



# Título del trabajo

## Materia

Daniela Arbeláez Montoya  
Jefferson Gamboa Betancur  
Jean Paul Piedrahita García

Universidad Nacional de Colombia  
Ciencias, Escuela de Estadística  
Medellín, Colombia  
2021

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Análisis descriptivo</b>	<b>2</b>
2.1. Contorno de California . . . . .	4
<b>3. Procesos puntuales</b>	<b>6</b>
3.1. Análisis basado en densidad . . . . .	6
3.2. Densidad cuadrática . . . . .	6
<b>4. Análisis basado en la distancia</b>	<b>9</b>
4.1. Análisis de vecino más cercano promedio . . . . .	9
<b>5. Técnica Manual</b>	<b>10</b>
5.1. Gráfico de Patterson - Fry . . . . .	11
<b>6. Análisis de segundo orden para la incidencia de incendios en California en el 2017</b>	<b>11</b>
6.1. Función K . . . . .	12
6.2. Función F . . . . .	13
6.3. Función G . . . . .	14
6.4. Función de correlación por pares . . . . .	15
<b>7. Test para aleatoriedad completa</b>	<b>15</b>
7.1. Pointwise Envelopes . . . . .	15

# 1. Introducción

California es uno de los lugares que tiene las temporadas de incendios forestales más mortíferas y destructivas. El conjunto de datos contiene la lista de incendios forestales ocurridos en California entre 2013 y 2020; contiene la ubicación donde ocurrieron los incendios forestales, incluido el nombre del condado, los valores de latitud y longitud y también detalles sobre cuándo comenzó el incendio forestal.

Estos datos ayudan a generar información sobre qué lugares de California están bajo amenaza de incendio, a qué hora suelen ocurrir los incendios forestales y qué tan frecuentes y devastadores son.

La base de datos con la que se va realizar el análisis de patrones puntuales fue descargada de kaggle<sup>1</sup>.

Los paquetes a utilizar son los siguientes:

## 2. Análisis descriptivo

Para leer el archivo se utiliza la función “*read\_csv*” de la librería “*tidyverse*” y se visualiza la dimensión de la base de datos con la función “*dim*”.

Filas	Columnas
1636	40

Todos los incendios forestales registrados en la base de datos se encuentran inactivos o han sido contenidos. Debido a la gran cantidad de variables, seleccionamos algunas para visualizar un poco mejor las localizaciones de incendios, tales como:

- **AcresBurned:** representa el número de acres de tierra afectadas por los incendios forestales.
- **ArchiveYear:** año en el cual se desarrolló o se produjo el incendio forestal.
- **Name:** nombre o denominación asignada al incendio forestal.
- **Counties:** nombre del condado en el cual se produjo el incendio forestal.
- **Latitude y Longitude** corresponden a las coordenadas geográficas de cada uno de los incendios forestales

AcresBurned	ArchiveYear	Name	Counties	Latitude	Longitude
257314	2013	Rim Fire	Tuolumne	37.85700	-120.0860
30274	2013	Powerhouse Fire	Los Angeles	34.58559	-118.4232
27531	2013	Mountain Fire	Riverside	33.70950	-116.7288
27440	2013	American Fire	Placer	39.12000	-120.6500
24251	2013	Springs Fire	Ventura	0.00000	0.0000
22992	2013	Aspen Fire	Fresno	37.27900	-119.3180

Para saber el número de incendios que ocurrieron por año, visualizamos el número de registros con la función **table**. Se puede observar que el año 2017 fue el periodo donde más hubo incendios.

Año	Freq
2013	162
2014	117
2015	152
2016	185
2017	438
2018	316
2019	266

Se desea entonces observar y analizar el patrón de puntos de incendios forestales ocurridos durante el año 2017 en el estado de California. De esta manera realizamos el filtro correspondiente por año de ocurrencia y además, eliminamos aquellas observaciones con errores en sus coordenadas geográficas:

Sin embargo, la base de datos sigue registrando seis observaciones duplicadas de acuerdo a sus coordenadas geográficas. Eliminamos dichas observaciones de la siguiente manera:

```
> # ¿Hay datos duplicados?: 6 por valores de longitud y latitud
> sum(duplicated(firesCA2017 %>% select(Longitude, Latitude)))
```

```
## [1] 6
```

```
> firesCA2017 <- firesCA2017[!duplicated(firesCA2017 %>% select(Longitude, Latitude)),]
```

La función **project** del paquete **rgdal** ofrece una interfaz con la librería *PROJ.4* de funciones de proyección para datos de posición geográfica. Para poder utilizar esta función debemos extraer las coordenadas de cada uno de los incendios en una matriz. Realizamos la proyección correspondiente de acuerdo a la zona UTM en la cual se encuentra el estado de California y de acuerdo al elipsoide

---

<sup>1</sup><https://www.kaggle.com/ananthu017/california-wildfire-incidents-20132020>

