Análisis de Datos Funcionales (FDA): Una Introducción y Desarrollos Recientes

Basado en el documento de Jan Gertheiss et al.

2025-05-22

Contents

1 Introducción al Análisis de Datos Funcionales (FDA)

El Análisis de Datos Funcionales (FDA) es un marco estadístico diseñado para analizar curvas, imágenes o funciones en dominios de dimensiones superiores. En FDA, las observaciones y los objetos de interés son estas funciones o curvas. Aunque los objetivos son a menudo los mismos que para el análisis estadístico de datos escalares o multivariados (como análisis descriptivos, clasificación y regresión), el FDA presenta desafíos adicionales debido a la alta/infinita dimensionalidad de las observaciones y los parámetros.

A diferencia de métodos más simples que reducen las observaciones funcionales a valores de resumen escalares, el FDA retiene toda la información importante utilizando directamente las observaciones funcionales en el análisis. Los datos funcionales se presentan en muchos campos, incluyendo medicina, biofísica, neurociencia y química, y su relevancia crece debido a las tecnologías que permiten su recolección.

1.1 Datos Funcionales, Datos Longitudinales y Series Temporales

El documento distingue el FDA de los datos longitudinales (LD) y las series temporales.

- **Los datos longitudinales (LD)** o *panel data* en ciencias de la salud y sociales consisten en observaciones repetidas a lo largo del tiempo en las mismas unidades (sujetos). Usualmente involucran pocas mediciones repetidas por unidad de observación, potencialmente en puntos de tiempo diferentes.
- **Las series temporales** suelen centrarse en una única serie.
- **Los datos funcionales (FD)**, especialmente los densos, a menudo se registran en el mismo intervalo (de tiempo), con la misma frecuencia y presentan un gran número de mediciones por unidad de observación.

Las formas tradicionales de analizar estos tipos de datos difieren: LD a menudo usa efectos aleatorios en modelos paramétricos, mientras que FD trata las secuencias observadas como realizaciones de un proceso suave modelado con métodos no o semiparamétricos. Sin embargo, hay solapamiento, y métodos de (sparse) FDA se han utilizado para hacer los modelos de LD menos paramétricos y más flexibles. El LD puede verse como datos funcionales dispersos (y ruidosos).

Aquí tienes un cuadro conceptual para visualizar los tipos de datos y su relación:

1.1.1 Cuadro Conceptual 1: Tipos de Datos Relacionados

Tipo de Datos	Descripción Clave	Número de mediciones por unidad	Puntos de medición	Enfoque tradicional en modelado	Relación con FDA
Datos Longitudinales	Observaciones repetidas en las mismas unidades.	Pocas	Potencialmente diferentes	Modelos paramétricos (efectos aleatorios)	LD puede verse como datos funcionales dispersos/ruidosos; FDA se aplica para mayor flexibilidad
Series Temporales	Datos registrados secuencialmente en el tiempo.	Muchas (una única serie)	Generalmente equiespaciados	Modelos de series temporales	FDA (clasificación) aplicable a muestras de series temporales
Datos Funcionales (FD)	Observaciones que son curvas, imágenes o funciones.	Muchas	Iguales (densos) o diferentes (dispersos/irregulares)	Métodos no/semiparamétricos	Marco general para análisis de curvas/funciones
- FD Denso	Muchos puntos de medición, en la misma rejilla y frecuencia.	Grande	Igual, densa	N/A	Ideal para muchas técnicas de FDA
- FD Disperso	Pocos puntos de medición, rejillas específicas por curva.	Pequeño	Específica por curva	N/A	Requiere métodos que incorporen suavizado en el análisis
- FD Irregular	Rejillas diferentes pero relativamente densas.	Grande	Diferente, densa	N/A	Similar a FD denso pero con rejillas variables

En la práctica, los datos funcionales solo se observan en una rejilla discreta de puntos de medición. Un modelo típico asume que la observación es la función subyacente (suave) más un término de error de medición. Los datos pueden ser densos (rejilla igual y densa) o dispersos (rejillas específicas de cada curva y escasas).

