

Taller # 2

Sebastán Gómez Duque Alejandro ... Felipe ... Daniel Felipe Cendales

Información del conjunto de datos

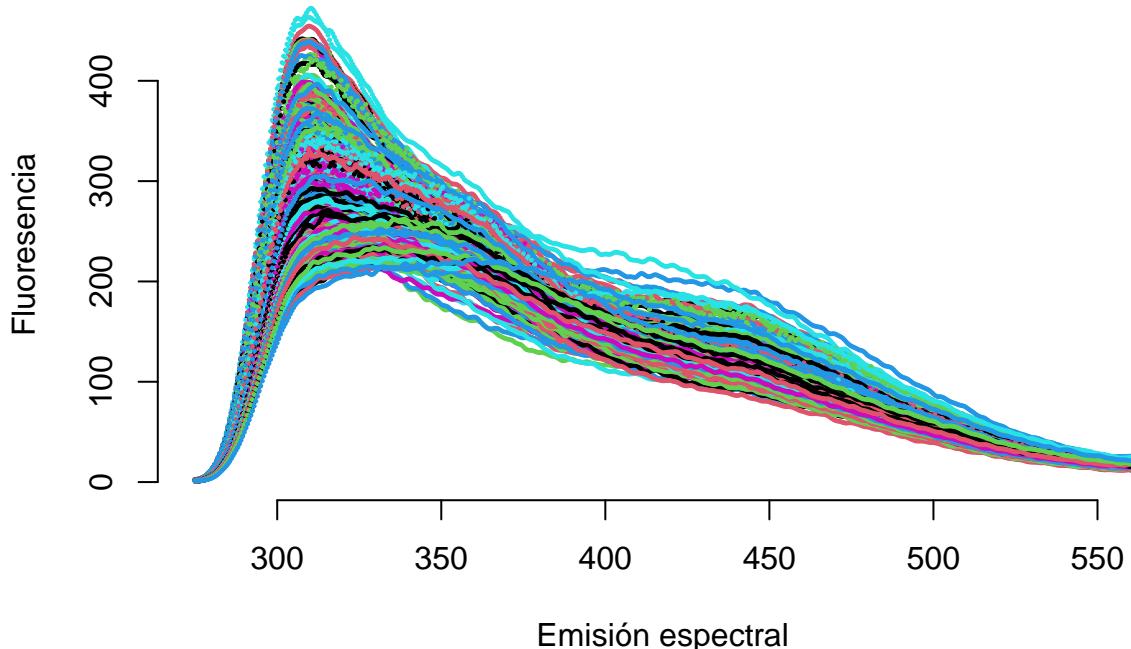
El azúcar se disolvió en agua sin amortiguar y la solución se midió espectrofluorométricamente en una cubeta de $10 \times 10 \text{ mm}^2$. Los datos en bruto se midieron con un fluorómetro. Para cada muestra, los espectros de emisión se midieron en intervalos de 0.5 nm en siete ondas de excitación.

Es decir, tenemos una muestra de tamaño 265, en 571 puntos para cada una de las 7 ondas de longitudes de excitación.

1. En nuestro caso, decidimos tomar el nivel de excitación correspondiente al nivel de onda 340.

Para entender un poco mejor los datos, visualicemos la información sin haberla tratado (i.e. en bruto).