**WGS84** (Sistema Geodésico Mundial).

```
> # Seleccionando las coordenadas  
> coordFiresCA2017 <- firesCA2017 %>% select(Longitude, Latitude) %>% as.matrix()  
> # Proyección de coordenadas a UTM  
> pointsFiresCA2017 <- project(coordFiresCA2017, "+proj=utm +zone=11N +ellps=WGS84") %>%  
+   as.data.frame() %>% rename(X = Longitude, Y = Latitude)
```

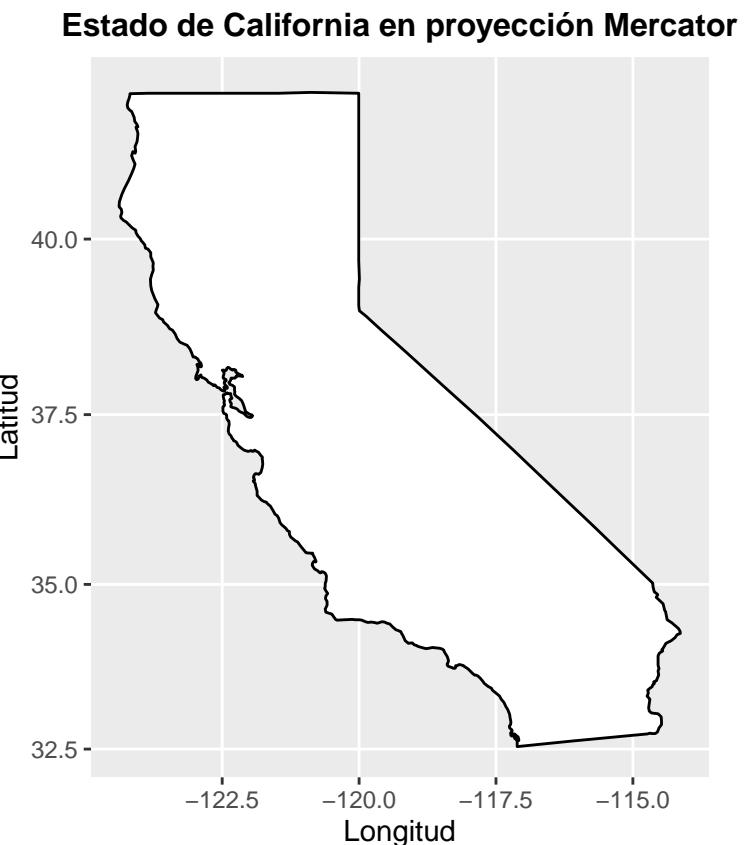
## 2.1. Contorno de California

El paquete **maps** contiene la base de datos **state** que produce un mapa de los estados del territorio continental de los Estados Unidos generado a partir de los datos del Departamento del Censo del mismo país; estos datos contiene las coordenadas geográficas de los polígonos correspondientes a cada uno de los estados.

```
## [1] "alabama"           "arizona"            "arkansas"  
## [4] "california"        "colorado"           "connecticut"  
## [7] "delaware"          "district of columbia" "florida"  
## [10] "georgia"           "idaho"              "illinois"  
## [13] "indiana"           "iowa"               "kansas"  
## [16] "kentucky"          "louisiana"         "maine"  
## [19] "maryland"          "massachusetts"     "michigan"  
## [22] "minnesota"         "mississippi"       "missouri"  
## [25] "montana"           "nebraska"          "nevada"  
## [28] "new hampshire"     "new jersey"        "new mexico"  
## [31] "new york"           "north carolina"    "north dakota"  
## [34] "ohio"               "oklahoma"          "oregon"  
## [37] "pennsylvania"      "rhode island"      "south carolina"  
## [40] "south dakota"       "tennessee"         "texas"  
## [43] "utah"               "vermont"            "virginia"  
## [46] "washington"        "west virginia"     "wisconsin"  
## [49] "wyoming"
```

Realizamos el filtro correspondiente al estado de California para obtener el polígono. La función **map\_data** del paquete **ggplot2** convierte fácilmente los datos del paquete de **maps** en un

marco de datos adecuado para trazar con **ggplot2**. La opción **Mercator** indica meridianos rectos igualmente espaciados, concordantes y rumbos rectos de la brújula.

