortadas será utilizado para dividir las diversas observaciones en varios grupos o clústers.

Se emplearán tanto para los conteos totales como las esperas con 8 nodos interiores, y como el número de elementos del b-spline cúbico es 4 más que el número de nodos, se tendrán 12 elementos en el b-spline cúbico.

Se decide realizar 2000 iteraciones al algoritmo de las k-medias recortadas.

### 6.2.1. Análisis clúster de los conteos totales

En primer lugar se analiza los conteos totales de peatones que pasan por el cruce. Mediante prueba y error se varía los valores del número de clústeres  $K$  y el porcentaje de recorte  $\alpha$  hasta obtener aquella combinación que proporciona una mejor diferenciación clara entre los clústeres desde un punto de vista cualitativo. Con  $K=5$  clústeres y un tamaño recorte de 0.1 se aprecia una clara diferenciación entre clústeres, como se puede ver en la figura 6.13.



*Figura 6.13. Curvas de los clústeres para los datos funcionales de los conteos totales de peatones con  $\alpha=0,1$  y  $K=5$  (considerando 8 nodos).*

Los días elegidos para cada clúster son los siguientes (los días de meses de septiembre a diciembre corresponden al año 2014 y los de enero a marzo corresponden al 2015):

**Clúster 1:** 15/9, 16/9, 17/9, 18/9, 22/9, 24/9, 25/9, 29/9, 30/9, 1/10, 2/10, 7/10, 14/10, 20/10, 21/10, 22/10, 26/10, 7/11, 18/11, 19/11, 20/11, 21/11, 25/11, 26/11, 28/11, 1/12, 2/12, 3/12, 4/12, 9/12, 11/12, 15/12, 16/12, 17/12, 18/12, 19/12, 8/1, 9/1, 12/1, 13/1, 14/1, 16/1, 19/1, 21/1, 22/1, 23/1, 26/1, 27/1, 28/1, 9/2, 10/2, 11/2, 12/2, 16/2, 19/2, 27/2, 2/3, 3/3.

**Clúster 2:** 19/9, 6/10, 8/10, 9/10, 13/10, 15/10, 16/10, 23/10, 27/10, 28/10, 30/10, 31/10, 3/11, 4/11, 5/11, 6/11, 10/11, 11/11, 12/11, 13/11, 14/11, 16/11, 17/11, 6/1, 15/1, 18/1, 20/1, 29/1, 30/1, 1/2, 2/2, 3/2, 4/2, 5/2, 6/2, 8/2, 13/2, 17/2, 18/2, 20/2, 23/2, 24/2, 25/2, 26/2, 4/3.

**Clúster 3:** 6/12, 7/12, 20/12, 21/12, 22/12, 23/12, 26/12, 30/12, 2/1, 4/1, 5/1.

**Clúster 4:** 28/9, 3/10, 5/10, 10/10, 17/10, 19/10, 1/11, 2/11, 8/11, 9/11, 15/11, 23/11, 14/12, 11/1, 25/1, 31/1, 7/2, 14/2, 22/2, 1/3.

**Clúster 5:** 20/9, 26/9, 4/10, 11/10, 18/10, 24/10, 25/10, 22/11, 29/11, 30/11, 5/12, 12/12, 13/12, 28/12, 1/1, 17/1, 24/1, 15/2, 21/2, 28/2.

**No asignados a ningún clúster:** 12/9, 13/9, 14/9, 21/9, 23/9, 27/9, 12/10, 24/11, 8/12, 24/12, 25/12, 27/12, 29/12, 31/12, 3/1, 7/1, 10/1.

En la figura 8.2 se puede observar la distribución de los días en cada clúster en el calendario:

Septiembre-14							Octubre-14							Noviembre-14						
L	M	X	J	V	S	D	L	M	X	J	V	S	D	L	M	X	J	V	S	D
1	2	3	4	5	6	7			1	2	3	4	5						1	2
8	9	10	11	12	13	14	6	7	8	9	10	11	12	3	4	5	6	7	8	9
15	16	17	18	19	20	21	13	14	15	16	17	18	19	10	11	12	13	14	15	16
22	23	24	25	26	27	28	20	21	22	23	24	25	26	17	18	19	20	21	22	23
29	30						27	28	29	30	31			24	25	26	27	28	29	30

Diciembre-15							Enero-15							Febrero-15						
L	M	X	J	V	S	D	L	M	X	J	V	S	D	L	M	X	J	V	S	D
1	2	3	4	5	6	7				1	2	3	4							1
8	9	10	11	12	13	14	5	6	7	8	9	10	11	2	3	4	5	6	7	8
15	16	17	18	19	20	21	12	13	14	15	16	17	18	9	10	11	12	13	14	15
22	23	24	25	26	27	28	19	20	21	22	23	24	25	16	17	18	19	20	21	22
29	30	31					26	27	28	29	30	31		23	24	25	26	27	28	

Marzo-15						
L	M	X	J	V	S	D
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	

Color del fondo:

Días sin datos no incorporados al estudio

Clúster 1 

Clúster 2 

Clúster 3 

Clúster 4 

Clúster 5 

Sin asignar a ningún clúster (recortados) 

Figura 6.14. Calendario de los clústers para el conteo total.

El primer clúster tiene principalmente días de lunes a viernes fuera del período de Navidad.

El segundo clúster se corresponde con lunes a viernes y algunos domingos con baja actividad peatonal también fuera del período navideño.

El tercer clúster agrupa días en los que muchas personas tienen vacaciones: 2 días del puente del 6 al 8 de diciembre y muchos días del período navideño. Se tratan de días en los que se realizan muchas compras y es el clúster con mayores picos de peatones tanto por la mañana como por la tarde.

El cuarto clúster junta bastantes sábados y domingos y algunos viernes con bajo flujo de peatones.

El quinto clúster tiene muchos sábados y algunos viernes y domingos con más flujo peatonal, además del 1 enero (jueves).

Se aprecia que el 10 % de días que no se asigna a ningún clúster corresponden en su mayoría a días especiales en los que se puede apreciar las causas del comportamiento anormal del flujo de peatones. Algunos de esos días se pueden ver representados en la figura 6.15. Estos son los sucesos que ocurrieron en algunos de esos días y que pueden explicar el comportamiento anómalo de los peatones de dichos días:

- El 12, 13 y 14 de septiembre era el último fin de semana de las ferias y fiestas locales en Valladolid. Estos 3 días hay mucha actividad de peatones por la noche.
- El 21 y 27 de septiembre hubo manifestaciones que en su itinerario pasaban por el cruce estudiado. [Fuentes: <http://www.elnortedecastilla.es/valladolid/201409/21/medio-millar-personas-exigen-20140921171622.html>, <http://www.elnortedecastilla.es/valladolid/201409/27/personas-sordas-reclaman-igualdad-20140927132758.html>. Fecha de consulta: Julio-2015]
- El 12 de octubre y el 8 de diciembre eran días de fiestas nacionales. Los picos de peatones de la mañana y de la tarde del día 12 de octubre son los de menor afluencia de todos los que no se incluyen en ningún clúster.
- El 24, 25, y 31 diciembre pertenecen al período navideño pero las tradicionales comidas o cenas que se producen estos días probablemente reducen el flujo de peatones en ciertas franjas horarias.
- El 7 de enero es el primer día de rebajas y uno de los días con mayor número de devoluciones tras el día de Reyes. En la figura 8.3 se aprecia el mayor flujo matinal respecto lo normal.

- El 10 de enero es el primer sábado de rebajas y además muchos moteros acudieron a Valladolid a pesar de que se suspendió la concentración motera “Pingüinos”. [Fuente: <http://www.elnortedecastilla.es/valladolid/pinguinos/201501/10/espiritu-pinguinos-sigue-vivo-20150110170715.html>. Fecha de consulta: Julio 2015]

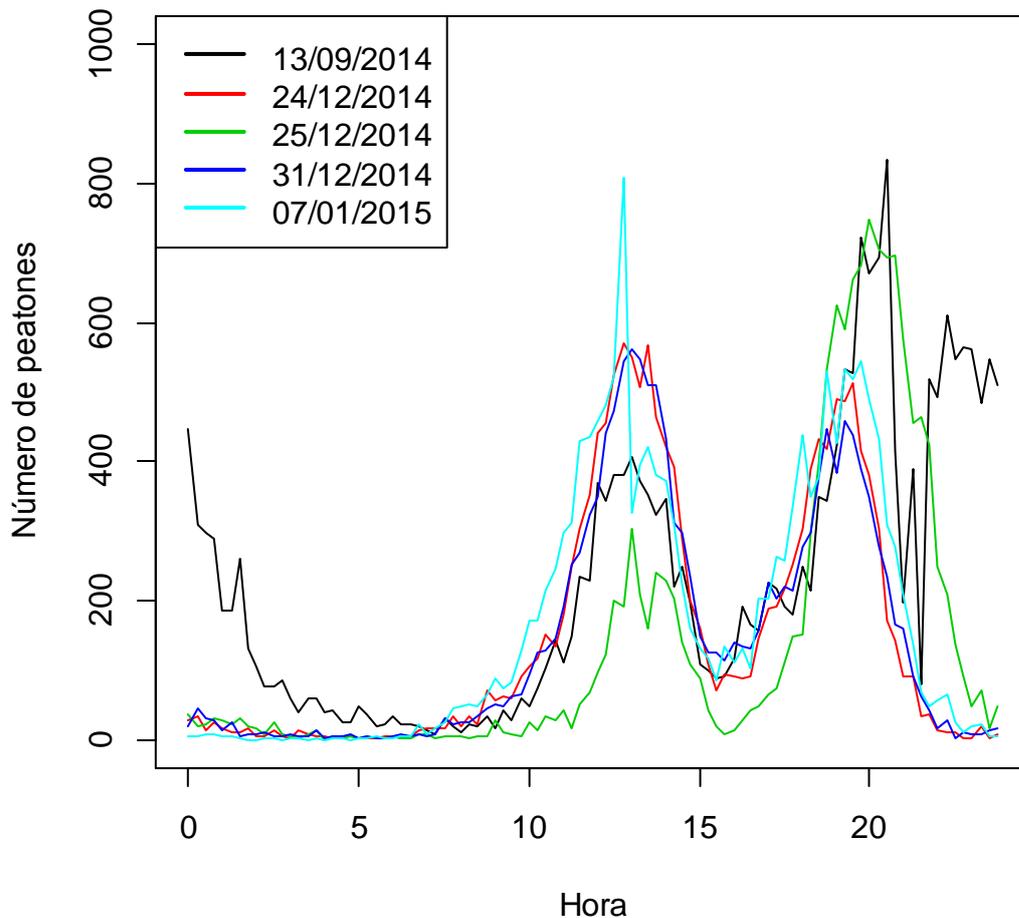
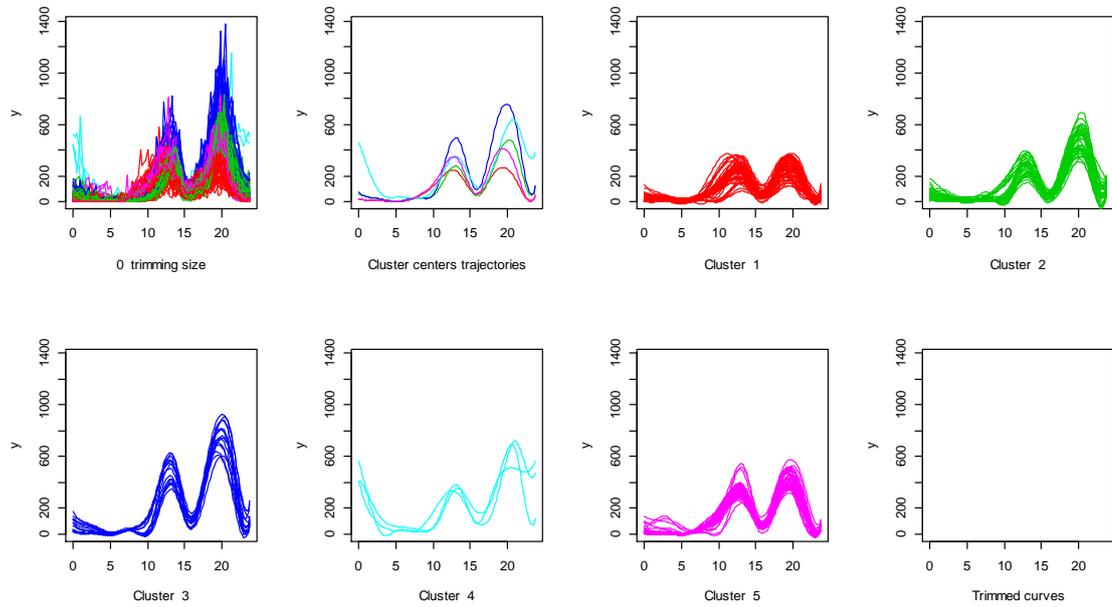


Figura 6.15. Gráfica sin suavizar de algunos de los días no incluidos en ningún clúster (10%) para los conteos totales.