2 Conceptos Fundamentales en FDA (Sección 2)

2.1 Descriptivos y Detección de Outliers

El primer paso suele ser un análisis descriptivo, incluyendo visualizaciones e identificación de observaciones atípicas (outliers).

En FDA, los outliers pueden diferir de la mayoría de la muestra por el rango de sus valores ("outliers de magnitud") o por su forma ("outliers de forma").

Es necesario definir una noción de "atipicidad" o "centralidad" para los datos funcionales, lo cual se hace usando las profundidades de datos funcionales. Existen diferentes definiciones de profundidad.

El documento menciona el *boxplot* funcional (basado en *band-depth*), el *bagplot* funcional y los *rainbow plots* como herramientas visuales para el análisis descriptivo y la detección de outliers. Sin embargo, señala que algunas implementaciones usan profundidades multivariadas basadas en componentes principales funcionales (FPCs) en lugar de profundidades funcionales "propias", capturando solo una parte de la variación.

Software: paquetes `fda` y `rainbow` en R para boxplots funcionales, bagplots y rainbow plots. `fdaoutlier` ofrece una colección de funciones para detección de outliers funcionales.

2.2 Suavizado para Datos Funcionales

Los datos funcionales observados a menudo incluyen error de medición y se registran en rejillas finitas. El objetivo del suavizado es reconstruir las curvas subyacentes "verdaderas".

Hay diferentes enfoques:

- **Presuavizado:** Suavizar cada curva individualmente antes del análisis. El documento señala que esto no funciona bien, especialmente para datos dispersos, ya que no considera la incertidumbre del suavizado ni permite "compartir información" entre funciones.
- **Incorporar el suavizado en el análisis:** Métodos más recientes integran el suavizado directamente en el procedimiento de análisis estadístico. Esto es particularmente útil para datos dispersos, ya que permite "compartir fuerza" (*borrowing of strength*) entre funciones. Un ejemplo destacado es el enfoque de Yao, Müller y Wang (2005) para FPCA, donde el suavizado se incorpora en la estimación de la función de covarianza. Ideas similares se han aplicado en regresión funcional.

El suavizado a menudo se realiza utilizando funciones base (como B-splines, ver Figura 3 en el documento) o a través de la estimación suavizada de la función de covarianza (como en FPCA).

Se pueden añadir penalizaciones a las derivadas de la función estimada o a las diferencias de coeficientes de base vecinos para fomentar la suavidad, similar a los P-splines.

2.3 Variación de Amplitud y Fase

Los datos funcionales a menudo presentan variación en amplitud (diferencias en los valores de la función) y fase (diferencias en la alineación temporal o el "tiempo"). La variación de fase puede oscurecer la variación de amplitud si las curvas no están alineadas.

El alineamiento funcional busca eliminar las diferencias de fase deformando el dominio de tiempo de las curvas para que los "eventos" clave ocurran en los mismos puntos temporales.

El documento menciona la distancia elástica como una métrica que permite medir la distancia entre funciones independientemente de la deformación temporal (*warping*). Un ejemplo visual (Figura 4) muestra curvas originales y las mismas curvas alineadas a su media.

El FPCA tradicional puede verse afectado por la variación de fase. Un FPCA alternativo podría alinear las funciones primero.

2.4 Análisis de Componentes Principales Funcionales (FPCA)

El FPCA es una técnica clave para la reducción de dimensionalidad y la representación de datos funcionales. Se basa en la expansión de Karhunen–Loève.

Representa una función aleatoria $X_i(t)$ como la suma de una función media $\mu(t)$ y una serie infinita de productos de scores aleatorios ($\xi_{m,i}$) y funciones propias ($\phi_m(t)$).

Las funciones propias $\phi_m(t)$ forman una base ortonormal (análoga a los vectores propios en PCA multivariado) y capturan los principales modos de variación alrededor de la media. Los scores $\xi_{m,i}$ cuantifican cuánto contribuye cada modo de variación a una función individual.