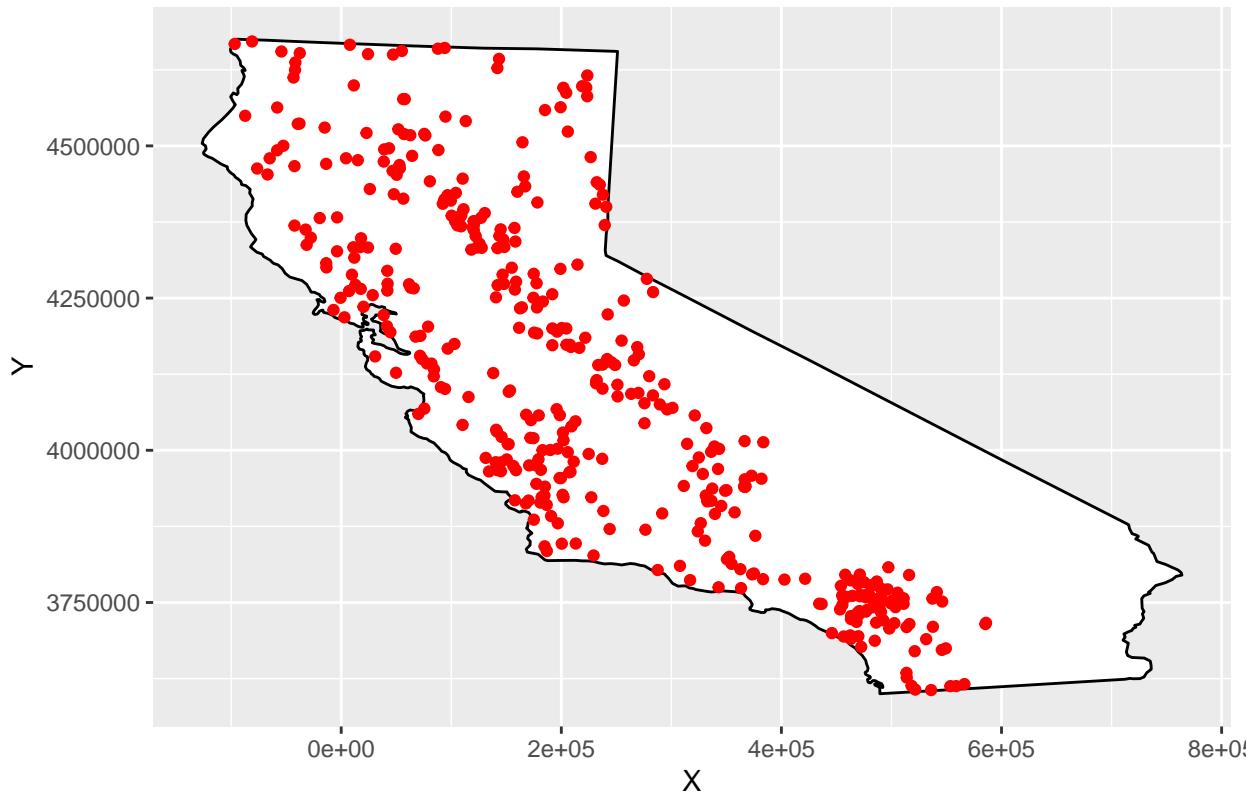


De la misma manera realizamos la proyección de las coordenadas geográficas en planas de acuerdo a la zona UTM correspondiente al estado de California.

Luego de la proyección, podemos crear el objeto **owin** correspondiente a la ventana de observación para el análisis del patrón de puntos espacial; sin embargo, es posible que algunos puntos o incendios forestales recaigan fuera de dicha ventana de observación, algo que no sería acorde o adecuado para el análisis. Para solucionar este inconveniente utilizamos la función **inside.owin** de la siguiente manera:

Visualizamos la proyección de los puntos en nuestra ventana de observación

## **Localización de Incendios en California, año 2017**



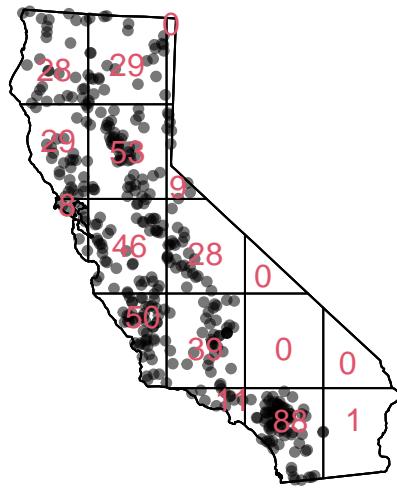
### **3. Procesos puntuales**

#### **3.1. Análisis basado en densidad**

#### **3.2. Densidad cuadrática**

El siguiente fragmento de código divide el estado de California en una cuadricula 9 filas, 6 columnas y luego se hace el conteo de las localizaciones que caen en cada cuadrante.