La importancia de usar el método de las medias recortadas se observa claramente en la figura 6.16, al llevar a cabo el mismo análisis clúster pero con  $\alpha = 0$ . Los clústeres no se encuentran tan bien definidos y se encuentran curvas que tienen un comportamiento muy diferente en ciertas franjas horarias a las del

resto de su clúster. Es llamativo que con  $\alpha = 0$ , los 3 días de fiestas (12/9, 13/9 y 14/9) son tan diferentes especialmente en las horas nocturnas al resto de días que pueden constituir un clúster propio (clúster 4).



*Figura 6.16. Curvas de los clústeres para los datos funcionales de los conteos totales de peatones con  $\alpha = 0$  y  $K = 5$  (considerando 8 nodos).*

## 6.2.2 Análisis clúster de la espera total

Para el análisis de la espera se obtiene, como con los conteos totales, mediante prueba y error la mejor combinación entre el número de clústeres  $K$  y el tamaño de recorte  $\alpha$ .

La mejor combinación es de nuevo 5 clústeres y un tamaño de recorte de 0.1. En la figura 6.17 se puede apreciar los diferentes clústeres seleccionados.

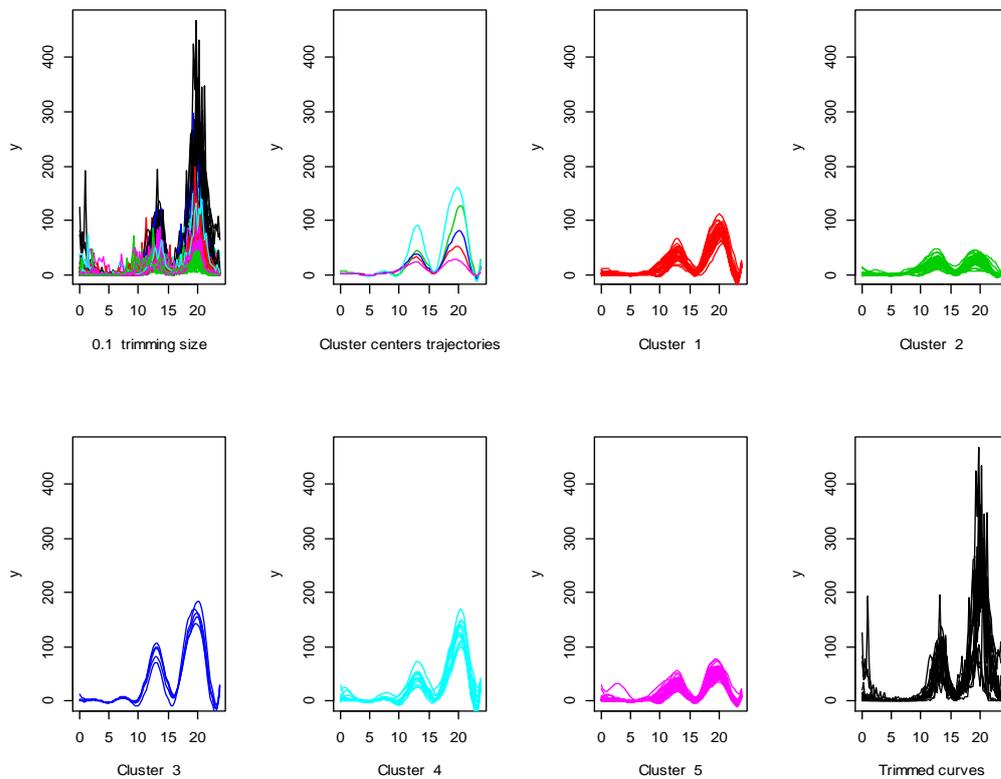


Figura 6.17. Curvas de los clústeres para los datos funcionales de la espera total de peatones con  $\alpha=0.1$  y  $K=5$  (considerando 8 nodos).

Los días seleccionados en cada clúster son los siguientes (los días de meses de septiembre a diciembre corresponden al año 2014 y los de enero a marzo corresponden al 2015):

**Clúster 1:** 15/9, 18/9, 24/9, 25/9, 26/9, 28/9, 29/9, 30/9, 1/10, 2/10, 3/10, 5/10, 10/10, 17/10, 21/10, 1/11, 8/11, 21/11, 28/11, 1/12, 4/12, 13/12, 15/12, 16/12, 18/12, 19/12, 23/12, 9/1, 17/1, 24/1, 31/1, 14/2, 15/2, 21/2, 27/2.

**Clúster 2:** 19/9, 23/9, 7/10, 8/10, 12/10, 13/10, 15/10, 27/10, 28/10, 30/10, 3/11, 4/11, 10/11, 11/11, 12/11, 13/11, 14/11, 16/11, 6/1, 11/1, 12/1, 13/1, 15/1, 18/1, 20/1, 21/1, 22/4, 27/1, 28/1, 29/1, 30/1, 1/2, 2/2, 3/2, 4/2, 5/2, 6/2, 7/2, 8/2, 9/2, 11/2, 13/2, 17/2, 18/2, 19/2, 20/2, 23/2, 24/2, 25/2, 26/2, 4/3.

**Clúster 3:** 8/12, 22/12, 29/12, 30/12, 2/1.

**Clúster 4:** 20/9, 4/10, 11/10, 18/10, 19/10, 25/10, 31/10, 22/11, 30/11, 5/12, 12/12, 28/12, 1/1, 28/2.

**Clúster 5:** 16/9, 17/9, 21/9, 22/9, 27/9, 6/10, 9/10, 14/10, 16/10, 20/10, 22/10, 23/10, 24/10, 26/10, 2/11, 5/11, 6/11, 7/11, 9/11, 15/11, 17/11, 18/11, 19/11, 20/11, 23/11, 24/11, 25/11, 26/11, 2/12, 3/12, 9/12, 11/12, 14/12, 17/12, 8/1, 10/1, 14/1, 16/1, 19/1, 23/1, 25/1, 26/1, 10/2, 12/2, 16/2, 22/2, 1/3, 2/3, 3/3.

**Sin asignar a ningún clúster:** 12/9, 13/9, 14/9, 29/11, 6/12, 7/12, 20/12, 21/12, 24/12, 25/12, 26/12, 27/12, 31/12, 3/1, 4/1, 5/1, 7/1.

Septiembre-14							Octubre-14							Noviembre-14						
L	M	X	J	V	S	D	L	M	X	J	V	S	D	L	M	X	J	V	S	D
1	2	3	4	5	6	7			1	2	3	4	5						1	2
8	9	10	11	12	13	14	6	7	8	9	10	11	12	3	4	5	6	7	8	9
15	16	17	18	19	20	21	13	14	15	16	17	18	19	10	11	12	13	14	15	16
22	23	24	25	26	27	28	20	21	22	23	24	25	26	17	18	19	20	21	22	23
29	30						27	28	29	30	31			24	25	26	27	28	29	30

Diciembre-14							Enero-15							Febrero-15						
L	M	X	J	V	S	D	L	M	X	J	V	S	D	L	M	X	J	V	S	D
1	2	3	4	5	6	7				1	2	3	4							1
8	9	10	11	12	13	14	5	6	7	8	9	10	11	2	3	4	5	6	7	8
15	16	17	18	19	20	21	12	13	14	15	16	17	18	9	10	11	12	13	14	15
22	23	24	25	26	27	28	19	20	21	22	23	24	25	16	17	18	19	20	21	22
29	30	31					26	27	28	29	30	31		23	24	25	26	27	28	

Marzo-15						
L	M	X	J	V	S	D
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	

Color del fondo:

Días sin datos no incorporados al estudio

Clúster 1 

Clúster 2 

Clúster 3 

Clúster 4 

Clúster 5 

Sin asignar a ningún clúster (recortados) 

Figura 6.18. Calendario de los clústeres para la espera total.

En el primer clúster no se observa un patrón claro, ya que en él se encuentran días de cualquier mes y de cualquier día de la semana. Hay algunos viernes y sábados más que el resto de días de la semana.

El segundo clúster se corresponde con días en el que los peatones han esperado muy poco tiempo, tanto por la mañana como por la tarde. Los días incluidos son principalmente de enero y febrero, de domingo a viernes (solamente hay un sábado en este clúster).

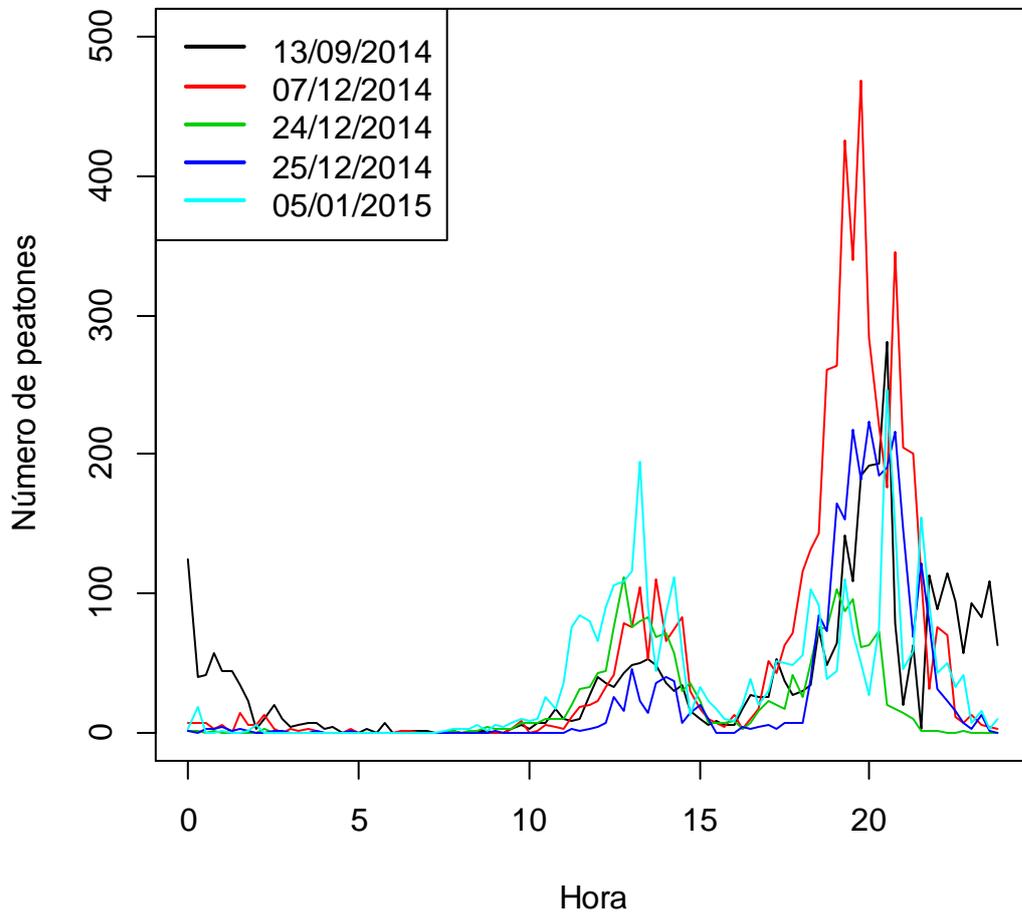
El tercer clúster es muy reducido con solo 5 días. Los días recogidos son de gran afluencia en la calle Santiago debido a las compras navideñas, y por tanto es lógico que sea el clúster con mayores picos de espera por la mañana y por la tarde.