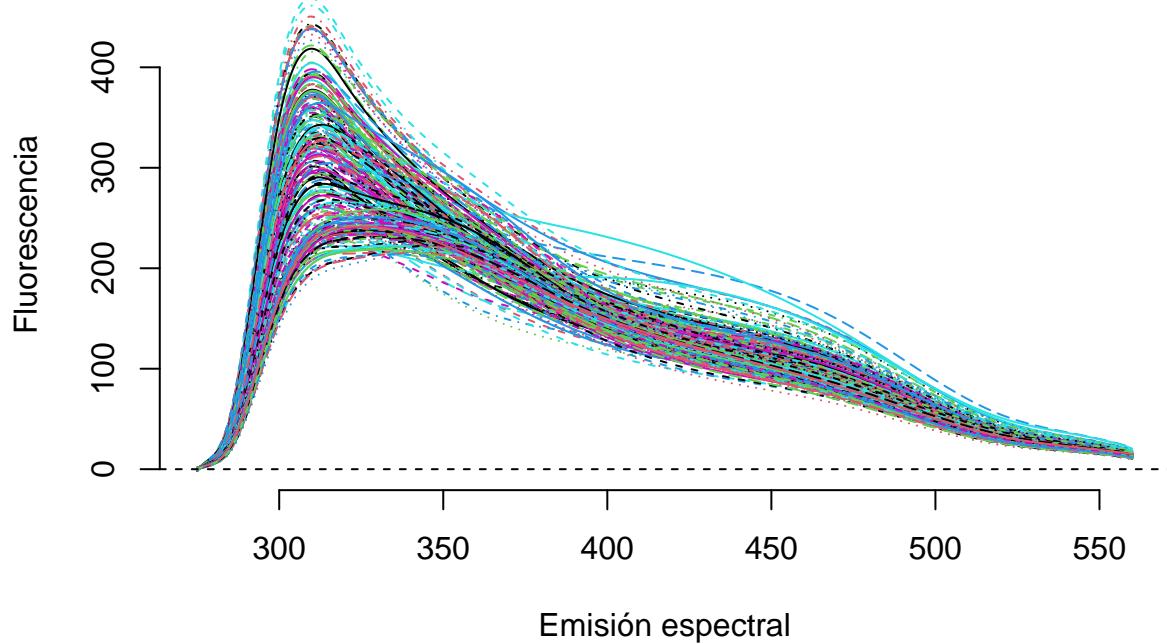
Datos observados



Después de visualizar las funciones (sin suavizar) consideramos como más acertado usar una base de funciones *B-Spline* para llevar a cabo el suavizado, pues en los datos no se perciben comportamientos cílicos (o sinusoidales). Como apunte adicional, mencionamos que para llevar a cabo el ajuste nos bastó con usar polinomios de grado 3 (orden 4) y para la parte que presentaba un apuntamiento muy alto usamos más *breakpoints*.

A continuación podemos ver las versiones suavizadas de la muestra en cuestión.

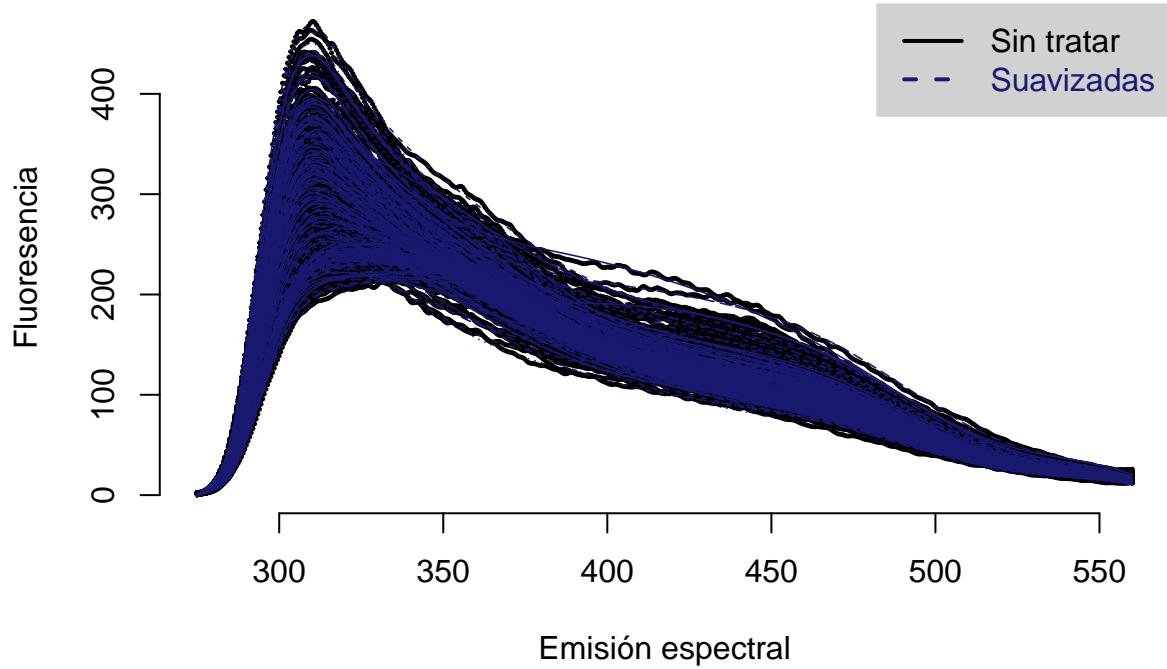
Curvas Suavizadas



```
## [1] "done"
```