## Conteo por cuadrantes

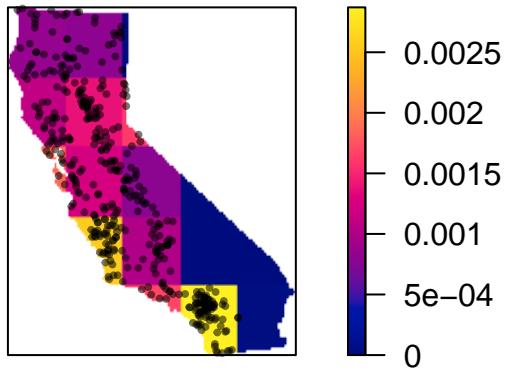


Se puede calcular la densidad de puntos dentro de cada cuadrante pero cada cuadrante esta en metros cuadrados, por ello para una visualización numérica se decidió pasar las localizaciones en  $km^2$  de la siguiente manera:

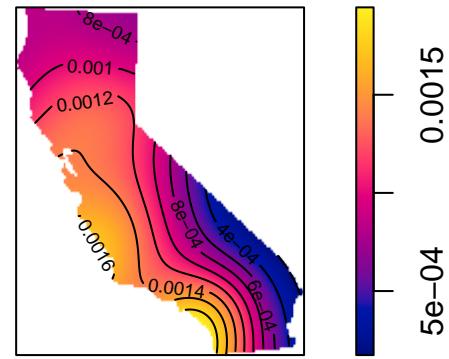
```
> # Reescalamiento a kilometros  
> FiresCA2017_ppp.km <- rescale(FiresCA2017_ppp, 1000, "km")  
>  
> # Conteo por cuadrantes en km  
> Q <- quadratcount(FiresCA2017_ppp.km)  
> # Intensidad  
> Q.d <- density(FiresCA2017_ppp.km)  
>  
> par(mfrow = c(1,2), no.readonly = T)  
> # plot de la densidad  
> plot(intensity(Q, image = T), main = "Densidad de incendios forestales \n en California",  
+       las = 1)  
> plot(FiresCA2017_ppp.km, pch = 20, cex = 0.6, col = rgb(0,0,0,.5), add = T, cols = "#00008B")  
> # Intensidad estimada
```

```
> plot(Q.d, main = "Intensidad Estimada")  
> contour(Q.d, add = T)
```

**Densidad de incendios forestales en California**

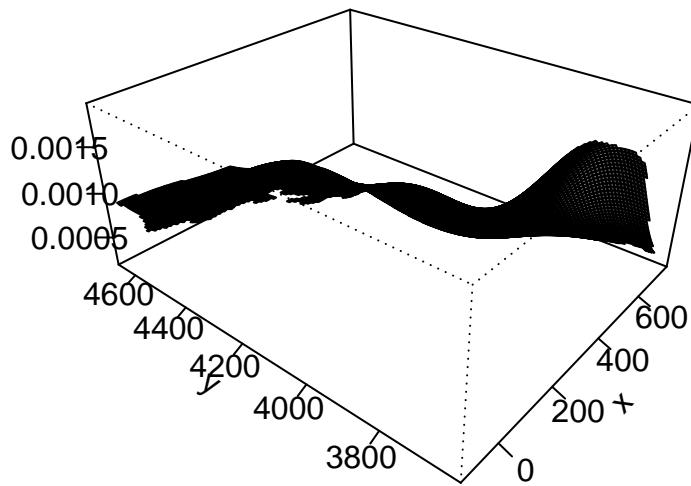


**Intensidad Estimada**



La intensidad espacial estimada para los incendios en California en el año 2017, proporciona una idea de lo que podemos esperar que pase en esa misma región en un futuro no muy lejano.

## **Superficie para la intensidad de incendios 2017**



## **4. Análisis basado en la distancia**

### **4.1. Análisis de vecino más cercano promedio**

A continuación, calcularemos las distancias promedio de vecino más cercano (ANN) entre la ocurrencia de incendios del año 2017 en California.

Para calcular la distancia promedio del primer vecino más cercano (en acres), se establece  $k = 1$ :

```
## [1] 10.59479
```