El cuarto clúster incluye especialmente sábados y vísperas de festivos, además de algún viernes y domingo. Se trata del segundo clúster con mayor espera tras el clúster 3.

El quinto clúster recoge días de cualquier día de la semana y mes, con una espera intermedia entre los clúster 1 y 2 que también recogían días de cualquier día de la semana.

En la figura 6.18 se puede observar la distribución de los clústers en el calendario.

Los días festivos correspondientes al último fin de semana de las fiestas de Valladolid, al puente de Diciembre y días del período de vacaciones de Navidad son los que se quedan sin asignar a ningún clúster. La mayor parte de esos días coinciden con los que no se incluían en ningún clúster también en los conteos totales. Por tanto, muchas de esas explicaciones dadas para el flujo del conteo total de dichos días, también son aplicables para los días no incluidos en ningún clúster para la espera total. El 5 de enero no estaba incluido en un clúster en los conteos totales pero en la espera total sí. Es bastante posible que pasara la cabalgata de Reyes por esa calle y la gente se encontrase parada viéndola modificándose el comportamiento normal de la espera. En la figura 6.19, se puede observar algunos de los días no incluidos en ningún clúster de la espera total.



*Figura 6.19. Gráfica sin suavizar de algunos de los días no incluidos en ningún clúster (10%) para la espera total.*

De nuevo se puede comprobar la importancia de utilizar el método de las k-medias recortadas en la figura 6.20, ya que al elegir un  $\alpha=0$ , los clúster quedan peor definidos y las curvas de cada clúster no son tan similares entre sí.

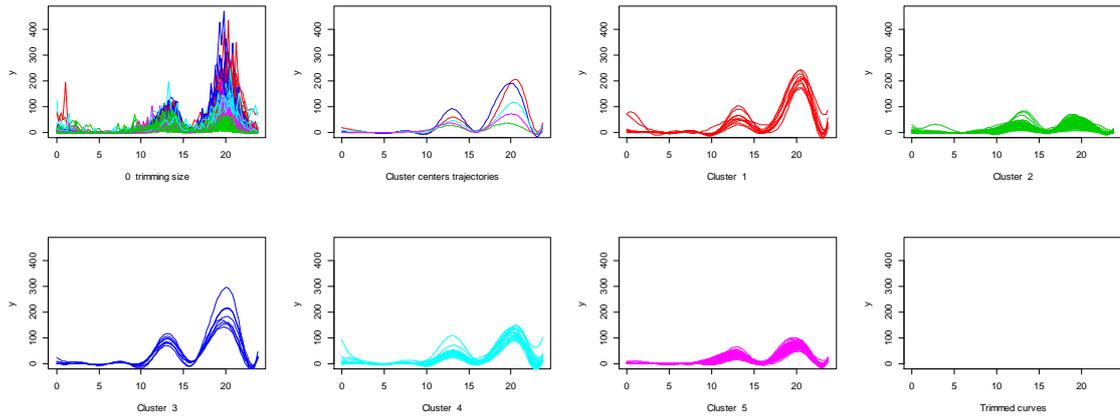


Figura 6.20. Curvas de los clústeres para los datos funcionales de la espera de peatones con  $\alpha=0$  y  $K=5$  (considerando 8 nodos).

### 6.2.3. Relación entre clústeres de conteos totales y esperas totales

En la tabla 6.2 se adjunta la relación de los días en común entre los diferentes clústeres para los conteos y esperas totales. De color rojo se ha resaltado aquellos clústeres que tienen en común mayor número de días.

		Espera total						
		Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Sin asignar	
Conteos totales	Clúster 1	19	10			29		58
	Clúster 2		37		1	7		45
	Clúster 3	1		3			7	11
	Clúster 4	9	2		1	8		20
	Clúster 5	6			12	1	1	20
	Sin asignar		2	2		4	9	17
		35	51	5	14	49	17	

Tabla 6.2. Tabla de la relación de días en común de los clústeres de los conteos y espera totales.

Debido a que los clústeres tienen distintos tamaños puede ser más interesante ver las relaciones entre clústeres en porcentajes. En la figura 6.3 está la tabla del porcentaje de cada clúster de los conteos totales en común con los clústeres de la espera total. Se ha resaltado de color rojo los porcentajes máximos. Los porcentajes máximos son altos, por encima del 45 % en todos los casos.

		Espera total					
		Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Sin asignar
Conteos totales	Clúster 1	32.76%	17.24 %			50%	
	Clúster 2		82.22 %		2.22 %	15.56 %	
	Clúster 3	9.09 %		27.27 %			63.64 %
	Clúster 4	45 %	10 %		5 %	40 %	
	Clúster 5	30 %			60 %	5 %	5 %
	Sin asignar		11.76 %	11.76 %		23.53 %	52.94 %

*Tabla 6.3. Porcentaje de cada clúster del conteo total en común con los clústers de la espera total.*

En la tabla 6.4 se realiza al contrario, se observa el porcentaje de observaciones de los clústeres de la espera total que se tiene en común con los clústeres del conteo total. Se resalta de color rojo el porcentaje máximo en cada clúster de la espera total. De nuevo los porcentajes máximos son muy altos en este caso por encima del 52 %.

		Espera total					
		Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Sin asignar
Conteos totales	Clúster 1	54.29%	19.61%			59.18%	
	Clúster 2		72.55%		7.14%	14.29%	
	Clúster 3	2.86%		60%			41.18%
	Clúster 4	25.71%	3.92%		7.14%	16.33%	
	Clúster 5	17.14%			85.71%	2.04%	5.88%
	Sin asignar		3.92%	40%		8.16%	52.94%

*Tabla 6.4. Porcentaje de cada clúster de la espera total en común con los clústers de los conteos totales.*

La relación entre los clústeres de los conteos y esperas totales es muy fuerte como se ha visto. Especialmente en los 3 casos en los que los porcentajes máximos de un clúster para conteos y esperas coincidían:

- Clúster 2 conteo total y clúster 2 espera total.
- Clúster 5 conteo total y clúster 4 espera total.
- Observaciones sin asignar a ningún clúster del conteo total y espera total.

Por último, en la tabla 6.5 se muestra el porcentaje de observaciones del total que tiene en común cada uno de los pares de clústeres de los conteos totales y espera total.

		Espera total						
		Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Sin asignar	
Conteos totales	Clúster 1	11,11%	5,85%			16,96%		33,92%
	Clúster 2		21,64%		0,58%	4,09%		26,32%
	Clúster 3	0,58%		1,75%			4,09%	6,43%
	Clúster 4	5,26%	1,17%		0,58%	4,68%		11,70%
	Clúster 5	3,51%			7,02%	0,58%	0,58%	11,70%
	Sin asignar		1,17%	1,17%		2,34%	5,26%	9,94%
		20,47%	29,82%	2,92%	8,19%	28,65%	9,94%	

*Tabla 6.5. Porcentaje de observaciones del total en común de cada par de clústeres de los conteos totales y espera total.*

### 6.3. Análisis ACP funcional

Para el análisis de componentes principales funcional se emplearán 96 funciones de base (92 nodos interiores) para proyectar las curvas. Se buscarán los principales modos de variación de los conteos totales y de la espera total, así como reducir el problema a un número de dimensiones inferior manteniendo la mayor parte de la variabilidad

#### 6.3.1. ACP de los conteos totales

La primera componente explica el 63.5 % de la variabilidad. En la figura 6.21 se observa que un efecto positivo aumenta los 2 picos de la mañana y la tarde, y el negativo disminuye dichos picos.

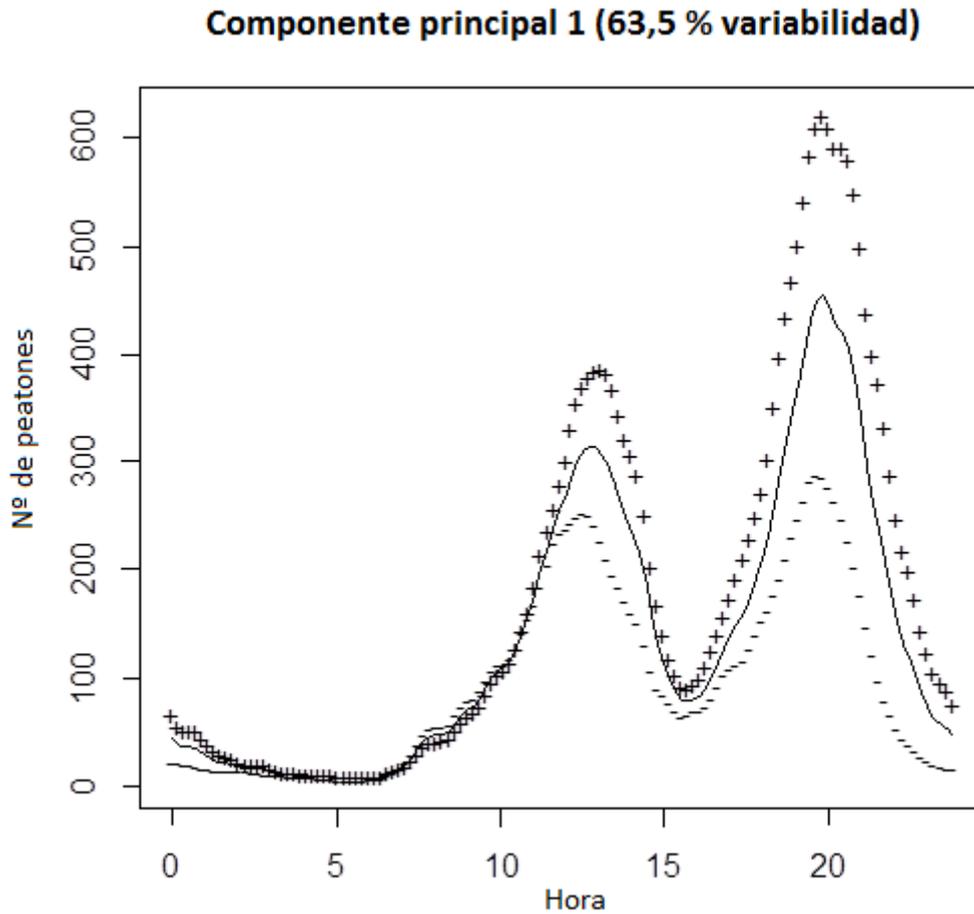
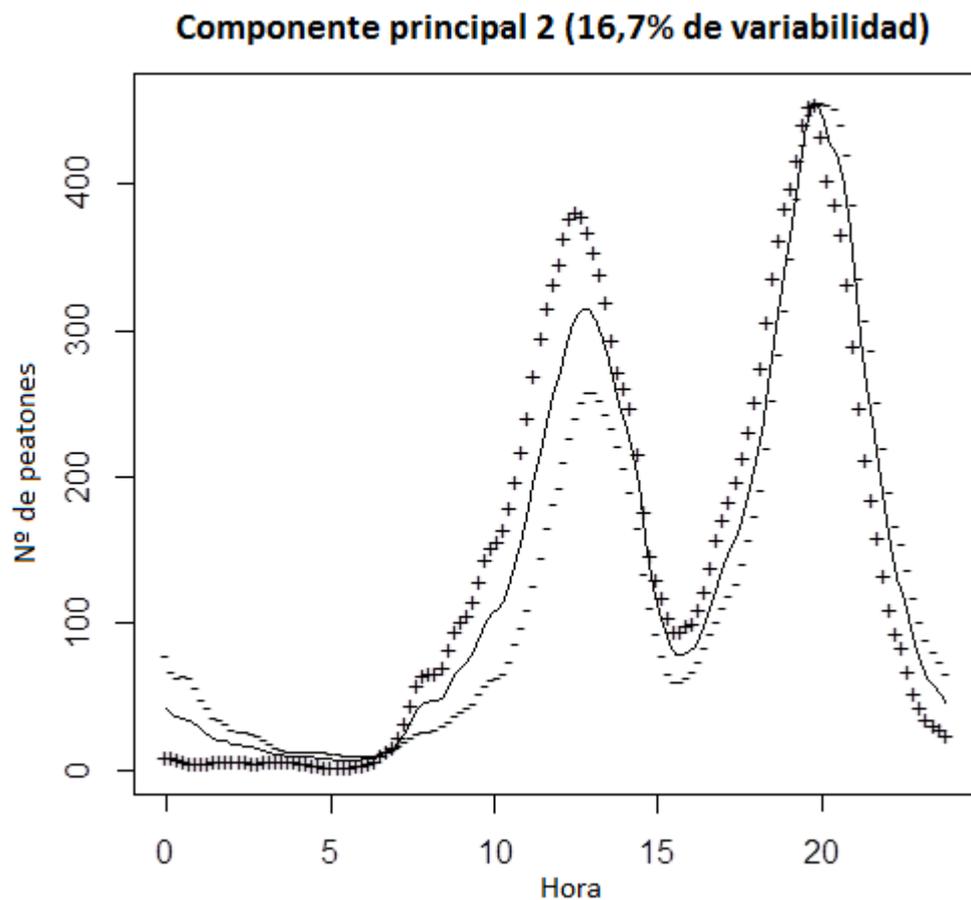


