

## **Análisis de Datos Longitudinales**

Código: 104879

Créditos ECTS: 6

| Titulación                   | Tipo | Curso | Semestre |
|------------------------------|------|-------|----------|
| 2503852 Estadística Aplicada | OT   | 4     | 0        |

La metodología docente y la evaluación propuestas en la guía pueden experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

### **Contacto**

Nombre: Juan Ramón González Ruiz

Correo electrónico: JuanRamon.Gonzalez@uab.cat

### **Uso de idiomas**

Lengua vehicular mayoritaria: español (spa)

Algún grupo íntegramente en inglés: No

Algún grupo íntegramente en catalán: No

Algún grupo íntegramente en español: No

### **Prerequisitos**

Es recomendable, pero no imprescindible porque se realizará una clase para homogeneizar el nivel de los alumnos, conocer los modelos lineales generalizados y el modelo de Cox para el análisis de supervivencia. La asignatura sobre estudios observacionales contiene una introducción complementaria al tema que versará sobre el análisis de la evolución temporal de tasas de incidencia y mortalidad en el que se introducirá el concepto de tasa.

Es recomendable pero no imprescindible conocer el paquete estadístico R

### **Objetivos y contextualización**

Los objetivos principales de esta asignatura son:

- Conocer los modelos estadísticos para el análisis de datos longitudinales (información que se obtiene a partir de mediciones efectuadas a lo largo del tiempo) que suelen aparecer frecuentemente en ciencias de la salud (biología, medicina, farmacología, toxicología, química y/o ingeniería)
- Conocer los modelos estadísticos para analizar la evolución temporal de las tasas de incidencia y mortalidad de una enfermedad para detectar cambios temporales y porqué son debidos
- Conocer los modelos estadísticos para analizar el tiempo hasta la ocurrencia de un evento de interés que aparece de forma recurrente (recaídas tumorales, migraña, infartos, ...) teniendo en cuenta el efecto de covariables, el efecto de la intervención y/o el efecto de observar varios eventos con anterioridad
- Conocer los modelos estadísticos para analizar datos obtenidos a partir de medidas repetidas a lo largo del tiempo utilizando modelos lineales (rehospitalizaciones, recaída de una enfermedad, ...)
- Conocer los modelos estadísticos para analizar datos obtenidos a partir de medidas repetidas a lo largo del tiempo utilizando modelos no lineales (crecimiento tumoral en ratas, evolución del peso de los niños tras nacer, ...)
- Ser capaces de leer de forma crítica un artículo científico en el que se plantea el análisis de un estudio en el que se disponga de información recogida a lo largo del tiempo.

- Ser capaces de identificar el modelo estadístico necesario para analizar un conjunto de datos que se presentarán en ejercicios prácticos y que pertenecen a estudios reales.
- Saber cómo realizar todos estos análisis utilizando R mediante las librerías adecuadas.

## Competencias

- Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otras personas.
- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- Formular hipótesis estadísticas y desarrollar estrategias para confirmarlas o refutarlas.
- Identificar la utilidad y la potencialidad de la estadística en las distintas áreas de conocimiento y saber aplicarla adecuadamente para extraer conclusiones relevantes.
- Interpretar resultados, extraer conclusiones y elaborar informes técnicos en el campo de la estadística.
- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
- Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
- Utilizar correctamente un amplio espectro del software y lenguajes de programación estadísticos, escogiendo el más apropiado para cada análisis y ser capaz de adaptarlo a nuevas necesidades.
- Utilizar eficazmente la bibliografía y los recursos electrónicos para obtener información.

## Resultados de aprendizaje

1. Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otros.
2. Diseñar y llevar a cabo tests de hipótesis en los diferentes campos de aplicación estudiados.
3. Elaborar informes técnicos que expresen claramente los resultados y las conclusiones del estudio utilizando vocabulario propio del ámbito de aplicación.
4. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
5. Extraer conclusiones coherentes con el contexto experimental propio de la disciplina, a partir de los resultados obtenidos.
6. Interpretar los resultados estadísticos en contextos aplicados.
7. Justificar la elección de cada método particular dentro del contexto en que se aplica.
8. Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
9. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
10. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
11. Reconocer la importancia de los métodos estadísticos estudiados dentro de cada aplicación particular.
12. Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
13. Utilizar distintos programas (tanto libres como comerciales) asociados a las distintas ramas aplicadas.
14. Utilizar eficazmente bibliografía y recursos electrónicos para obtener información.