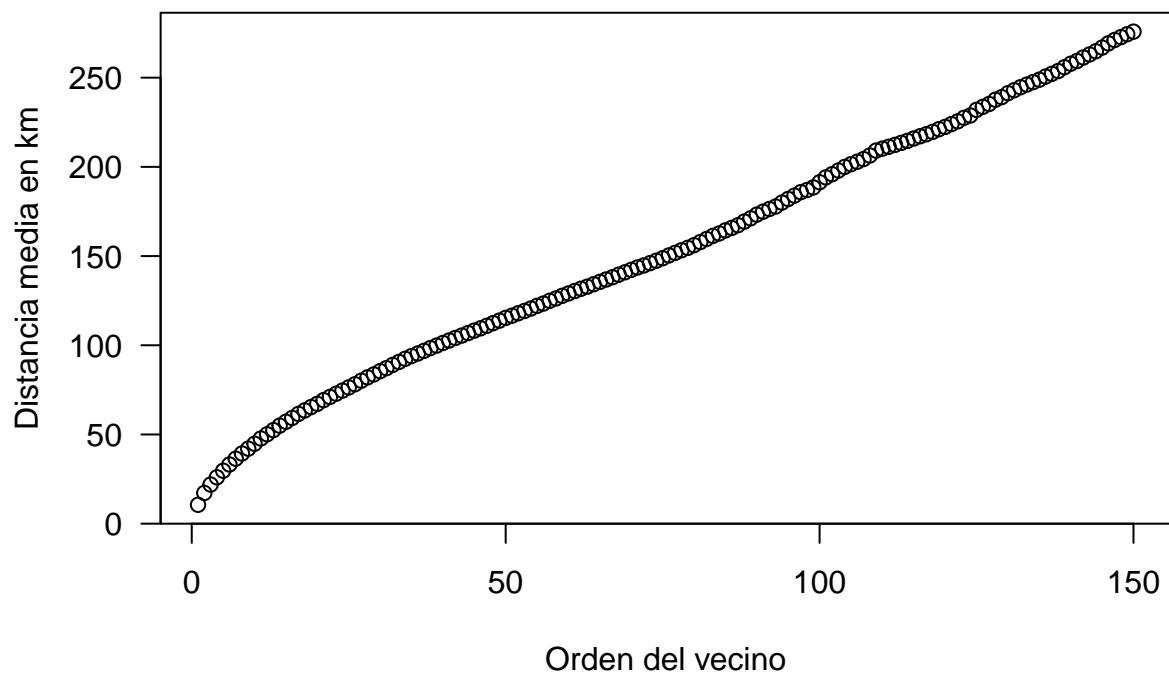
Para calcular el conjunto de distancia promedio del segundo vecino mas cercano, se usa:

```
## [1] 17.22054
```

El parámetro  $k$  puede tomar cualquier orden vecino (hasta  $n - 1$  donde  $n$  es el número total de puntos).

A continuación se traza ANN como una función del orden de vecinos para los primeros 150 vecinos mas cercanos.