Figura 6.21. Efectos sobre los conteos totales de la primera componente principal

La segunda componente explica el 16.7 % de la variabilidad. Un efecto positivo aumenta el pico matinal y reduce la actividad nocturna. Un efecto negativo, por el contrario, provoca que el pico de la mañana se realice más tarde, y con menor afluencia de peatones. Sin embargo, el efecto negativo aumenta la actividad nocturna. Las 2 primeras componentes principales explican un 80.2 % de la variabilidad. En la figura 6.22, se puede observar el efecto de la segunda componente principal. Posteriormente, se comprobó si la utilización de las componentes principales rotadas facilitaba la interpretación. Como su interpretación era bastante confusa se decidió no emplear dichas componentes y quedarnos con las componentes sin rotar.



*Figura 6.22. Efectos sobre los conteos totales de la segunda componente principal*

En la figura 6.23 se grafica el valor de la puntuación en las 2 primeras componentes de todas las observaciones coloreadas según el tipo de día. Se

puede apreciar que las observaciones se agrupan según el tipo de día. Se pueden explicar las causas de los valores de cada grupo:

- **Laborable no escolar:** se trata de días de vacaciones de Navidad o un día de fiesta que se traslada por caer domingo a un laborable. Tienen una alta puntuación positiva en la componente 1, es decir, grandes picos matinales y vespertinos, y alta puntuación en la componente 2, lo que implica que hay prácticamente nula actividad nocturna y gran afluencia por la mañana. Estas puntuaciones se pueden explicar por las compras navideñas tanto por la mañana y por la tarde, y por las bajas temperaturas por la noche en esos días.
- **Laborable:** son días con la componente 1 en torno a cero con más observaciones negativas y una componente 2 principalmente positiva. Se trata de días con poca actividad de noche.
- **Sábados:** tienen la componente 2 negativa lo que quiere decir es que tienen alto flujo de peatones de madrugada y bajo flujo por la mañana, y tiene la componente 1 ligeramente positiva que implicaría mayor flujo matinal y/o vespertino. Como el flujo matinal es bajo por tener la componente 2 negativa, los sábados tienen una gran afluencia de peatones por la tarde y un bajo flujo por la mañana.
- **Domingo-festivo:** tienen la componente 2 negativa como los sábados por la gran cantidad de personas que pasa de madrugada y la poca gente que pasa a primeras horas de la mañana. La componente 1 se encuentra en torno a 0 tendiendo a ser negativa: en este caso el flujo vespertino no es tan grande para compensar el efecto negativo de la baja afluencia matinal como ocurría en muchos sábados.

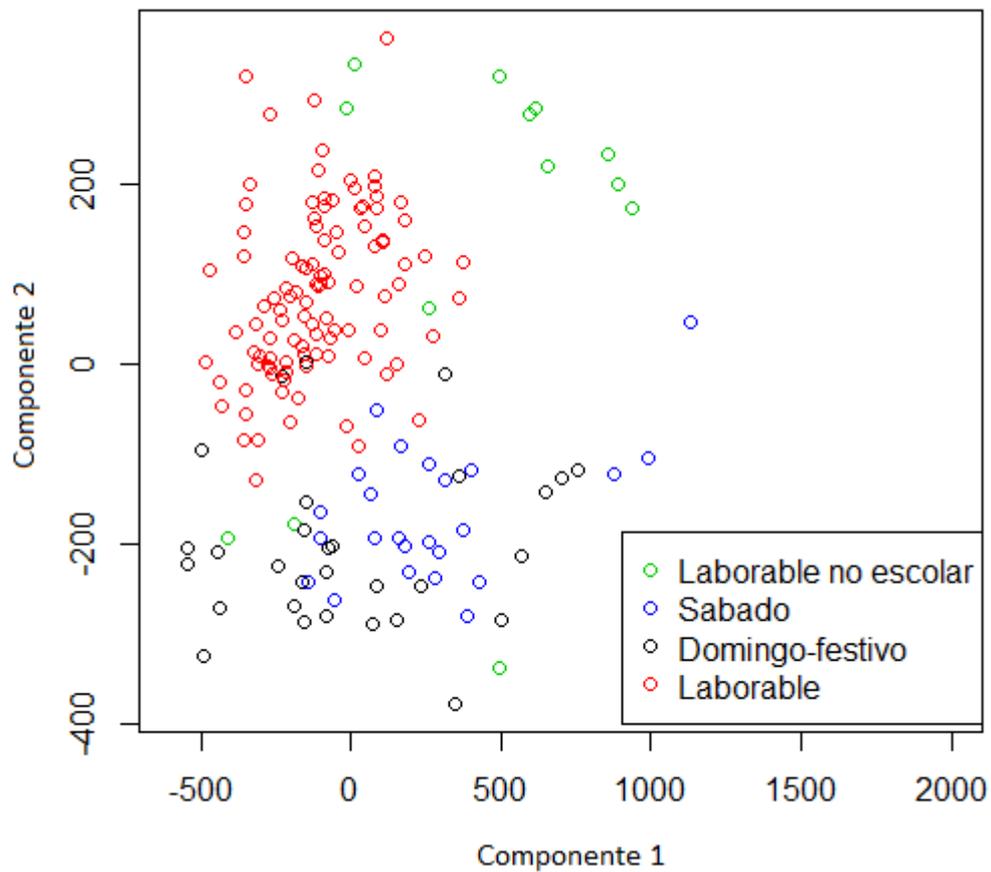


Figura 6.23. Puntuaciones de las observaciones de conteos totales en las 2 componentes principales

### 6.3.2 ACP de la espera total

El 77.2 % de la variabilidad de la espera total es explicada por la componente principal 1. En la figura 6.24 se puede observar el efecto de los valores de esta 1ª componente principal. La primera componente tenderá a ser mayor cuanto mayores sean los 2 picos por la mañana y por la tarde. Por el contrario, será menor cuanto menores sean los picos de gente esperando por la mañana y por la tarde.

### Componente principal 1 (77,2 % de variabilidad)

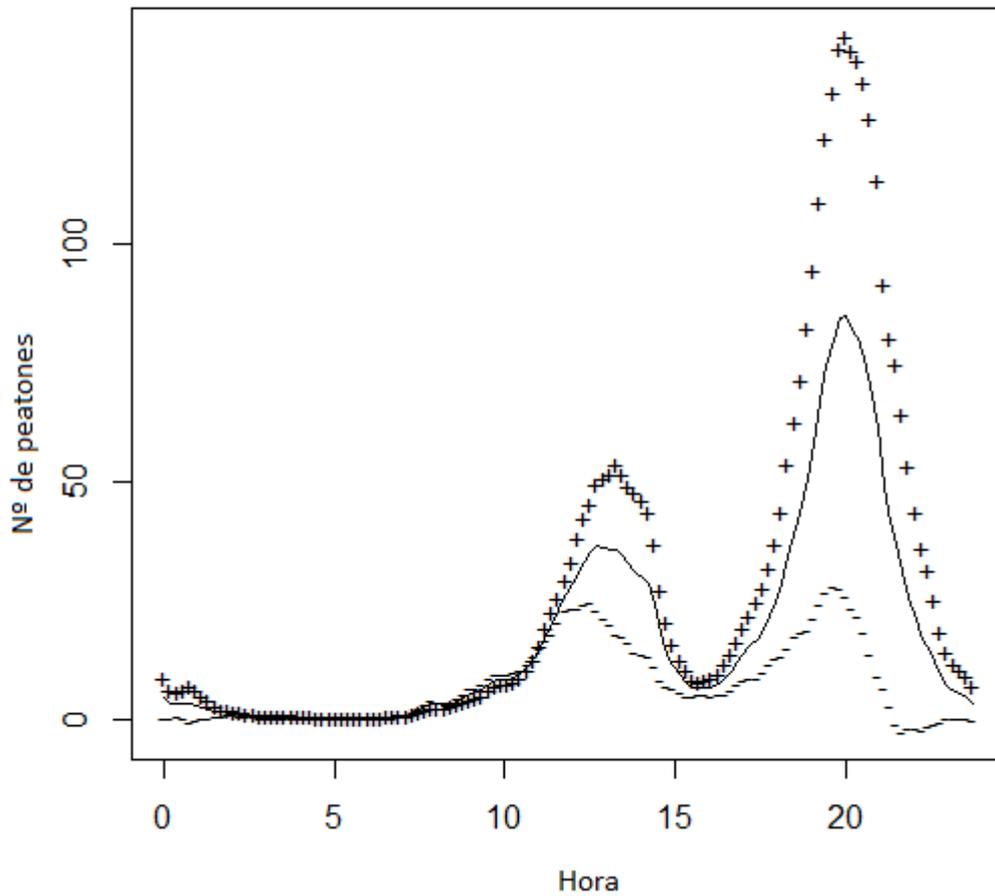


Figura 6.24. Efecto de la primera componente principal de la espera total

La segunda componente principal de la espera total explica un 6.6 % de la variabilidad. Entonces, las 2 primeras componentes principales describen el 83.8 % de la variabilidad. La figura 6.25 muestra cómo cambia el valor de la 2ª componente en función del número de personas que espera. Se tendrá un valor mayor en esta componente principal especialmente cuanto mayor sea el pico matinal. Aumentará ligeramente el valor de la 2ª componente principal si se adelanta el pico de la tarde o si la espera nocturna y de madrugada es nula. Como con los conteos totales, la utilización de componentes rotadas no mejora la interpretación, y por tanto se decide que no se incluyan.