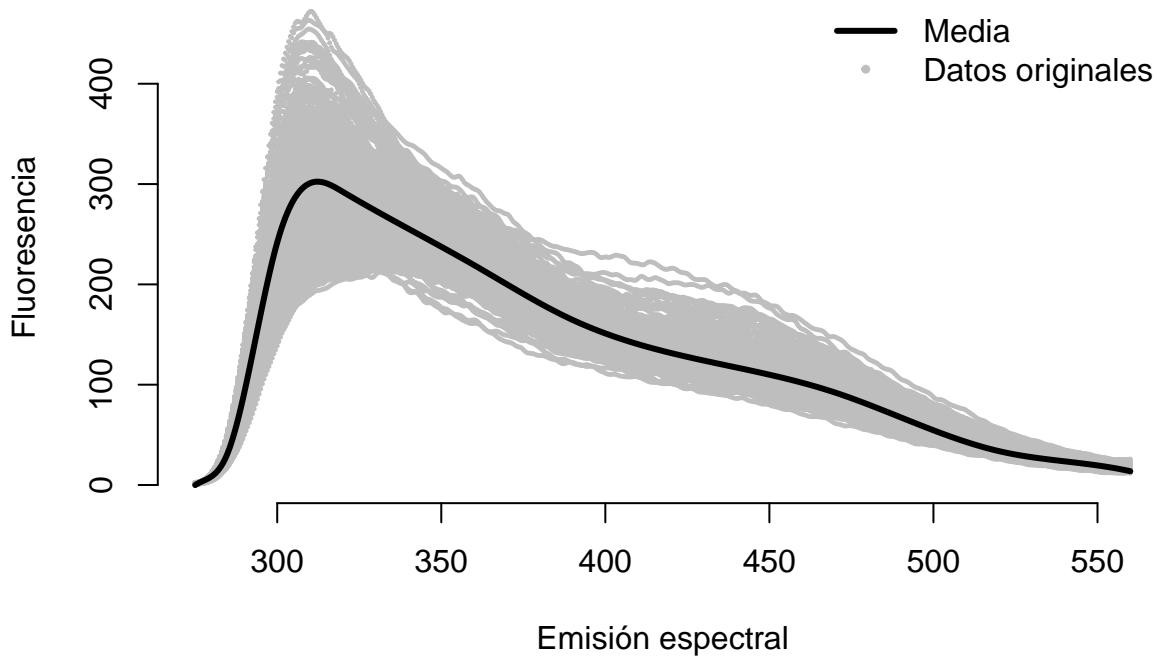
Para contrastar las estimaciones, vamos a superponer los datos sin tratar, junto con las respectivas curvas suavizadas.

Muestras y curvas suavizadas asociadas



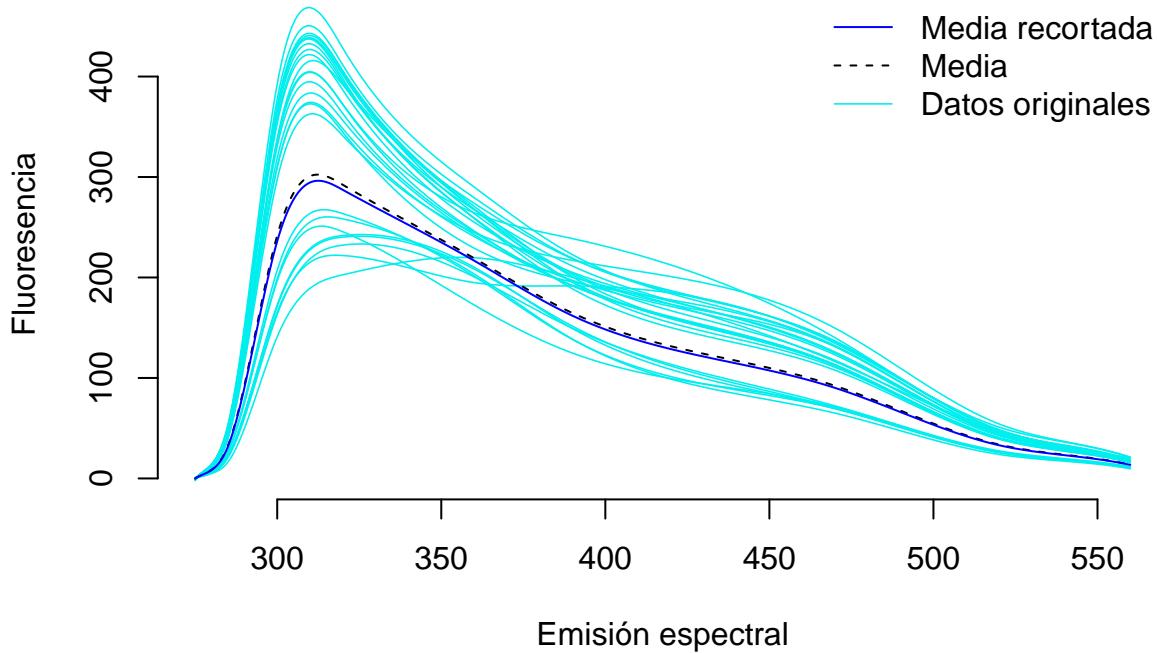
- **Función de media:** la función de media $\bar{x}(t)$ se presenta a continuación