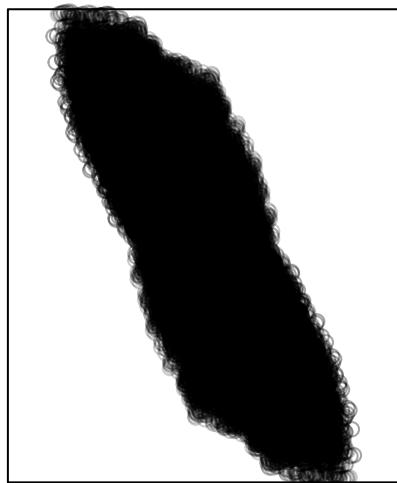
### **Vecinos mas cercanos**



## **5. Técnica Manual**

### **5.1. Gráfico de Patterson - Fry**

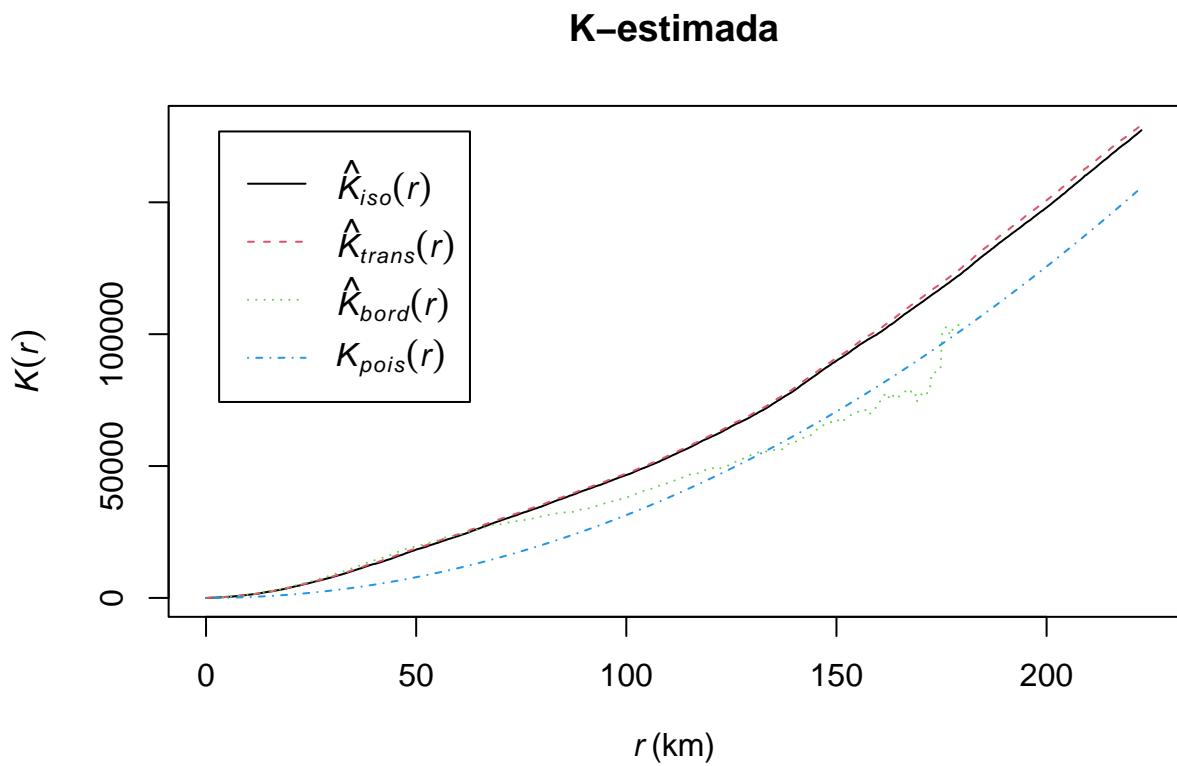
**Gráfico de Fry**



Al realizar el gráfico de Patterson Fry se observa que todos los puntos muestran una alta aglomeración de puntos alrededor del centro, por ello se presume un patrón de agregación.

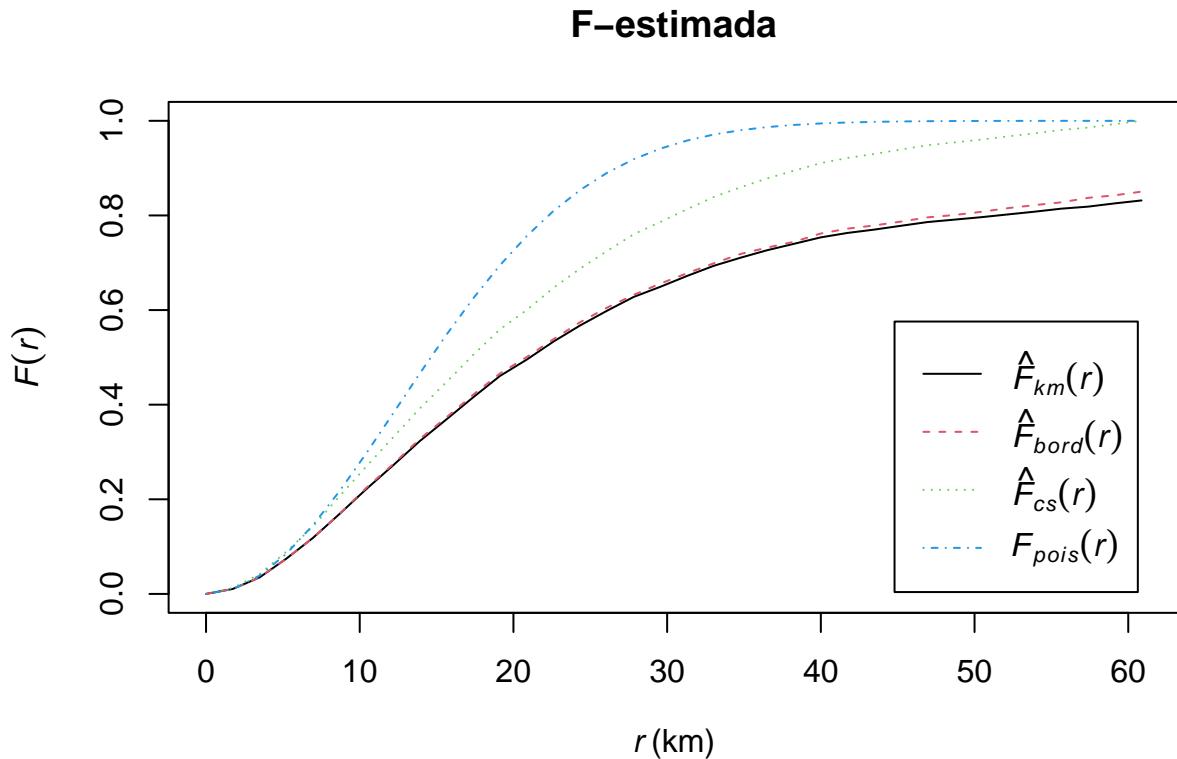
## **6. Análisis de segundo orden para la incidencia de incendios en California en el 2017**