## Contenido

Estos son los contenidos propuestos\*:

1. Introducción al curso

1.1 Introducción a R markdown: creación de informes automatizados y reproducibles

1.2 Tidyverse

2. Análisis de tasas de incidencia y mortalidad

a. Introducción

b. Definición de tasa

c. Cálculo de tasas estandarizadas

d. Análisis de tendencias temporales

e. Modelos de regresión 'jointpoint'

f. Modelos edad-periodo-cohorte

3. Análisis de supervivencia para datos con eventos recurrentes

a. Introducción

b. Modelos no paramétricos

i. Modelo Peña-Strawderman-Hollander

ii. Modelo de Chan-Wang

iii. Modelo de fragilidad ('frailty model')

c. Modelos semi-paramétricos

i. Modelo condicional (Prentice-William-Peterson)

ii. Modelo marginal (Wei-Lin-Weisfeld)

iii. Modelo de fragilidad ('frailty')

iv. Modelo general (Peña-Hollander)

v. Modelo de cáncer (González-Peña-Slate)

d. Modelo con evento terminal

i. Estimación mediante verosimilitud penalizada

4. Análisis de datos longitudinales mediante modelos lineales

a. Introducción

b. Diseños con medidas repetidas

c. ANOVA de medidas repetidas

d. MANOVA

e. Modelo lineal mixto.

f. Diagnóstico del modelo

5. Análisis de datos longitudinales mediante modelos no-lineales

- a. Introducción
  - b. Inspección gráfica de los datos
  - c. Estimación de un modelo no lineal
  - d. Diagnóstico del modelo
  - e. Soluciones cuando no se cumplen las hipótesis del modelo
  - f. Selección del modelo
  - g. Modelo no lineal mixto
1. Introducción al curso
    - 1.1 Introducción a R markdown: creación de informes automatizados y reproducibles
    - 1.2 Tidyverse
  2. Análisis de tasas de incidencia y mortalidad
    - a. Introducción
    - b. Definición de tasa
    - c. Cálculo de tasas estandarizadas
    - d. Análisis de tendencias temporales
    - e. Modelos de regresión 'jointpoint'
    - f. Modelos edad-período-cohorte
  3. Análisis de supervivencia para datos con eventos recurrentes
    - a. Introducción
    - b. Modelos no paramétricos
      - i. Modelo Peña-Strawderman-Hollander
      - ii. Modelo de Chan-Wang
      - iii. Modelo de fragilidad ('frailty model')
    - c. Modelos semi-paramétricos
      - i. Modelo condicional (Prentice-William-Peterson)
      - ii. Modelo marginal (Wei-Lin-Weisfeld)
      - iii. Modelo de fragilidad ('frailty')
    - iv. Modelo general (Peña-Hollander)
    - v. Modelo de cáncer (González-Peña-Slate)
  - d. Modelo con evento terminal
    - i. Estimación mediante verosimilitud penalizada

4. Análisis de datos longitudinales mediante modelos lineales

- a. Introducción
- b. Diseños con medidas repetidas
- c. ANOVA de medidas repetidas
- d. MANOVA
- e. Modelo lineal mixto.
- f. Diagnóstico del modelo

5. Análisis de datos longitudinales mediante modelos no-lineales

- a. Introducción
- b. Inspección gráfica de los datos
- c. Estimación de un modelo no lineal
- d. Diagnóstico del modelo
- e. Soluciones cuando no se cumplen las hipótesis del modelo
- f. Selección del modelo
- g. Modelo no lineal mixto

\*A menos que las restricciones impuestas por las autoridades sanitarias obliguen a una priorización o reducción de estos contenidos.

## Metodología

### Sesiones teóricas presenciales:

En estas sesiones se presentarán los principales conceptos de cada tema, así como el análisis de datos con ejemplos reales en los que se mostrará el código necesario de R para llevar a cabo dicha tarea. Las diapositivas (creadas con R Markdown - que garantiza reproducibilidad de los resultados) incluirán los conceptos teóricos, el análisis de datos y la interpretación de resultados y las conclusiones que se extraen a partir de ellos.

### Sesiones prácticas presenciales:

En estas sesiones se plantearán unos ejercicios guiados que el alumno deberá de resolver de forma individual. Cada alumno dispondrá de un conjunto de datos individual para el mismo problema. Se generará una base de datos aleatoria para el mismo problema real de forma independiente para cada alumno. Con esta metodología se pretende que el alumno investigue y aprenda cómo analizar un conjunto de datos reales sin importar que pregunte a un compañero cómo se debe llevar a cabo, ya que cada alumno tendrá que analizar sus datos y obtener sus conclusiones a partir de sus propios resultados.

### Asistencia a seminarios

Excepcionalmente, si se dan ciertas coincidencias horarias y si los alumnos ven factible poder realizarlo, los alumnos, junto con el profesor, asistirán a algún seminario organizado por el Servei d'Estadística Aplicada de la UAB o por algún otro centro de investigación cercano a la Universidad. Esta asistencia no será obligatoria, pero será altamente recomendada por el profesor, ya que los alumnos podrían ver cómo la metodología que están aprendiendo se utiliza en estudios reales y podrían ver cómo el trabajo de un estadístico tiene una implicación crucial en la finalización de dichas investigaciones.