En la práctica, la expansión se trunca a un número finito M de componentes. La proporción de varianza explicada por los primeros M componentes se puede calcular.

La estimación práctica del FPCA requiere estimar la función media, las funciones propias, los valores propios y los scores. Para datos dispersos o con error, se recomienda incorporar el suavizado en la estimación de la función de covarianza, a partir de productos cruzados de observaciones de cada curva. La descomposición ortogonal se realiza numéricamente.

Los scores estimados $\hat{\xi}_i$ proporcionan un vector resumen multivariado de las funciones individuales, permitiendo la reducción de dimensionalidad y la aplicación de muchos métodos multivariados a datos funcionales. También permiten predecir la función completa incluso a partir de datos individuales dispersos, compartiendo información entre funciones.

Software: la función `fpca.face` en el paquete `refund` en R. Otros paquetes como `fdapace` también están disponibles.

3 Regresión Funcional (Sección 3)

La regresión funcional se aplica cuando al menos una variable (respuesta o covariable) es funcional. El documento distingue tres configuraciones principales:

3.0.1 Cuadro Conceptual 2: Tipos de Regresión Funcional

Tipo de Regresión	Descripción	Variables	Modelo Típico	Interpretación del efecto funcional
Scalar-on-Function(s) (SOFR)	Una respuesta escalar se modela en función de una o más covariables funcionales (y opcionalmente escalares).	Respuesta: Escalar; Covariable(s): Funcional(es) (y/o Escalar)	Modelo lineal funcional: $Y_i = \alpha + \int x_i(t)\beta(t)dt + \epsilon_i$. GLM funcional (para respuesta no Gaussiana): $g(\mu_i) = \eta_i = \alpha + \int x_i(t)\beta(t)dt + ...$	La función coeficiente $\beta(t)$ muestra cómo los valores de la covariable funcional en diferentes puntos t se asocian con la respuesta escalar.
Function-on-Scalar (FOSR)	Una respuesta funcional se modela en función de una o más covariables escalares.	Respuesta: Funcional; Covariable(s): Escalar(es)	Modelo lineal: $Y_i(t) = \alpha(t) + z_i\gamma(t) + \epsilon_i(t)$ (Ejemplo simple, no directamente del texto principal pero implícito)	Los coeficientes (p.ej., $\alpha(t), \gamma(t)$) son funciones que describen cómo las covariables escalares influyen en la forma de la función de respuesta en cada punto t.
Function-on-Function(s) (FOFR)	Una respuesta funcional se modela en función de una o más covariables funcionales (y opcionalmente escalares).	Respuesta: Funcional; Covariable(s): Funcional(es) (y/o Escalar)	Modelo lineal (concurrent): $Y_i(t) = \alpha(t) + x_i(t)\beta(t) + \epsilon_i(t)$ (solo efecto "local" en el mismo t). Modelos más generales también existen.	La función coeficiente $\beta(t)$ en un modelo concurrente muestra la asociación local en cada punto t. Modelos más complejos consideran la integral de la covariable funcional.

El documento se centra principalmente en los enfoques semiparamétricos utilizando funciones base y enfoques no paramétricos utilizando kernels.

3.1 Enfoques Semiparamétricos y el Enfoque de Funciones Base

La idea central es representar las funciones coeficientes desconocidas (como $\beta(t)$) como una combinación lineal de un conjunto de funciones base predefinidas (como B-splines o funciones propias de FPCA).

Las integrales en los modelos (como en SOFR) se aproximan típicamente mediante integración numérica.

Al usar el enfoque de funciones base en modelos lineales (generalizados) funcionales, el modelo funcional se transforma en un modelo lineal (generalizado) estándar donde los coeficientes de base son los coeficientes de regresión desconocidos a estimar.

Cuando se utilizan muchas funciones base (como con B-splines), las funciones coeficientes pueden sobreajustar los datos. Una solución popular es añadir una penalización a las derivadas (o diferencias de coeficientes) para fomentar la suavidad. Esto lleva a modelos lineales (generalizados) penalizados.