Media y datos originales



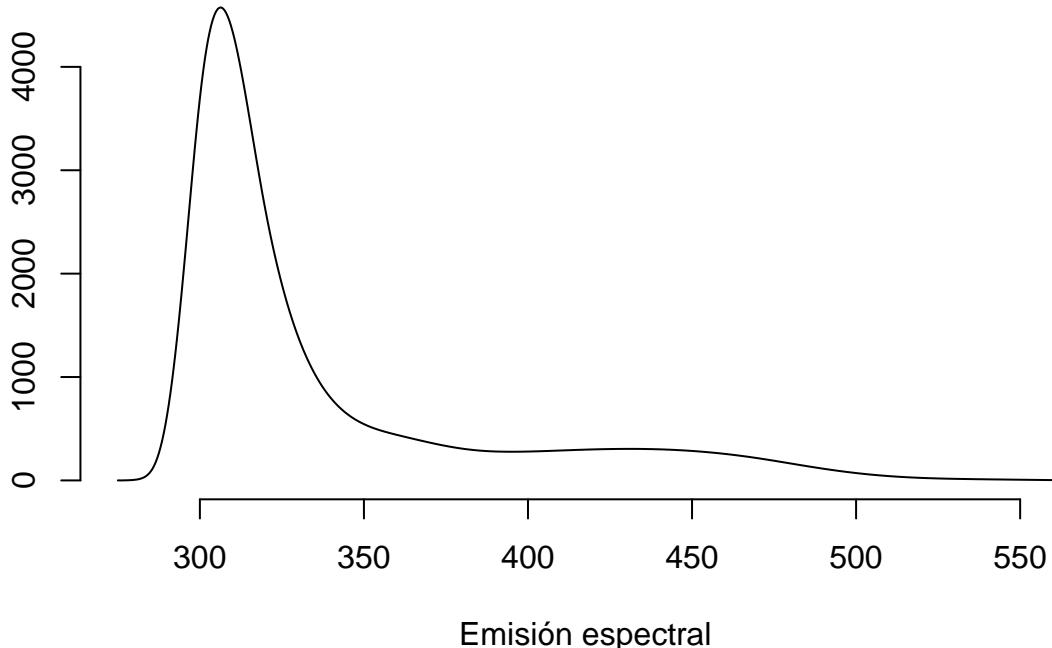
- **Función de media recortada al 10%:** veremos la media recortada, que corresponde al cálculo de la media después de haber recortado una cantidad de observaciones (en este, el 10% de las observaciones *menos centrales*). Vemos que las funciones menos centrales, parecen mostrar un comportamiento, como más apuntados en alguna parte o la parte central en lugar de ser decreciente crece.