## Trabajo individualizado

Las soluciones de todas las prácticas que se realicen de forma presencial deberán ser entregadas al profesor. Como ya se ha comentado estas prácticas se realizarán de forma individualizada, pues cada alumno dispondrá de una base de datos personalizada sobre el mismo problema a tratar. Además de estas prácticas presenciales, el alumno tendrá que resolver otras cuatro prácticas en casa y entregar la solución numérica, así como el código de R utilizado para obtener dichos resultados. Tanto las prácticas presenciales como las que se realicen en casa formarán parte de su evaluación continuada.

El alumno tendrá acceso a todo el material didáctico a partir de una web docente en la que también existirá un foro en el que plantear dudas que idealmente podrían ser resueltas por sus compañeros y que el profesor supervisará y/o solucionará, en caso de ser necesario.

NOTA: La metodología docente propuesta puede experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

## Actividades

| Título                | Horas | ECTS | Resultados de aprendizaje                     |
|-----------------------|-------|------|---|
| Tipo: Dirigidas       |       |      |   |
| Sesiones presenciales | 150   | 6    | 1, 4, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 |

## Evaluación

Preguntas de autoevaluación:

Tras cada sesión teórica se propondrán una serie de preguntas generales para evaluar si el alumno ha aprendido los conceptos básicos del tema tratado durante esa sesión.

Entrega de prácticas:

Durante el curso el alumno deberá solucionar varias prácticas en las que el profesor podrá guiar ya que se harán de forma presencial. Se deberá enviar la solución numérica así como el código de R utilizado para resolverlas.

Examen:

Los alumnos realizarán un examen presencial tipo test para evaluar si ha adquirido los conceptos teóricos y prácticos mínimos sobre la asignatura. Esta prueba contendrá preguntas conceptuales sobre los modelos tratados en clase y salidas de R similares a las obtenidas en los análisis que se han realizado durante el curso, sobre el que se realizarán preguntas de interpretación de resultados.

La nota final de la asignatura se será:

$0,5*NE + 0,5*NPA$

donde NE corresponde a la nota del examen y NPA a la nota de prácticas y preguntas de autoevaluación. Para poder hacer media se necesita que  $NE \geq 5$ .

NOTA: La evaluación propuesta puede experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

## Actividades de evaluación

| Título  | Peso | Horas | ECTS | Resultados de aprendizaje              |
|---|------|-------|------|--|
| Actividades individualizadas (prácticas y autoevaluación) | 50%  | 0     | 0    | 1, 4, 2, 3, 6, 7, 8, 9, 10, 12, 13, 14 |
| Examen presencial   | 50%  | 0     | 0    | 5, 6, 7, 11                            |

## Bibliografía

Material principal:

Todos los modelos y conceptos que se presentan en las sesiones teóricas se dispondrán en diapositivas.

El código de R para analizar los datos de estas sesiones teóricas también estará disponible para los alumnos.

Todo este material se podrá bajar de este repositorio:  
[https://github.com/isglobal-brge/TeachingMaterials/tree/master/Longitudinal\\_data\\_analysis](https://github.com/isglobal-brge/TeachingMaterials/tree/master/Longitudinal_data_analysis)

Bibliografía complementaria :

### Artículos (accesibles en la web de la asignatura)

González JR, Llorca F, Moreno V. Algunos aspectos metodológicos sobre los modelos edad-período-cohorte. Aplicación a las tasas de mortalidad por cáncer. *Gaceta Sanitaria*, 2002;16:267-273

Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine*, 2000;19:335-51

Fernandez E, Gonzalez JR, JM Borras, et al. Recent decline in cancer mortality in Catalonia (Spain). A Joint point regression analysis. *European Journal of Cancer*, 2001;37:2222-2228.

Gonzalez JR, Peña E, Slate E. Modelling intervention effects alter cancer relapses. *Statistics in Medicine*, 2005;24:3959-1975

V Rondeau, Gonzalez JR. Frailtypack: a computer program for the análisis of correlated failure time data using penalized likelihood estimation. *Computer Methods and Programs in Biomedicine*, 2005;80:154-164.

González JR, Peña E. Estimación no paramétrica de la función de supervivencia para datos con eventos recurrentes. *Revista Española de Salud Pública*, 2004;78:211-220

### Libros

Gonzalez JR. Modelling recurrent event data with application to cancer research. VDM Verlag, Saarbrken, Germany, 2009 (pdf del libro accesible en la web de la asignatura)

Therneau T and Grambsch P. Modeling Survival Data: Extending the Cox Model. Springer-Verlag, New York, 2000.

Duchateau, L and Janssen, P. The Frailty model. Springer-Verlag, New York 2008

Ritz C and Streibig JC. Nonlinear Regression with R. Use R! Springer, New York 2009

## Software

Tanto las clases teóricas como las prácticas se llevarán a cabo con R.