Una alternativa es usar los scores de FPCA como covariables en un modelo lineal (generalizado) estándar. Como pocos componentes principales suelen capturar la información principal, el número de coeficientes de regresión es pequeño y no requiere regularización adicional. Sin embargo, el FPCA es un método no supervisado (no considera la respuesta), por lo que los scores pueden no ser las mejores dimensiones para el problema de regresión específico.

La perspectiva de modelos mixtos es una forma popular y elegante de tratar las penalizaciones cuadráticas. Equivale a interpretar los coeficientes de base como efectos aleatorios. Esto permite estimar los parámetros de suavizado (parámetros de penalización) utilizando máxima verosimilitud o verosimilitud restringida. También proporciona herramientas para inferencia estadística (pruebas, intervalos de confianza).

La selección de variables es importante, especialmente con muchas covariables potenciales. Se han propuesto enfoques que utilizan penalizaciones que inducen escasez (*sparsity-inducing penalties*). También se pueden usar métodos más clásicos como la selección *forward/backward* basada en pruebas estadísticas. La selección de variables a veces se interpreta como seleccionar los puntos o regiones más predictivos en el dominio de las funciones.

3.2 Enfoques No Paramétricos

Estos enfoques ofrecen aún mayor flexibilidad en cuanto a la estructura de asociación entre la respuesta y las covariables.

Los métodos no paramétricos (como la regresión por kernels) suelen depender de una medida de distancia o proximidad entre las observaciones funcionales.

La elección de la (semi-)métrica d es crucial. Las opciones incluyen la norma $\mathcal{L}_2$ (una métrica) o semi-métricas basadas en derivadas (que no implican $X_i = x$ si $d(X_i, x) = 0$). La elección depende de la forma de los datos y el objetivo del análisis. Los resultados pueden variar significativamente según la métrica elegida.

El FPCA también puede usarse para definir distancias en el espacio de scores reducidos.

Los métodos no paramétricos presentados en el documento suelen estar restringidos a casos con una sola covariable funcional. Existen extensiones para múltiples covariables.

Software: Rutinas `npfda` en R.

Software para Regresión Funcional: Los paquetes R `fda` y `refund` (que usa metodología de `mgcv`) permiten ajustar una amplia gama de modelos de regresión funcional. `mgcv` también puede usarse directamente. `funGp` y `GPFDA` para enfoques basados en procesos Gaussianos. `FDboost` para modelos de boosting.

4 Inferencia Estadística con Datos Funcionales (Sección 4)

La inferencia estadística, como las pruebas de hipótesis y los intervalos de confianza, es un área importante en FDA.

4.1 Análisis de Varianza Funcional (ANOVA)

Permite comparar las funciones medias de datos funcionales entre diferentes grupos. Es un caso especial de prueba de hipótesis lineal en FOSR. Se han propuesto diversas pruebas (Figura 10 muestra p-valores para diferentes pruebas).

Puede generalizarse a ANOVA de múltiples factores o a datos funcionales multivariados.

4.2 Pruebas e Intervalos de Confianza en Regresión Funcional

Para FOSR, se han propuesto pruebas F y basadas en la norma $\mathcal{L}_2$ para hipótesis lineales.

Para SOFR, las pruebas comunes incluyen:

- Probar si una covariable funcional tiene algún efecto (es decir, si la función coeficiente es cero).
- Probar si el modelo lineal es adecuado.

Se han propuesto diferentes procedimientos de prueba, incluyendo pruebas de ratio de verosimilitud restringida (RLRT) usando la perspectiva de modelos mixtos, pruebas tipo Wald, Score, Likelihood Ratio y F extendidas usando FPCA, y pruebas de bondad de ajuste basadas en proyecciones aleatorias.

Los intervalos de confianza puntuales para la función coeficiente en el modelo lineal (generalizado) funcional a menudo se basan en la perspectiva Bayesiana del proceso de suavizado (usando la perspectiva de modelos mixtos). Se calculan de forma similar a los intervalos para coeficientes en modelos lineales, pero aplicados a los coeficientes de base transformados a los puntos de evaluación. La Figura 6 en el documento muestra intervalos de confianza puntuales sombreados para funciones coeficiente estimadas.