Media recortada y datos originales



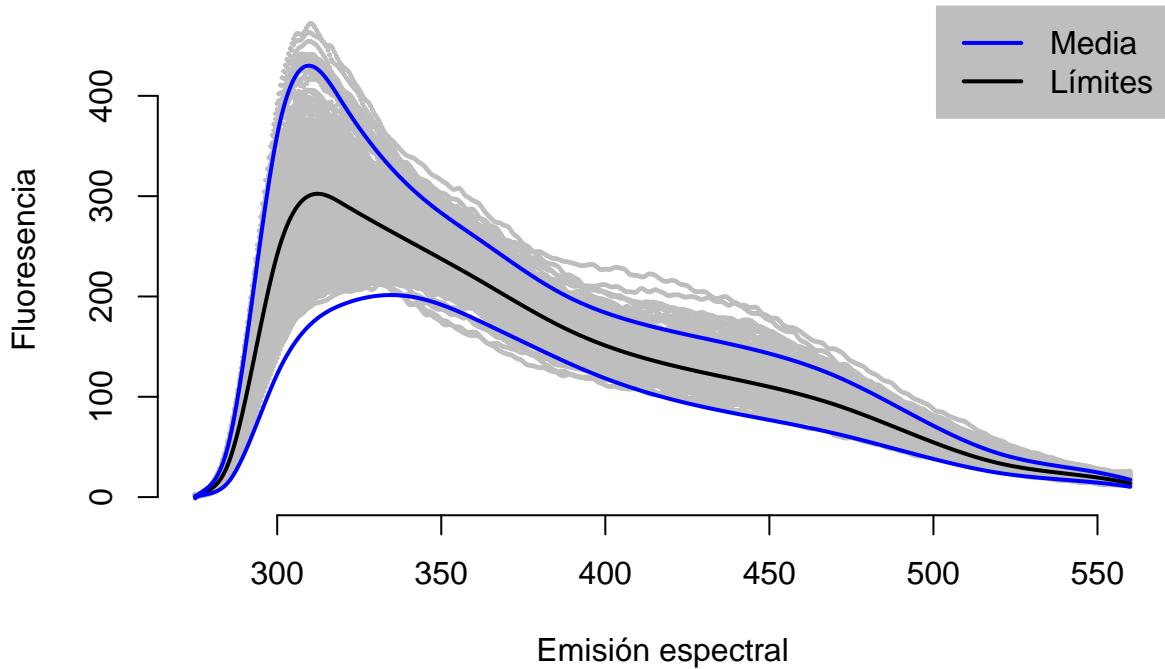
- **Función de varianza:** se calculó usando todos los datos.

Varianza estimada



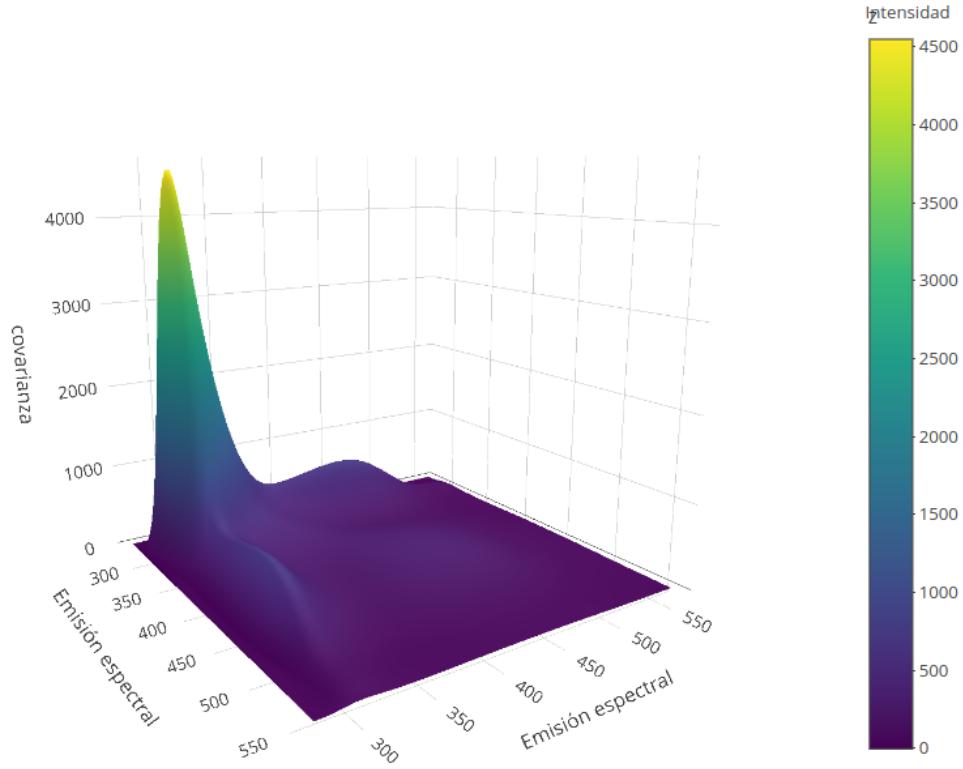
Podemos ver igualmente, una gráfica de una “banda de confianza” (abusamos de términología para orientar la construcción del gráfico, pero no es tal).