### Componente principal 2 (6,6 % variabilidad)

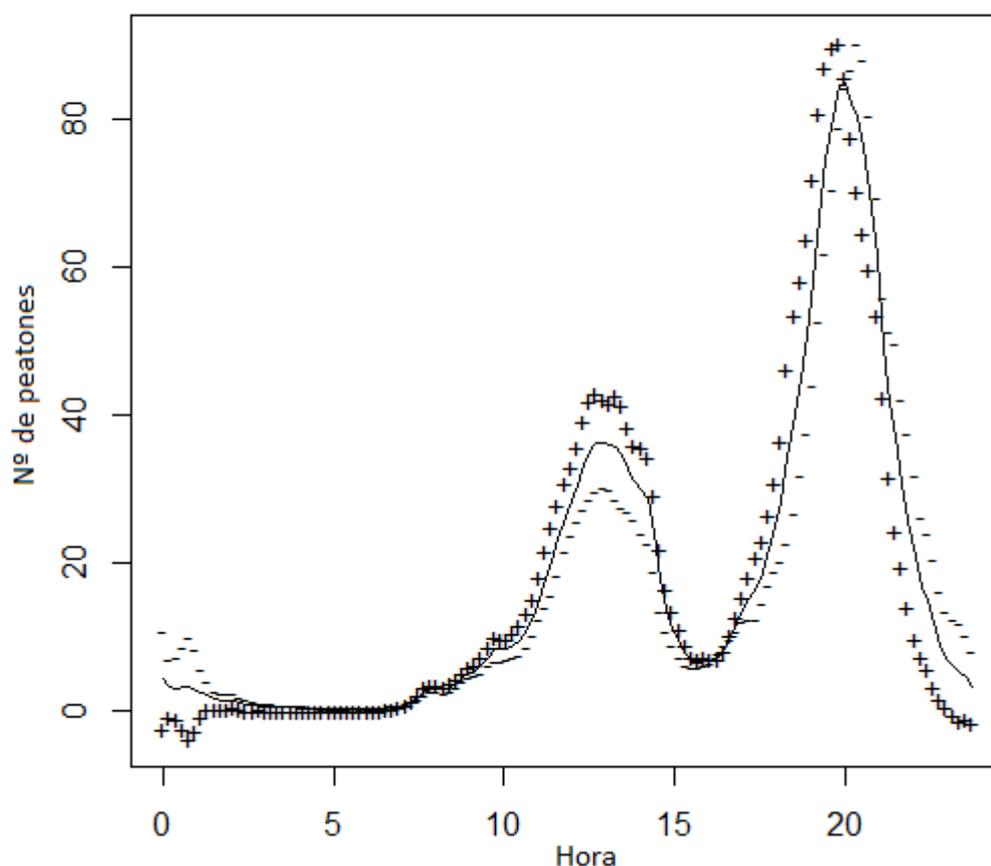


Figura 6.25. Efectos sobre las esperas totales de la segunda componente principal

En la figura 6.26, se incluye, al igual que en los conteos totales, las puntuaciones de las observaciones en las 2 primeras componentes coloreadas según el tipo de día. En este caso los grupos no se encuentran tan definidos como con los conteos totales pero sí que se observan claras tendencias en algunos grupos:

- **Laborables no escolares:** los días en este grupo están muy dispersos pero tienen una puntuación positiva en la 1ª componente principal, excepto 2 días: Nochebuena y Nochevieja. Estos días se caracterizan por las compras y es normal el mayor pico de espera por la mañana y por la tarde. Las excepciones están justificadas ya que esos días el flujo por la tarde se ve muy reducido por las típicas cenas.

- **Sábados:** tienen una puntuación negativa en la 2ª componente principal y la mayor parte de los días tienen puntuación positiva en la 1ª componente principal. Entonces, los sábados tienen un pico mayor de peatones esperando más tarde, más gente esperando por la noche y de madrugada, así como un pico matinal menor.
- **Domingos-festivos:** las observaciones se encuentran muy dispersas entre sí y es complejo sacar conclusiones.
- **Laborables:** las observaciones se encuentran bastante próximas entre sí. La puntuación en las 2 componentes principales se encuentran en torno a cero, habiendo ligeramente más observaciones con puntuación negativa en la componente principal 1.

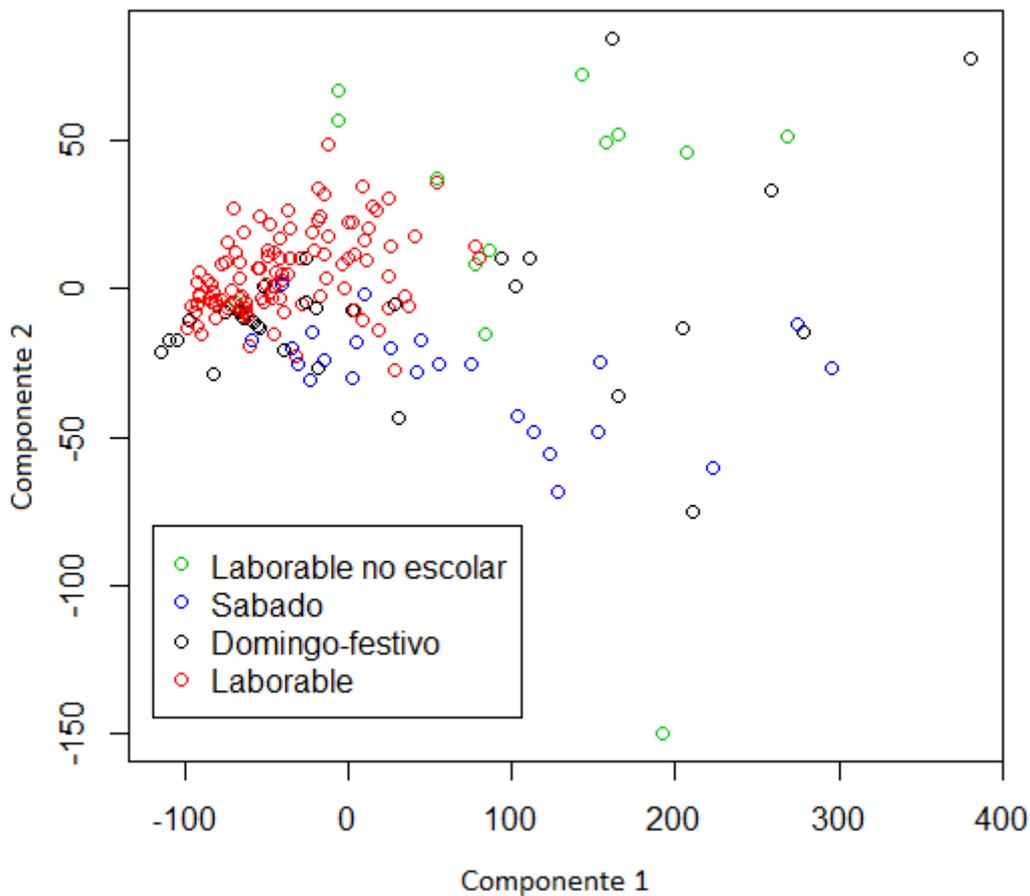


Figura 9.6. Puntuaciones de las observaciones de las esperas totales en las 2 componentes principales agrupadas por tipo de día.

## 7. Análisis de los tiempos de espera

Para comprobar la efectividad del semáforo inteligente en la reducción de los tiempos de espera se realiza el test chi-cuadrado de homogeneidad para comprobar si existen diferencias significativas en la función de distribución de los tiempos de espera dependiendo de si el semáforo inteligente se encuentra activado o desactivado. Se decide no hacer el análisis para la franja horaria 14-19:30 debido a que es un momento de escaso flujo peatonal como vimos en el análisis de flujos de la espera total, y por tanto, es muy probable que no se cumpla en ningún momento la condición de tener 17 o más peatones esperando a cruzar en el paso para que el semáforo inteligente aumente el tiempo paso de los peatones. Se realiza entonces:

- En primer lugar se compara el lunes 28 de abril de 2014 (semáforo inteligente activado) con el lunes 5 de mayo de 2014 (semáforo inteligente desactivado) en las franjas horarias entre las 19:30 y las 24:00.
- En segundo lugar se compara los tiempos de espera del martes 29 de abril de 2014 (semáforo inteligente activado) con el martes 6 de mayo de 2014 (semáforo inteligente desactivado) entre las franjas horarias de las 0:00 y las 14:00.

### Caso 1

En este caso las medias de los tiempos de espera son:

- 9.11236 segundos para el 28 de abril de 19:30 a 14 h.
- 9.690972 segundos para el 5 de mayo de 19:30 a 24h.

Se observa, como cabe esperar, que el tiempo de espera medio es inferior en el período con el semáforo inteligente activado. A continuación, se procederá a comprobar si esa diferencia es significativa a través del test de homogeneidad chi-cuadrado para 2 poblaciones.

En la tabla 7.1 se encuentra el número de observaciones de las 2 poblaciones en las 8 distintas clases que se han considerado. La elección de las clases se ha realizado procurando que las clases tengan un número de observaciones parecido.

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3-4</b>	<b>5-7</b>	<b>8-12</b>	<b>13-23</b>	<b>24-80</b>	
<b>28 abril</b>	72	60	39	37	26	38	38	46	356
<b>5 mayo</b>	70	33	18	33	41	25	33	35	288
	142	93	57	70	67	63	71	81	

*Tabla 7.1. Distribución de los tiempos de espera entre los 8 grupos el 28 de abril y 5 de mayo por la noche*

En la figura 7.1 se representa los tiempos de espera en un histograma. De color morado están las partes comunes del histograma. De color azul se encuentra las partes del histograma del 28 de abril con mayor frecuencia que el 5 de mayo. De color carne están los tiempos de espera del 5 de mayo cuya frecuencia es mayor que la del 28 de abril.

### Histogramas 28 abril y 5 mayo por la noche

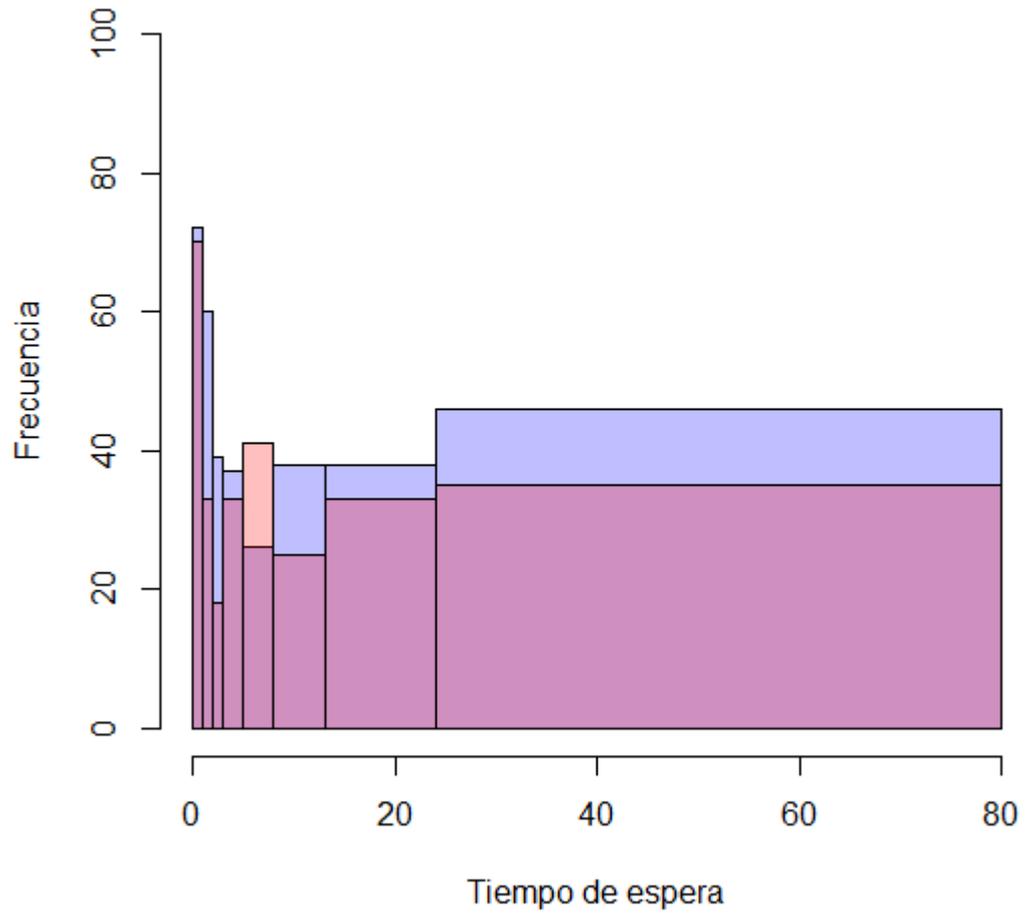


Figura 7.1. Histograma de las frecuencias relativas de los tiempos de espera del 28/4 y 5/5 por la noche.