5 Clasificación y Clustering Funcional (Sección 5)

Estos son problemas de aprendizaje, siendo la clasificación supervisada (etiquetas de clase conocidas) y el clustering no supervisado (descubrir grupos).

5.1 Clasificación de Datos Funcionales

El objetivo es asignar nuevas observaciones funcionales a una de varias clases distintas utilizando un modelo entrenado con datos de entrenamiento con etiquetas de clase conocidas.

Se pueden usar modelos como el modelo logit funcional (un caso de GLM funcional) para estimar probabilidades de clase condicionales. La Figura 6 (derecha) muestra una función coeficiente de un modelo logit funcional para predecir el sexo a partir de perfiles de aceleración.

Otros enfoques incluyen:

- Clasificación por kernel usando (semi-)métricas para medir la proximidad entre curvas y ponderar las observaciones de entrenamiento.
- k-Nearest Neighbor (kNN) funcional, asignando una nueva curva a la clase más frecuente entre sus k vecinos más cercanos (definidos por una métrica/semi-métrica).
- Análisis Discriminante Lineal (LDA) funcional.
- Usar scores de FPCA como entrada para algoritmos de clasificación multivariada estándar (bosques aleatorios, redes neuronales, etc.).
- Usar datos funcionales directamente en algunas situaciones, como entrada para una Red Neuronal Convolucional (CNN).

Software: El paquete `refund` puede ajustar el modelo logit funcional (vía `pfr()`). Rutinas para clasificación por kernel y kNN pueden encontrarse en el sitio de Ferraty y Vieu.

5.2 Clustering Funcional

Es un problema de aprendizaje no supervisado cuyo objetivo es agrupar datos funcionales de manera que los grupos (clusters) sean homogéneos.

Los algoritmos de clustering funcional pueden clasificarse según si son basados en datos/observaciones vs. basados en modelos, y métodos de datos crudos vs. reducción de dimensionalidad.

- **Métodos de datos crudos:** Los clusters se forman basándose en los puntos de evaluación de las curvas.
- **Reducción de dimensionalidad:**
  - **Métodos de filtrado:** Primero aproximan las funciones (p.ej., usando funciones base) y luego usan los coeficientes de base para el clustering.
  - **Métodos adaptativos:** Realizan la reducción de dimensionalidad y el clustering simultáneamente, permitiendo que las bases o el número de bases dependan del cluster. El clustering basado en modelos (como modelos de mezcla basados en expansión de bases o FPCA específica de cluster) entra en esta categoría.

El documento ilustra el clustering jerárquico en datos de ejemplo, mostrando un dendrograma y la distribución de los clusters en el espacio de los primeros dos componentes principales.

Software: Paquetes estándar como `cluster` y `mclust` (para clustering basado en modelos multivariado). Paquetes funcionales específicos como `funHDDC` para clustering basado en modelos con FPCA.

6 Enfoques de Machine Learning (Sección 6)

FDA también se cruza con el Machine Learning (ML) y el Deep Learning (DL). El ML a menudo se centra en el rendimiento predictivo.

6.1 Enfoques Implícitos

Estos enfoques no tienen en cuenta explícitamente la naturaleza funcional de los datos (p.ej., la suavidad).

- Tratan las covariables funcionales como secuencias o series temporales multivariadas.
- Calculan estadísticas de resumen escalares (media, moda, etc.) de las funciones y las usan como entradas para métodos ML clásicos.
- Representan funciones multivariadas como "imágenes" (concatenando puntos de evaluación) para usar CNNs pre-entrenadas, basado en transferencia de aprendizaje.
- Para resultados funcionales, a menudo se les llama regresión de trayectoria o modelos *sequence-to-sequence*.

6.2 Machine Learning Clásico para Datos Funcionales

Algunos métodos ML se han adaptado para datos funcionales explícitamente.