## 6.1. Función K



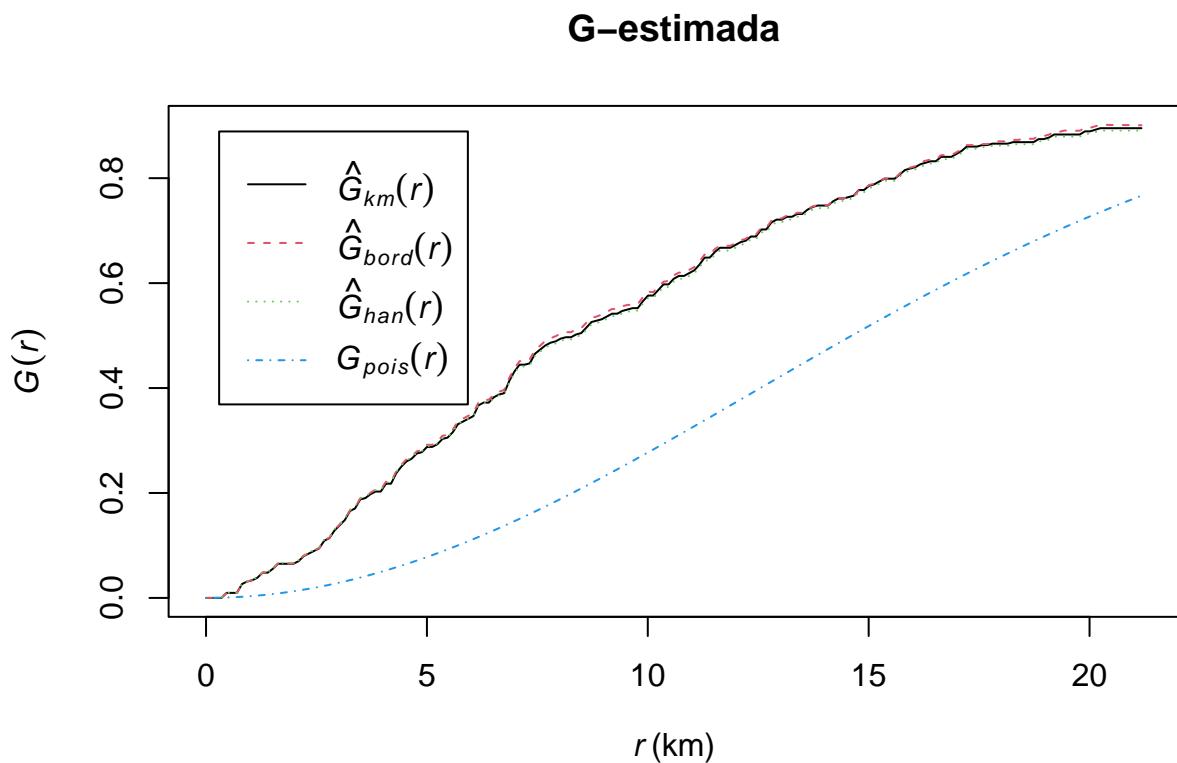
Debido a que las estimaciones estan por encima de la curva teórica, se puede pensar que se trata de un proceso de agregación.

## 6.2. Función F



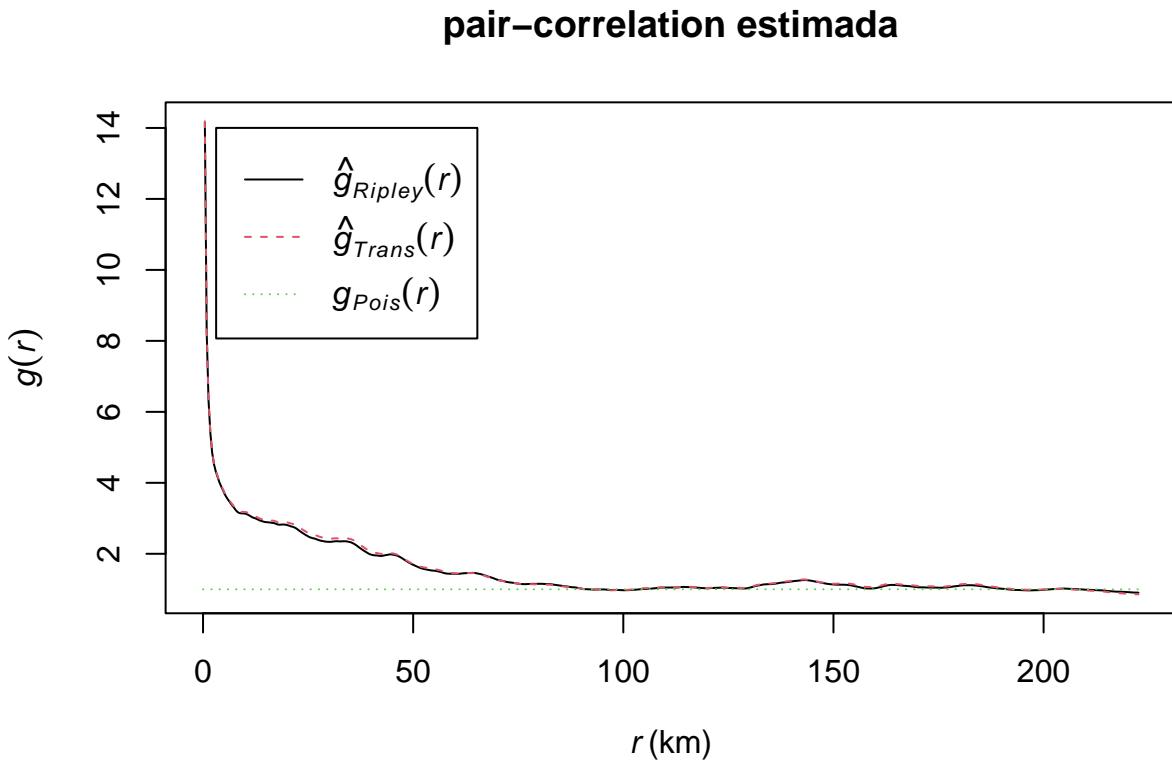
Cómo las curvas estimadas están por debajo de la curva teórica (Azul) se supone que es un patrón de agregación.