"Banda de confianza"



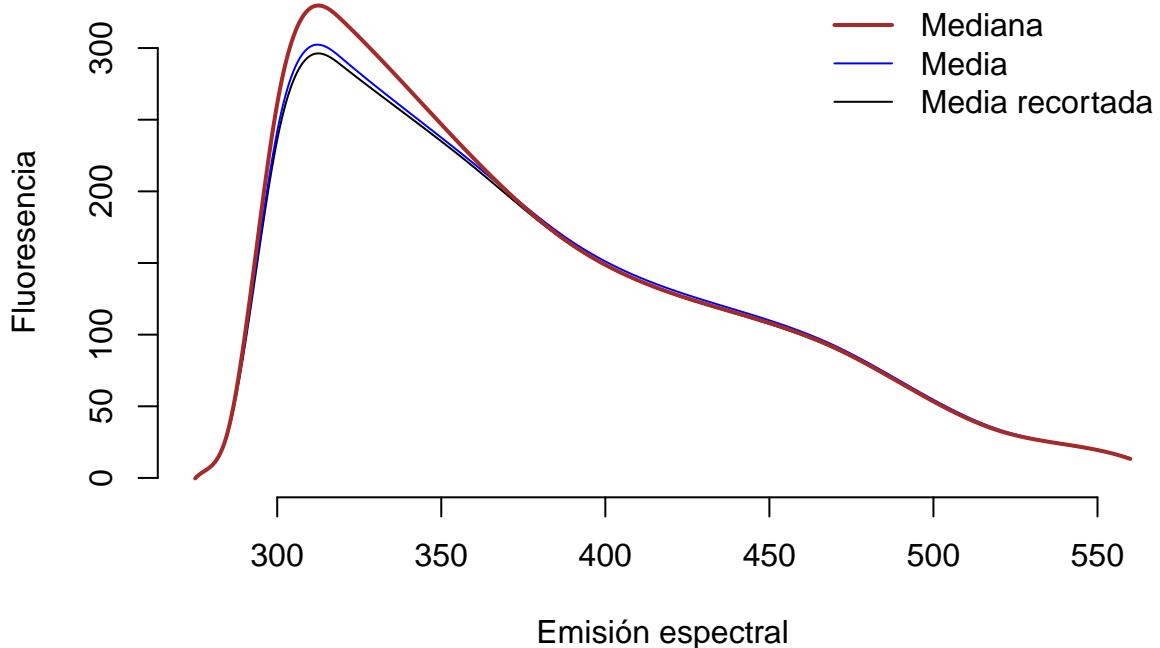
- **Función de covarianza:** mostramos la superficie obtenida. Con la función *persp* obtenemos una

figura no muy agradable. Sin embargo, usaremos la función *plot_ly* que graficará la superficie en una página web y tomamos capturas de pantalla.