- **Support Vector Machines (SVM):** Usan kernels para medir la distancia entre observaciones. Se ha propuesto mapear funciones a un espacio euclidiano donde se pueden aplicar kernels clásicos.

- **Gaussian Processes (GPs):** Adaptados para regresión funcional (GPFR), combinando la estimación de la media de un modelo de regresión funcional con un GP multivariado. Los efectos funcionales definen la media, y la información temporal/funcional se usa para estimar la covarianza del GP.
- **Boosting:** Enfoques de *gradient boosting* para modelos de regresión funcional. Implementaciones como `FDboost` utilizan un descenso de gradiente por coordenadas que induce escasez.
- **Métodos basados en árboles:** Adaptados para covariables funcionales (bosques aleatorios) o para resultados funcionales (functional random forests).

6.3 Deep Learning y Redes Neuronales para Datos Funcionales

Se han combinado redes neuronales con datos funcionales, como *multilayer perceptrons* (MLPs) funcionales.

Las Redes Neuronales Convolucionales (CNNs) pueden usar datos funcionales directamente, a menudo representados como imágenes.

Software para ML/DL: Paquetes R como `FDboost`, `FuncNN`, `funGp`, `GPFDA`. El paquete `scikit-fda` en Python/R proporciona transformaciones de datos funcionales para ML implícito.

7 Campos Relacionados y Perspectivas (Sección 7)

7.1 FDA y Campos Relacionados

Como se mencionó, FDA ha influido en el análisis de datos longitudinales y series temporales. Los métodos FDA semiparamétricos son útiles en estos campos.

La visión de LD como datos funcionales dispersos permite el uso de FPCA (con scores aleatorios) para obtener modelos más flexibles que los modelos lineales mixtos paramétricos estándar.

FDA también se relaciona con datos espaciotemporales, que pueden verse como campos aleatorios espaciales repetidos o curvas correlacionadas espacialmente.

Datos funcionales pueden tener otras estructuras como series temporales funcionales o datos funcionales longitudinales. Estos y otros se denominan "datos funcionales de segunda generación".

7.2 Más allá de las Funciones Unidimensionales

FDA se está expandiendo para ver los datos funcionales como un caso especial de "datos de objeto" (*object data*), donde cada observación es un objeto en un espacio más complejo que $\mathbb{R}^d$. Esto se llama a veces análisis de datos orientado a objetos.

Ejemplos incluyen:

- **Imágenes y datos con valores de superficie:** Vistos como funciones en un dominio de dimensión superior ($d > 1$).
- **Funciones multivariadas:** Definidas en un intervalo pero con valores vectoriales en $\mathbb{R}^d$ ($d > 1$). Ejemplos: trayectorias de movimiento.

Muchos métodos de FDA unidimensional (como FPCA y regresión) se están extendiendo a imágenes y datos funcionales multivariados.

Otras áreas relacionadas incluyen:

- **Análisis de forma (*Shape Analysis*):** Análisis de objetos invariante a rotación, traslación y escala. Se cruza con FDA en el análisis de forma funcional (*shape of a curve*). Las formas viven en variedades (*manifolds*) o espacios cociente.
- **Análisis de datos composicionales:** Vectores de entradas no negativas que suman una constante. Se cruza con FDA al ver las densidades de probabilidad como datos funcionales con restricciones o como composiciones infinitas, analizadas en el espacio de Bayes–Hilbert u otros espacios como los espacios de Wasserstein.
- **Datos funcionales generalizados:** Donde los valores observados, condicionados a una curva media suave, provienen de una distribución no Gaussiana (p.ej., binaria, conteo).
- **Otros tipos de datos de objeto:** datos direccionales, datos con valores de covarianza (matrices u operadores), datos con valores de procesos puntuales, y datos con valores de árboles o redes.

El documento concluye enfatizando que el FDA es un campo activo y emocionante con muchos desarrollos actuales, particularmente en la generalización a "datos de objeto".