Se realiza el test chi-cuadrado según lo visto en el punto 4.5.

Las hipótesis en este caso son:

$H_0: p_{\text{espera en 0 28 abril noche}} = p_{\text{espera en 0 5 mayo noche}}$

$p_{\text{espera en 1 28 abril noche}} = p_{\text{espera en 15 mayo noche}}$

$p_{\text{espera en 2 28 abril noche}} = p_{\text{espera en 2 5 mayo noche}}$

$p_{\text{espera en 3-4 28 abril noche}} = p_{\text{espera en 3-4 5 mayo noche}}$

$p_{\text{espera en 5-7 28 abril noche}} = p_{\text{espera en 5-7 5 mayo noche}}$

$p_{\text{espera en 8-12 28 abril noche}} = p_{\text{espera en 8-12 5 mayo noche}}$

$p_{\text{espera en 13-23 28 abril noche}} = p_{\text{espera en 13-23 5 mayo noche}}$

$p_{\text{espera en 24-80 28 abril noche}} = p_{\text{espera en 24-80 5 mayo noche}}$

$H_1$ : Al menos un par de los anteriores no es cierto

Se calcula el valor del estadístico chi-cuadrado y resulta ser 16.725. El número de grados de libertad es  $(8-1)(2-1)=7$ . La distribución chi-cuadrado para 7 grados de libertad nos indica que para un valor del estadístico chi-cuadrado el p-valor es 0.01926 como se puede ver en la figura 7.2. Entonces, se rechaza la hipótesis nula: existen diferencias significativas, al menos, entre uno de las clases de los tiempos de espera.

### Función de densidad chi-cuadrado para 7 g.l

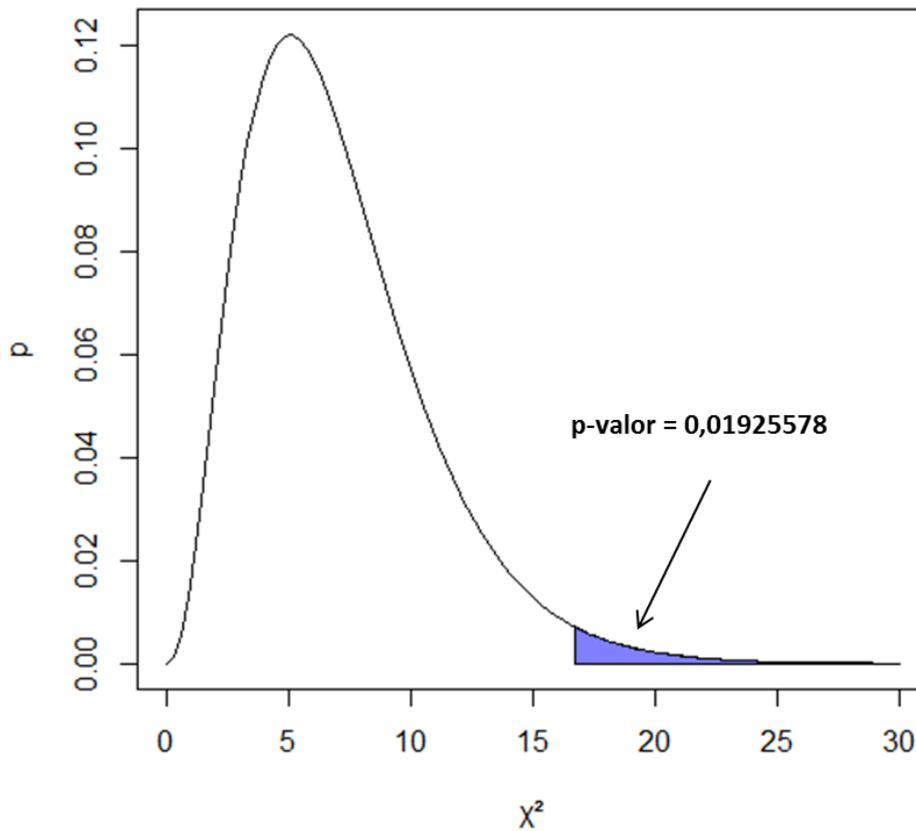


Figura 7.2. Función de densidad chi-cuadrado para el caso 1

### Caso 2

En este caso las medias de los tiempos de espera son:

- 8.711864 segundos para el 29 de abril de 0:00 a 14 h.
- 9.042403 segundos para el 6 de mayo de 0:00 a 14 h.

Como en el caso anterior el tiempo de espera medio con el semáforo inteligente activo es inferior que con el semáforo inteligente desactivado. En la tabla 10.2 se recoge el número de observaciones que hay en los 8 grupos durante los 2 martes por la mañana.

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3-4</b>	<b>5-7</b>	<b>8-12</b>	<b>13-23</b>	<b>24-80</b>	
<b>29 abril</b>	42	55	27	25	34	31	52	29	295
<b>6 mayo</b>	42	34	27	32	39	42	34	33	283
	84	89	54	57	73	73	86	62	578

*Tabla 7.2 Distribución de los tiempos de espera entre los 8 grupos el 29 de abril y 6 de mayo por la mañana.*

En la figura 7.3, se muestra los histogramas de los tiempos de espera del 29 de abril y el 6 de mayo. De color morado están las partes comunes del histograma. De color azul se encuentra las partes del histograma del 29 de abril con mayor frecuencia que el 6 de mayo. De color carne están los tiempos de espera del 6 de mayo cuya frecuencia es mayor que la del 29 de abril.

## Histogramas 29 abril y 6 mayo por la mañana

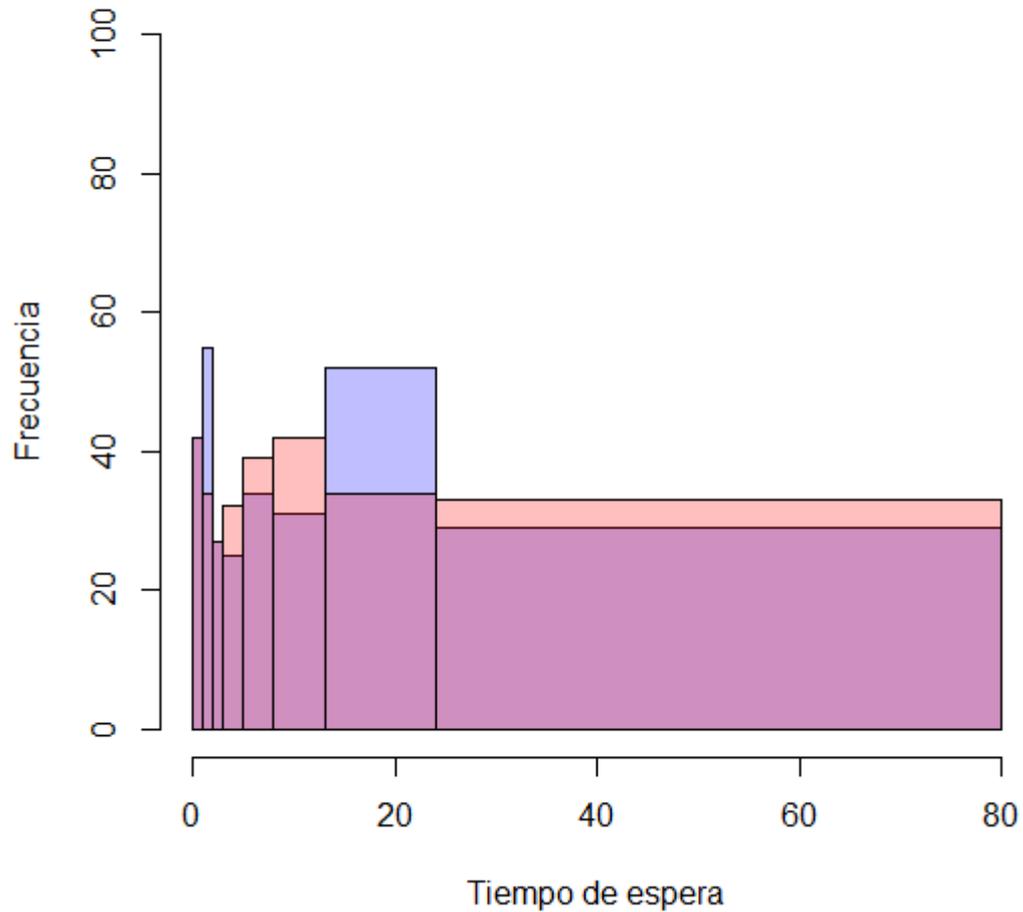


Figura 7.3 Histograma de los tiempos de espera el 29 de abril y el 6 de mayo por la mañana

Las hipótesis en este caso son:

$H_0: p_{\text{tespera en 0 29 abril mañana}} = p_{\text{tespera en 0 6 mayo mañana}}$

$p_{\text{tespera en 1 29 abril mañana}} = p_{\text{tespera en 16 mayo mañana}}$

$p_{\text{tespera en 2 29 abril mañana}} = p_{\text{tespera en 2 6 mayo mañana}}$

$p_{\text{tespera en 3-4 29 abril mañana}} = p_{\text{tespera en 3-4 6 mayo mañana}}$

$p_{\text{tespera en 5-7 29 abril mañana}} = p_{\text{tespera en 5-7 6 mayo mañana}}$

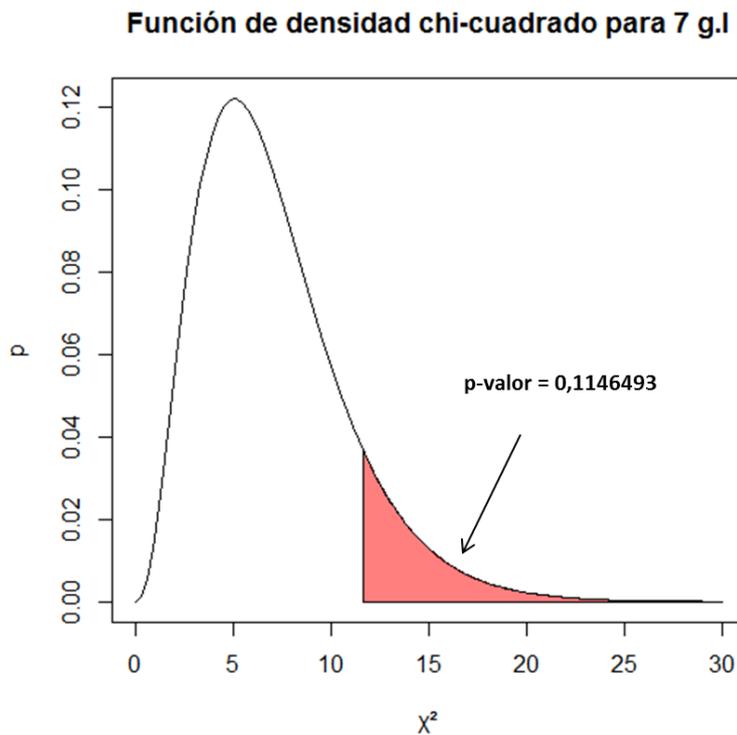
$p_{\text{tespera en 8-12 29 abril mañana}} = p_{\text{tespera en 8-12 6 mayo mañana}}$

$p_{\text{tespera en 13-23 29 abril mañana}} = p_{\text{tespera en 13-23 6 mayo mañana}}$

$p_{\text{tespera en 24-80 29 abril mañana}} = p_{\text{tespera en 24-80 6 mayo mañana}}$

$H_1$ : Al menos un par de los anteriores no es cierto

El valor del estadístico chi-cuadrado es 11.59607. Entonces para una distribución chi-cuadrado con 7 grados de libertad se corresponde con un p-valor de 0.1146493 como se puede apreciar en la figura 7.4. Por tanto, se acepta la hipótesis nula y en este caso, no existen diferencias significativas entre los distintos grupos de tiempos de espera considerados.