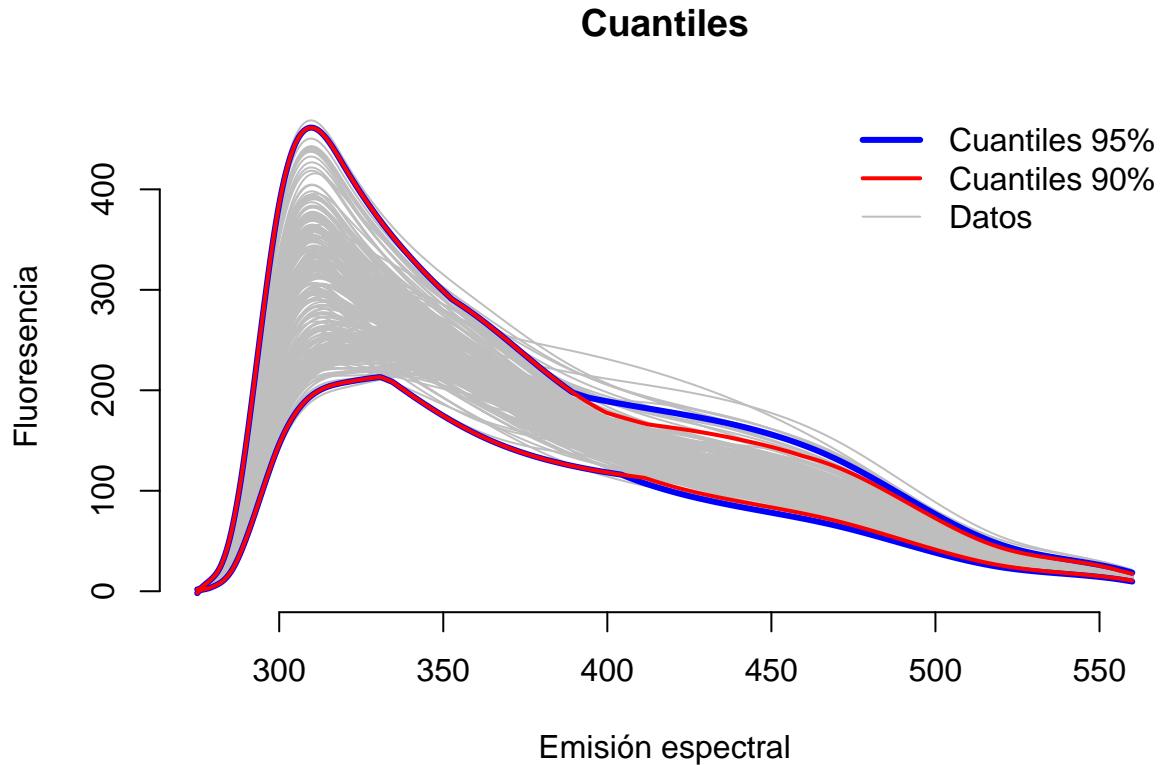


- La función de mediana: corresponde a la estimación con la mayor densidad de todas las estimaciones.

Medidas de tendencia central

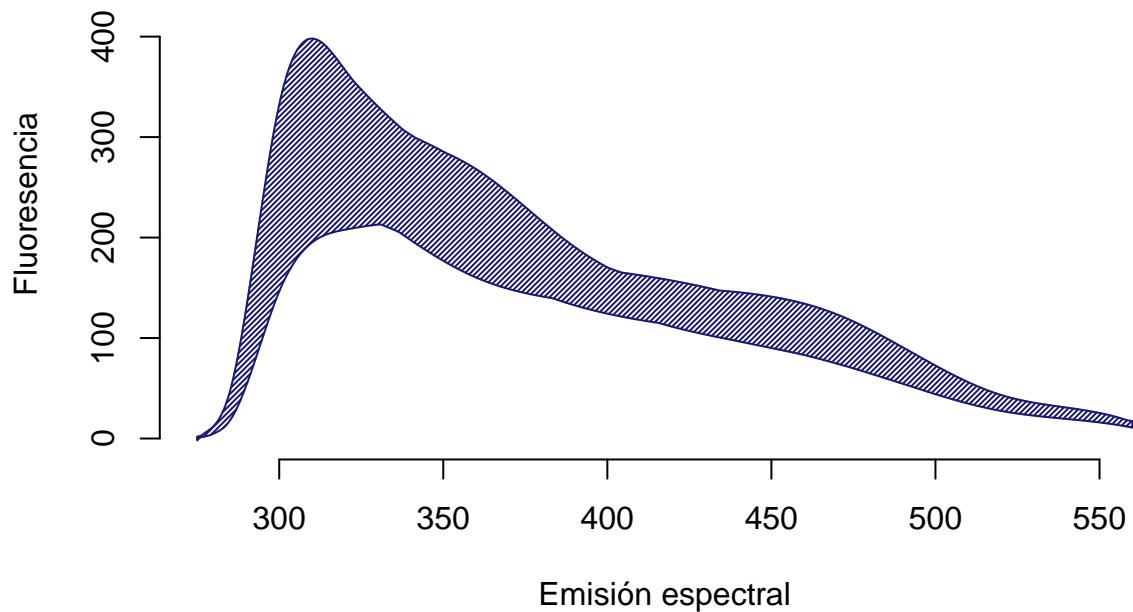


- **Funciones cuantiles 90% y 95%:** en el mismo gráfico presentamos las funciones de los cuantiles solicitados.