7.2.1 Cuadro Conceptual 3: Enfoques Clave de Análisis en FDA

Área de Análisis	Descripción / Objetivo	Métodos Comunes / Ideas Clave	Software Típico (en R)
Descriptivos y Outliers	Visualizar y resumir datos funcionales; identificar curvas atípicas.	Profundidades funcionales; Boxplots funcionales; Bagplots; Rainbow plots.	`fda`, `rainbow`, `fdaoutlier`

Área de Análisis	Descripción / Objetivo	Métodos Comunes / Ideas Clave	Software Típico (en R)
Suavizado	Reconstruir curvas subyacentes a partir de observaciones discretas y ruidosas.	Presuavizado (generalmente menos recomendado para datos dispersos); Incorporar suavizado en el análisis (p.ej., en estimación de covarianza para FPCA, o en regresión); Funciones base (B-splines); Penalizaciones.	`fda`, `refund` (integra suavizado), `mgcv` (para penalizaciones).
FPCA	Reducción de dimensionalidad; encontrar modos principales de variación; construir una base *data-driven*.	Expansión de Karhunen–Loève; Estimación de covarianza suavizada; Estimación de scores (integración vs. expectativas condicionales/modelo mixto).	`fda`, `refund` (`fpca.face`), `fdapace`, `MFPCA` (para datos multivariados).
Regresión Funcional	Modelar la relación entre variables funcionales y/o escalares.	SOFR, FOSR, FOFR. Enfoques semiparamétricos (funciones base, penalizaciones, perspectiva de modelo mixto); Enfoques no paramétricos (kernels, (semi-)métricas). Selección de variables.	`fda`, `refund`, `mgcv`, rutinas `npfda`, `funGp`, `GPFDA`, `FDboost`.
Inferencia Estadística	Realizar pruebas de hipótesis y construir intervalos de confianza.	ANOVA funcional (comparación de medias); Pruebas para efectos de covariables funcionales (si $\beta(t) = 0$); Pruebas de bondad de ajuste; Intervalos de confianza/credibilidad (p.ej., puntuales usando perspectiva Bayesiana/modelo mixto).	`fdatest` (interval testing), métodos implementados en `refund`/`mgcv`.
Clasificación	Asignar observaciones funcionales a clases conocidas (aprendizaje supervisado).	Modelos logit funcionales; Clasificación por kernel; kNN funcional; LDA funcional; Uso de scores FPCA en ML multivariado; CNNs.	`refund` (logit funcional), rutinas `npfda`.

Área de Análisis	Descripción / Objetivo	Métodos Comunes / Ideas Clave	Software Típico (en R)
Clustering	Agrupar observaciones funcionales en clusters desconocidos (aprendizaje no supervisado).	Basado en datos (datos crudos, scores FPCA con k-means/jerárquico); Basado en modelos (modelos de mezcla, FPCA específica de cluster). Uso de distancias/semi-métricas.	`cluster`, `mclust`, `funHDDC`.
Machine Learning	Usar métodos ML para problemas de FDA, a menudo centrándose en la predicción.	Implícito (series temporales, estadísticas resumen, imágenes); Explícito (SVM funcionales, GPs funcionales, Boosting funcional, Bosques aleatorios funcionales). Deep Learning (MLP, CNNs).	`FDboost`, `FuncNN`, `funGp`, `GPFDA`, `scikit-fda` (transformaciones).
Datos de Objeto (Generalización)	Extender FDA a datos más complejos como imágenes, funciones multivariadas, formas, densidades, redes.	Extensiones de FPCA, regresión y otros métodos a estos tipos de datos y geometrías asociadas (variedades, espacios cociente, espacios de Hilbert/Wasserstein).	`MFPCA`, `elasdics`, `fdasrvf` (elástico), software específico por tipo de objeto.

Este resumen cubre los puntos principales del documento, explicando los conceptos fundamentales, las técnicas de análisis clave (regresión, clasificación, clustering), su relación con otros campos de análisis de datos temporales y la dirección de investigación hacia el análisis de datos de objeto más generales. Los cuadros conceptuales buscan ofrecer una estructura clara para ayudarte a visualizar y entender las diferentes facetas del FDA presentadas en el texto.