*Figura 7.4. Función de densidad chi-cuadrado para el caso 2.*

A partir de los análisis anteriores se puede afirmar que los tiempos de espera se reducen en ambos casos anteriores. Por la tarde, esa reducción es significativa modificándose el comportamiento de la función de distribución de los tiempos de espera. Por la mañana, sin embargo, no se puede concluir que se modifica la función de distribución aunque disminuya el tiempo de espera. Esto seguramente se deba a que el semáforo inteligente añadiera en muy pocas ocasiones segundos adicionales a la fase verde por no cumplirse la condición de que 17 peatones o más esperen a cruzar, ya que como vimos en los análisis anteriores el pico matinal de peatones es menor. Por ello, probablemente si se realizara el análisis entre 2 sábados en lugar de días de diario, la reducción de los tiempos de espera sería mayor y la modificación de la función de distribución de los tiempos de espera de los peatones sería también significativa por la mañana.

## 8. Conclusiones globales y líneas futuras

Como conclusiones se pueden señalar:

- El flujo de peatones en la calle Santiago presenta dos momentos de mayor afluencia cada día: el primero en torno a las 12 de la mañana y el segundo en torno a las 19.30 de la tarde, siendo este segundo pico el mayor normalmente.
- Del análisis de la varianza se ha apreciado que existen diversos factores como el día de la semana, el mes, el tipo de día, la presencia de precipitaciones que son muy significativos y tienen una gran influencia en el comportamiento de los flujos de peatonales totales y de los que se detienen a esperar en el semáforo.
- El análisis clúster nos ha permitido agrupar los días con otros días de similares características, siendo los grupos lo más heterogéneos entre sí. Se han detectado ciertos días que por diversos motivos (gran afluencia por compras navideñas, manifestaciones o concentraciones, fiestas locales) son muy diferentes al resto de días y no encajan en ninguno de los grupos. Se observa que los sábados y los días de diciembre son los mayor afluencia en la calle Santiago.
- El análisis de componentes principales nos ha permitido comprobar que todos los datos de los flujos peatonales se podían reducir a 2 dimensiones manteniéndose más del 80 % de variabilidad tanto en los conteos totales como la espera total. Además, se comprueba que al graficar las puntuaciones de las observaciones en esas dos dimensiones, las observaciones se encuentran bastante agrupadas según el tipo de día.
- En el análisis de los tiempos de espera con test chi-cuadrado de homogeneidad se ha comprobado que existen diferencias significativas de que el semáforo reduce los tiempos de espera en el período vespertino del día analizado. En el período matinal analizado, sin embargo, no se encontraron diferencias significativas entre los tiempos de espera. Esto último no quiere significar el semáforo no funcione correctamente, sino que por la mañana el efecto del semáforo inteligente es menor debido a que, como se vio en el análisis de la espera total de peatones, por la mañana hay menos peatones que esperan en el paso, y

por tanto en menos ocasiones se modificará el ciclo del semáforo al no cumplirse la condición de que haya al menos 17 peatones esperando.

Lo anteriormente señalado puede servir para valorar el diseño de ciclos de semáforo adaptados a las características que aquí han sido detectadas como significativas en el tránsito de peatones. Es decir que como, por ejemplo, los días festivos (o los lluviosos) se comportan de forma diferente a los que no lo son, parece conveniente el establecimiento de patrones de semáforo adaptados a estas características para mejorar los flujos de tránsito. Este diseño queda, obviamente, más allá de los objetivos de este trabajo

También sería recomendable emplear los resultados del estudio para modificar el ciclo de semáforo en las horas con nula o casi nula espera de los días de un clúster, aumentando el tiempo de ciclo en verde para vehículos para evitar esperas innecesarias de éstos. Como en el caso anterior, es un diseño que se encuentra fuera de los objetivos del trabajo.

El presente proyecto del análisis de los flujos de tránsito de la calle Santiago deja abiertas diversas vías a desarrollar en un futuro.

En primer lugar, el análisis de los flujos peatonales se encuentra incompleto ya que el estudio se ha realizado de medio año. Sería interesante prolongar el análisis al año completo, especialmente en momentos de gran flujo peatonal como son los días de Semana Santa y la semana de fiestas locales de Valladolid que no se recogió en este estudio. Es bastante posible que existan causas meteorológicas, sociales, vacacionales o de cualquier otra índole que modifiquen el comportamiento del flujo de peatones en los meses de primavera y verano, respecto a los estudiados de otoño e invierno.

En segundo lugar, si se mantiene la intención de instalar más semáforos inteligentes en otras intersecciones como Plaza de Zorrilla, Ferrari, plaza de España, Colón o Fuente Dorada como se comentó desde el ayuntamiento en junio de 2014 [Fuente:

<http://www.elnortedecastilla.es/valladolid/201406/02/valladolid-instalara-semaforos-inteligentes-20140602133304.html>. Fecha de consulta: Julio-2015], puede ser interesante realizar otro estudio como el realizado en este TFG y comparar los resultados con los de este estudio para analizar si los picos en el flujo de peatones se producen en otras horas o existe algún comportamiento diferente al de la calle Santiago.

En tercer lugar, el análisis de los tiempos de espera se ha realizado comparando 2 lunes y 2 martes. Sería muy recomendable realizar la comparación entre días de mayor flujo especialmente los sábados, en los que se intuye que probablemente el efecto del semáforo inteligente sea mayor (recordemos que el semáforo inteligente se activaba cuando había 17 o más personas esperando).

Por último, podría ser de gran interés la instalación de este sistema de conteos de peatones en otros lugares de Valladolid que tengan características diferentes a los de la calle Santiago (como zonas residenciales, calles con gran concentración de bares y/o restaurantes, recintos deportivos, colegios, centros comerciales) en los que el flujo peatonal sea importante y un análisis estadístico como el realizado en el presente TFG pueda ser interés.



## 9. Apéndices: Implementación en R

En este TFG se ha utilizado el software estadístico R para la realización de todos los análisis: análisis clúster funcional, análisis de la varianza funcional, análisis de componentes principales funcional y test chi-cuadrado de homogeneidad para 2 muestras.

R es un paquete estadístico que posee las siguientes características:

- Es libre, gratuito y multiplataforma (Windows, Mac, Linux).
- Dispone de gran cantidad de paquetes y librerías escritos por los propios usuarios que amplían las herramientas para realizar análisis estadísticos avanzados.
- Es uno de los programas más utilizados para realizar análisis estadísticos, especialmente entre la comunidad científica.

A continuación se adjuntan los scripts utilizados para realizar los diferentes análisis.

### 9.1 ANOVA funcional

El Análisis de la Varianza Funcional se ha utilizado con el paquete de R `fda.usc`:

```
#Carga de librerías
library(fda)
library(fda.usc)

#Lectura datos
ct_def<-read.table("ct_def_bis.txt", header=T, sep="\t", dec=".", na.strings="NA")
esp_def<-read.table("esp_def_bis.txt", header=T, sep="\t", dec=".", na.strings="NA")
hora<-(0:95)/4

#crear datos
total <- ct_def[,7:102]
espera<- esp_def[,7:102]
ct1<-fdata(total,hora,fdata2d=FALSE)
esp1<-fdata(espera,hora,fdata2d=FALSE)

#ANOVA días de la semana conteo total

ct1$class<- ct_def[,3]

a<-anova.onefactor(ct1,ct1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA meses de la semana conteo total

ct1$class<- ct_def[,4]
```

```

b<-anova.onefactor(ct1,ct1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA festivos conteo total

ct1$class<- ct_def[,6]
c<-anova.onefactor(ct1,ct1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA tipo día conteo total

ct1$class<- ct_def[,103]
d<-anova.onefactor(ct1,ct1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA ubicación del día en el mes conteo total

ct1$class<- ct_def[,104]
e<-anova.onefactor(ct1,ct1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA precipitación acumulada mayor que 1 mm conteo total

ct1$class<- ct_def[,105]
f<-anova.onefactor(ct1,ct1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA días de la semana espera total

esp1$class<- esp_def[,3]
g<-anova.onefactor(esp1,esp1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA meses de la semana espera total

esp1$class<- esp_def[,4]
h<-anova.onefactor(esp1,esp1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA festivos espera total

esp1$class<- esp_def[,6]
i<-anova.onefactor(esp1,esp1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA tipo día espera total

esp1$class<- esp_def[,103]
j<-anova.onefactor(esp1,esp1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA ubicación del día en el mes espera total

esp1$class<- esp_def[,104]
k<-anova.onefactor(esp1,esp1$class,10000,plot=TRUE,verbose=TRUE)

#ANOVA precipitación acumulada mayor que 1 mm conteo espera total

esp1$class<- esp_def[,105]
l<-anova.onefactor(esp1,esp1$class,10000,plot=TRUE,verbose=TRUE)

```

## 9.2. Análisis clúster funcional

Para el análisis clúster de los flujos de peatones se ha utilizado el script en R de A.Gordaliza y L.A García Escudero, que introdujeron en el artículo A Proposal for Robust Clustering.

El script resultante es:

*#Análisis clúster con el método de las k-medias recortadas de Gordaliza A. y García Escudero L.A.*

*#Carga de librerías*

```
library(splines)
```

*#Lectura datos*

```
ct_def<-read.table("ct_def.txt", header=T, sep="\t", dec=".", na.strings="NA")
```

```
esp_def<-read.table("esp_def.txt", header=T, sep="\t", dec=".", na.strings="NA")
```

```
hora<-(0:95)/4
```

*#Función para realizar el análisis clúster funcional*

```
trimm2 <- function(X,K,alfa){
```

```
  niter <- 2000
```

```
  csteps <- 10
```

```
  on <- rep(1,p)
```

```
  un <- (1:K)
```

```
  ll <- (1:K)
```

```
  dist <- (1:n)
```

```
  ind <- (1:n)
```

```
  recor <- floor(n*alfa)
```

```
  #initialize the objective function by a large enough value
```

```
  vopt <- 100000000
```

```
  #Ramdon restarts
```

```
  for (iter in 1:niter){
```

```
    cini <- X[sample(1:n,size=K,replace=F),]
```

```
    dim(cini) <- c(K,p)
```

```
    #C-steps
```

```
    for (t in 1:csteps){
```

```
      for (h in 1:n){
```

```
        for (k in 1:K){
```

```
          ll[k] <- sum((X[h,]-cini[k,])^2)
```

```
        }
```

```
        dist[h] <- min(ll)
```

```
        ind[h] <- which.min(ll)
```

```
      }
```

```
      distt <- sort(dist)[n-recor]
```

```
      qq <- (1:n)[dist<=distt]
```

```
      xmod <- as.matrix(cbind(X[qq,],ind[qq]))
```

```

#Calculus of the new k centers

for (k in 1:K){
  ni <- sum(xmod[,p+1]==k)
  if (ni>1){
    cini[k,]<-apply(xmod[xmod[,p+1]==k,1:p],2,mean)
  }
}

}

#Calculus of the trimmed k-variance

obj <- 0
for (l in 1:(n-recor)){
  obj <- obj+sum((xmod[l,1:p]-cini[xmod[l,p+1,]])^2 )
}
obj <- obj/(n-recor)

#Change the optimal value if a reduction in the objective function happens
if (obj <vopt){
  vopt <- obj
  print(vopt)
  copt <- cini
}

}

aa <- rep(1,K)
dista <- rep(1,n)
indi <- rep(1,n)
asig <- rep(1,n)
for (h in 1:n){
  for (k in 1:K){
    aa[k] <- sqrt(sum((X[h,]-copt[k,])*(X[h,]-copt[k,])))
  }
  dista[h] <- min(aa)
  indi[h] <- which.min(aa)
}

ord <-sort(dista)
ropt <- ord[n-recor]
for (h in 1:n){
  ifelse( dista[h]>ropt, asig[h] <-0, asig[h] <-indi[h])
}
v <- c(1:n)
for (k in 1:K){
  print(k)
  print(v[asig==k])
}
print(v[asig==0])