- **Región central 0.75:**

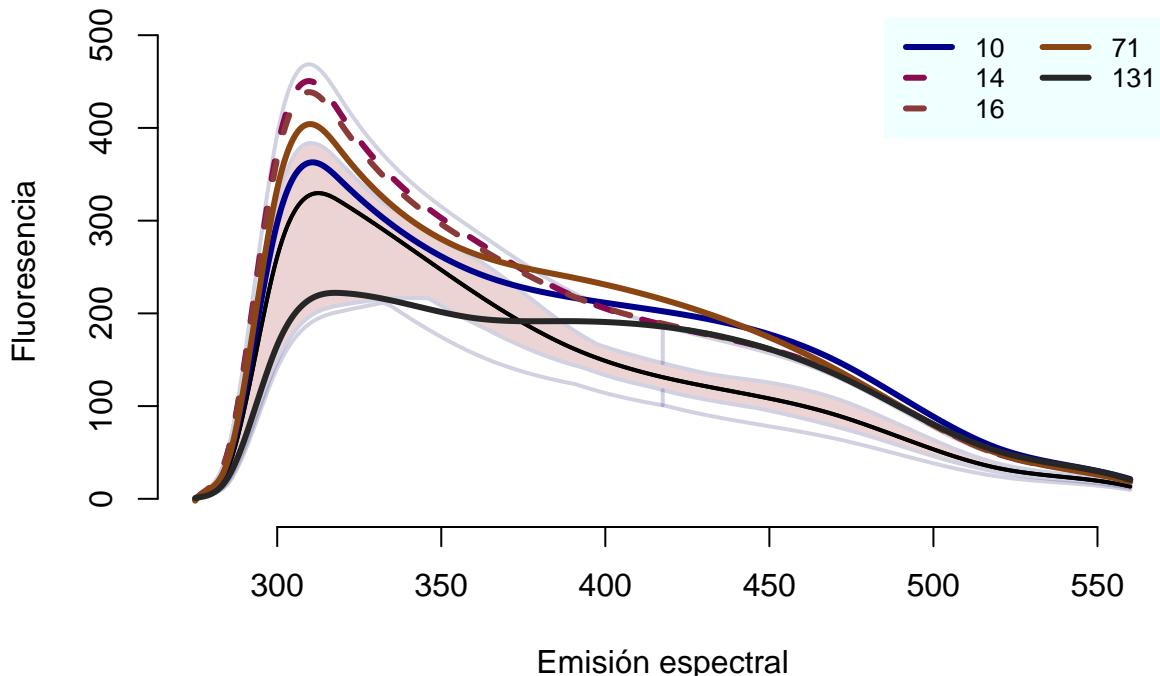
Región 0.75



- **Boxplot funcional y outliers:** los números que aparecen en la leyenda son los identificadores de las

funciones \hat{x} consideradas como atípicas. En este caso, podemos considerar que se tratan de *outliers parciales*. Esto lo podemos ver con las observaciones 131, 71 y 10, pues hay una región (intervalo 350-450 aproximadamente) en la que las curvas o bien decrecen de forma muy suave o bien se comportan de forma creciente. En cambio, el comportamiento de los individuos 14 y 16 es más complejo de entender porqué son *outliers*.

Boxplot funcional



Podemos resaltar que los individuos 14 y 16 están entre los 3 elementos con menor profundidad (aproximadamente un 1.2%).

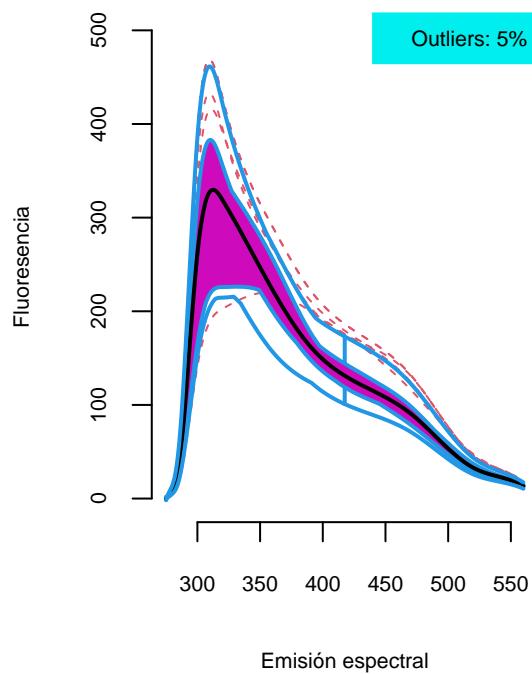
- **Boxplot funcional ajustado y outliers:** la estimación robusta de la matriz de covarianza la llevamos a cabo apartando las funciones que arriba consideramos como atípicas. Sin embargo, la simulación de datos no la logramos hacer como se menciona en las diapositivas.

Como alternativa, proponemos basarnos en técnicas de remuestreo para estimar el boxplot sin usar los datos atípicos y definir el valor del *hiperparámetro F*.

Los gráficos nos ayudan a ver el efecto de haber cambiado el factor F de 1.5 a 1.83 y vemos que la probabilidad de designar un dato atípico cuando no lo es se reduce a un 2%.

```
## [1] 17 40 55 69 85
## [1] 55 69
```

Remuestreo



Reajuste de factor F