### 6.3. Función G



Como las funciones estimadas para la G están por encima de la teórica se presume que el patrón es de agración.

## 6.4. Función de correlación por pares



Debido a que las estimaciones se alejan por arriba de la teórica en distancias pequeñas se puede asumir que la hipótesis del patrón puntual es de agregación.

## 7. Test para aleatoriedad completa

### 7.1. Pointwise Envelopes

```
## Generating 39 simulations of CSR with fixed number of points ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
##
## Done.

## Generating 39 simulations of CSR with fixed number of points ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
##
## Done.

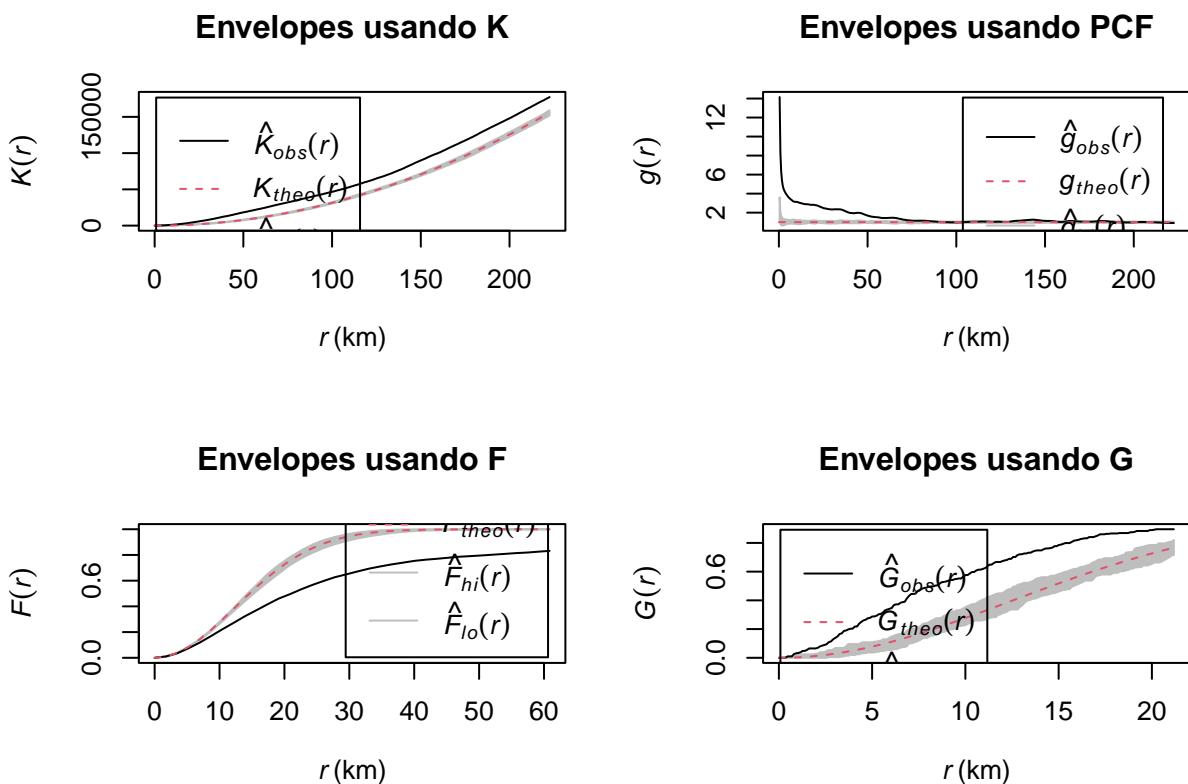
## Generating 39 simulations of CSR with fixed number of points ...
```

```

## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
##
## Done.

## Generating 39 simulations of CSR with fixed number of points ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
##
## Done.

```



Con el test para la aleatoriedad completa se puede concluir que el patrón se aleja por arriba de la curva teórica y además la sombra gris no contiene la curva negra que es la función K-estimada, por ello se asume que es un patrón de agregación. Para la función g-estimada a distancias cortas se sobre ajusta a la curva teórica y además las sombras grises no cubren a la función estimada a distancias pequeñas intuyendo un patrón de agregación.