```

```

a <- list(vopt=vopt,copt=copt,ropt=ropt,asig=asig)

return(a)

}

#crear datos
Xori <- ct_def[,7:102]

n <- dim(Xori)[1]
pp <- dim(Xori)[2]

#representar base splines
nodos <- 8
X <- matrix(rep(0,n*(nodos+4)),nrow=n)
df <- nodos+4
xval <- (1:pp)/pp
b <- bs(xval, df = df,intercept=T)
for (u in 1:n){
fit <- lm(t(Xori[u,]) ~ -1+b)
X[u,] <- fit$coefficients
}

p <- dim(X)[2]
K <- 5
alfa <- 0.1
a <- trimm2(X,K,alfa)

#Gráficas de los clústers
par(mfrow=c(2,floor(K/2)+2))
plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab=paste(alfa," trimming size"),ylab="y")
for (i in 1:n){
lines(hora,Xori[i,],col=a$asig[i]+1,pch=2)
}
plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab="Cluster centers trajectories",ylab="y")

for (i in 1:K){
lines(hora,b[,i]*%a$copt[i],col=i+1,pch=2)
}

for (k in 1:K){

plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab=paste("Cluster ",k),ylab="y")

for (i in 1:n){
if (a$asig[i]==k) lines(hora,b[,i]*%X[i,],col=k+1,pch=2)
}

}

plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab="Trimmed curves",ylab="y")
for (i in 1:n){

```

```

        if (a$asig[i]==0) lines(hora,Xori[i,],col=1,pch=2)

    }

#crear datos
Xori <- esp_def[,7:102]

n <- dim(Xori)[1]
pp <- dim(Xori)[2]

#representar base splines
nodos <- 8
X <- matrix(rep(0,n*(nodos+4)),nrow=n)
df <- nodos+4
xval <- (1:pp)/pp
b <- bs(xval, df = df,intercept=T)
for (u in 1:n){
  fit <- lm(t(Xori[u,]) ~ -1+b)
  X[u,] <- fit$coefficients
}

p <- dim(X)[2]
K <- 5
alfa <- 0.1
a <- trimm2(X,K,alfa)

dev.new()
par(mfrow=c(2,floor(K/2)+2))
plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab=paste(alfa," trimming size"),ylab="y")
for (i in 1:n){
  lines(hora,Xori[i,],col=a$asig[i]+1,pch=2)
}
plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab="Cluster centers trajectories",ylab="y")

for (i in 1:K){
  lines(hora,b[,i]*a$scpt[i,],col=i+1,pch=2)
}

for (k in 1:K){

plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab=paste("Cluster ",k),ylab="y")

  for (i in 1:n){
    if (a$asig[i]==k) lines(hora,b[,i]*X[i,],col=k+1,pch=2)
  }

}

}
plot(hora,c(range(Xori)[2],rep(range(Xori)[1],pp-1)),type='n',xlab="Trimmed curves",ylab="y")

```

```

for (i in 1:n){
    if (a$asig[i]==0) lines(hora,Xori[i,],col=1,pch=2)
}

```

### 9.3 ACP funcional

Para el Análisis de Componentes Principales funcional se ha utilizado el paquete `fda`. El comando que permite la realización del análisis de componentes principales funcional dentro del paquete `fda` es `pca.fd`.

```

library(fda)

#Lectura datos
ct_def<-read.table("ct_def_bis.txt", header=T, sep="\t", dec=".", na.strings="NA")
esp_def<-read.table("esp_def_bis.txt", header=T, sep="\t", dec=".", na.strings="NA")
hora<-(0:95)/4

#Conteos totales
#crear datos
Xori <- ct_def[,7:102]
aaa<-matrix(rep(0,171*96),nrow=171)
for (e in 1:171){
    for (j in 1:96){
        aaa[e,j]<- ct_def[e,6+j]
    }
}

#Creacion del b-spline cúbico
mybasis<-create.bspline.basis(c(0,23.75),nbasis=96,norder=4)
cfd<-fd(t(aaa),mybasis,fdnames=(ct_def[,3]))

#Realización del análisis de componentes principales
a<-pca.fd(cfd, nharm = 2, harmfdPar=fdPar(cfd), centerfns= TRUE)
plot.pca.fd(a)

#Puntuaciones de las observaciones en las 2 componentes principales
plot(a$scores, col=ct_def[,103])
legend(0,-50,unique(ct_def[,103]),col=ct_def[,103],pch=1)

#Componentes principales rotadas
rot_cfd<-varmx.pca.fd(a)
plot.pca.fd(rot_cfd)

#Espera total
#crear datos

```

```

Xori <- esp_def[,7:102]
aaa<-matrix(rep(0,171*96),nrow=171)
for (e in 1:171){
  for (j in 1:96){
    aaa[e,j]<- esp_def[e,6+j]
  }
}

#Creación del b-spline cúbico
mybasis<-create.bspline.basis(c(0,23.75),nbasis=96,norder=4)
cfd<-fd(t(aaa),mybasis)

#Realización del análisis de componentes principales
a<-pca.fd(cfd, nharm = 2, harmfdPar=fdPar(cfd), centerfns= TRUE)
plot.pca.fd(a)

#Gráfica de la puntuación de las observaciones en las 2 componentes principales
plot(a$scores, col=ct_def[,103])
legend(-120,-80,unique(ct_def[,103]),col=ct_def[,103],pch=1)

#Componentes rotadas
a2<-varmx.pca.fd(a)
plot.pca.fd(rot_cfd)

```

#### 9.4. Test chi-cuadrado de homogeneidad para 2 muestras

El test chi-cuadrado está implementado en R mediante el comando `chisqtest()`, no es necesario recurrir a ningún paquete adicional.

El script utilizado para el análisis es:

```

#test chi cuadrado de homogeneidad

#Carga de los tiempos de espera distribuidos en 8 grupos

L5_3<-read.table("C:/Users/f/Desktop/tespera2/grupos/L5_19.5-24.txt", header=F, sep="\n", dec=".",
na.strings="NA")

L28_3<-read.table("C:/Users/f/Desktop/tespera2/grupos/L28_19.5-24.txt", header=F, sep="\n", dec=".",
na.strings="NA")

M29_1<-read.table("C:/Users/f/Desktop/tespera2/grupos/M29_0-14.txt", header=F, sep="\n", dec=".",
na.strings="NA")

M6_1<-read.table("C:/Users/f/Desktop/tespera2/grupos/M6_0-14.txt", header=F, sep="\n", dec=".",
na.strings="NA")

#Lunes por la noche (28-4 y 5-5)

#Obtener frecuencias absolutas de los 2 días

```

```

a7<-table(L28_3)

b7<-table(L5_3)

#Juntar las frecuencias absolutas en una matriz

c7<- matrix(nrow=2,ncol=8)

for(i in 1:8){

    c7[1,i]<- a7[i]

    c7[2,i]<-b7[i]

}

#Realización del test chi-cuadrado de homogeneidad

chisq.test(c7)

#Martes por la mañana (29-4 y 6-5)

#Obtener frecuencias absolutas de los 2 días

a3<-table(M29_1)

b3<-table(M6_1)

#Juntar las frecuencias absolutas en una matriz

c3<- matrix(nrow=2,ncol=8)

for(i in 1:8){

    c3[1,i]<- a3[i]

    c3[2,i]<-b3[i]

}

#Realización del test chi-cuadrado de homogeneidad

chisq.test(c3)

```



## 10. Bibliografía

### Referencias

- Cuesta-Albertos J.A, Gordaliza A., Matran C. Trimmed k-means: An Attempt to Robustify Quantizers. *The Annals of Statistics*, 25, 553-576. (1997)
- Cuevas, M. Febrero, R. Fraiman. An anova test for functional data. *Computational Statistics & Data Analysis*, 47, 111-122. (2004)
- Ferraty F., Vieu P. *Nonparametric Functional Data Analysis*. Springer (2006)
- Fundación CIDAUT – Comunicación Personal
- Gordaliza A., García Escudero L.A. A Proposal for Robust Curve Clustering. *Journal of Classification*, 22, 185-201. (2005)
- Hartigan J.A, *Clustering Algorithms*. New York: Wiley. (1975)
- Mcqueen J. Some Methods for Classification and Analysis of Multivariate Observations. 5th Berkeley Symposium on Mathematics, Statistics and Probability, Vol 1, 281-298. (1967)
- Montgomery D.C. *Análisis y diseño de experimentos*. Limusa Wiley (2004)
- Ramsay J.O, Silverman B.W. *Applied Functional Data Analysis: Methods and Case Studies*. Springer (2002)
- Ramsay J.O, Silverman B.W. *Functional Data Analysis*. Springer (2005)
- Ramsay J.O, Hooker G, Graves S. *Functional data analysis with R and MATLAB*. Springer (2009)
- Zhang J.T. *Analysis of Variance for Funcional Data*. CRC Press (2013)

## Referencias electrónicas

- <http://www.elnortedecastilla.es/valladolid/201406/02/valladolid-instalara-semaforos-inteligentes-20140602133304.html> Fecha de consulta: Julio-2015
- <http://www.eltiempo.es/valladolid.html?v=historico> Fecha de consulta: Julio-2015
- [http://www.seio.es/BEIO/files/BEIOVol24n1\\_ES\\_MFebrero.pdf](http://www.seio.es/BEIO/files/BEIOVol24n1_ES_MFebrero.pdf) Fecha de consulta: Julio-2015
- <http://www.psych.mcgill.ca/misc/fda/> Fecha de consulta: Julio-2015
- <http://www.uantof.cl/facultades/csbasicas/matematicas/academicos/jreyes/DOCENCIA/APUNTES/APUNTES%20PDF/Unidad%205%20Prueba%20chi-cuadrado.pdf> Fecha de consulta: Julio-2015
- <http://stattrek.com/chi-square-test/homogeneity.aspx?Tutorial=AP> Fecha de consulta: Julio-2015
- <http://wiener.math.csi.cuny.edu/Statistics/R/simpleR/stat013.html> Fecha de consulta: Julio-2015
- <http://faculty.bscc.cornell.edu/~hooker/ShortCourseHandout.pdf> Fecha de consulta: Julio-2015
- <http://www.europapress.es/castilla-y-leon/noticia-semaforo-inteligente-reduce-21-espera-20000-peatones-transitan-dia-calle-santiago-20140602133507.html> Fecha de consulta: Julio-2015
- [http://www.interempresas.net/Smart\\_Cities/Articulos/132517-Los-semaforos-inteligentes-reducen-un-21-por-ciento-el-tiempo-medio-de-espera-de-los.html](http://www.interempresas.net/Smart_Cities/Articulos/132517-Los-semaforos-inteligentes-reducen-un-21-por-ciento-el-tiempo-medio-de-espera-de-los.html) Fecha de consulta: Julio-2015
- <http://www.eldiadevalladolid.com/noticia/Z284B046C-E044-B191-91CC076FA2560470/20140303/valladolid/prueba/primer/semaforo/inteligente/calle/santiago> Fecha de consulta: Julio-2015