

**Dades Longitudinals: Temes Avançats en Ciències  
 de la Salut**

Codi: 103173

Crèdits: 6

Titulació	Tipus	Curs	Semestre
2501919 Estadística Aplicada	OT	4	0

**Professor/a de contacte**

Nom: Juan Ramón González Ruíz

Correu electrònic: JuanRamon.Gonzalez@uab.cat

**Utilització d'idiomes a l'assignatura**

Llengua vehicular majoritària: espanyol (spa)

Grup íntegre en anglès: No

Grup íntegre en català: No

Grup íntegre en espanyol: Sí

**Prerequisits**

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.

Es recomendable pero no imprescindible conocer el paquete estadístico R.

**Objectius**

**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 plantee 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

## **Competències**

- Desenvolupar estratègies d'aprenentatge autònom.
- Desenvolupar l'interès pel coneixement i la interpretació de fenòmens susceptibles de quantificació.
- Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes
- Identificar els models estadístics i de recerca operativa més adequats per a cada context i que permeten la presa de decisions.
- Identificar i seleccionar les fonts d'obtenció de dades i depurar-les per tractar-les estadísticament.
- Identificar la utilitat i la potencialitat de l'estadística en les diferents àrees de coneixement i saber aplicar-la adequadament per a extreure'n conclusions rellevants.
- Interpretar resultats, extreure conclusions i elaborar informes tècnics.
- Reconèixer la utilitat de la inferència estadística i de la recerca operativa i aplicar-les adequadament.
- Utilitzar bibliografia o eines d'Internet específiques de l'estadística i d'altres ciències afins, tant en llengua anglesa com en la llengua pròpia.
- Utilitzar correctament una bona part del programari estadístic i de recerca operativa existent, escollir el més apropiat per a cada anàlisi estadística i ser capaç d'adaptar-lo a les noves necessitats.
- Utilitzar tecnologies de la informació i de la comunicació.

## **Resultats d'aprenentatge**

1. Adaptar el programari, modificant-lo poc o lleugerament, a situacions diferents de les estudiades.
2. Basar-se en l'exploració de les dades, en l'ajust dels models i en els resultats de la inferència per a interpretar les dades de salut i extreure'n conclusions.
3. Desenvolupar estratègies d'aprenentatge autònom.
4. Desenvolupar l'interès pel coneixement i la interpretació de fenòmens susceptibles de quantificació.
5. Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes.
6. Identificar els models estadístics de la supervivència.
7. Identificar i seleccionar un disseny experimental o observacional per a l'obtenció de la informació en ciències de la salut.
8. Identificar models estadístics de prevalença i taxa.
9. Identificar models estadístics dels mecanismes de resistència a malalties.
10. Reconèixer la utilitat i la potència dels procediments estadístics generals, en vista de com s'han aplicat a les àrees de la salut, la millora genètica o la psicometria.
11. Utilitzar correctament un ampli programari estadístic, específic de la salut i la psicometria, i triar el més apropiat per a cada problema i context.
12. Utilitzar eines TIC i bibliografia per completar i millorar la formació en les àrees de la salut.
13. Utilitzar la inferència estadística per a l'anàlisi de taxes i altres indicadors d'evolució de dades biomèdiques.
14. Utilitzar tecnologies de la informació i de la comunicació.

## **Continguts**

### **1. Introducción al curso**

- a. Introducción a 'Sweave': creación de informes automatizados y reproducibles con R

### **2. Análisis de la evolución de tasas de incidencia y mortalidad**

- a. Introducción
- b. Definición de tasa
- c. Análisis de tendencias temporales
- d. Modelos de regresión 'jointpoint'

### **3. Análisis de supervivencia para datos con eventos recurrentes**

- a. Repaso de análisis de supervivencia: métodos no paramétricos, comparación de curvas, modelo de Cox
- b. Introducción al análisis de supervivencia con eventos recurrentes
- c. Modelos no paramétricos
  - i. Modelo Peña-Strawderman-Hollander
  - ii. Modelo de Chan-Wang
  - iii. Modelo de fragilidad ('frailty model')
- d. 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)
  - vi. Modelo con evento terminal
- e. 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

## **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 'Sweave' - que garantiza reproducibilidad de los resultados) incluirán los conceptos teóricos, el análisis de datos, 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 personalizado para un 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.

Una práctica se basará en resolver un problema real en grupo y realizar una exposición oral de los resultados, de forma que el alumno se habitue a expresar resultados estadísticos en una investigación biomédica.

### **Seminarios:**

Se realizaran dos seminarios. Uno de ellos servirá para introducir Latex y Sweave para aprender a redactar informes con R de forma automática y garantizar la reproducibilidad de los resultados de las prácticas que los alumnos entregan. El otro seminario se llevará a cabo a mitad del curso y en él, el profesor ilustrará un problema con datos reales y los alumnos (en grupos de 3-4 personas) deberán discutir el plan de análisis y llevarlo a cabo 'in situ' para despues presentar sus conclusiones.

### **Trabajo individualizado:**

El alumno tendrá que resolver una práctica para cada uno de los temas que se traten. Estas prácticas se realizarán en clase con la tutorización del profesor y podrán acabarse en casa. Cada alumno deberá entregar la solución numérica, así como el código de R utilizado para obtener dichos resultados. Estas prácticas 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.

## **Activitats formatives**

Títol	Hores	ECTS	Resultats d'aprenentatge
<b>Tipus: Dirigides</b>			
Seminarios	4	0,16	1, 3, 6, 7, 12
Sesiones teóricas	18	0,72	6, 7, 8, 12, 13, 14
<b>Tipus: Supervisades</b>			
Sesiones prácticas	20	0,8	1, 2, 6, 9, 13
<b>Tipus: Autònomes</b>			

Sesiones prácticas individualizadas (virtuales)	20	0,8	2, 3, 6, 7, 8, 13
Trabajo personal	88	3,52	1, 2, 3, 6, 7, 8, 9, 12, 13, 14

## Avaluació

### Entrega de prácticas:

Durante el curso el alumno deberá solucionar varias prácticas que serán guiadas por el profesor y que llevarán a cabo en el aula de informática. Estas prácticas se harán de forma presencial y el profesor solucionará los problemas que los alumnos vayan encontrando. El alumno dispondrá de 1 semana para redactar un pequeño informe con la solución numérica, así como el código de R utilizado para resolverlas. Estas prácticas estarán distribuidas a lo largo del curso y servirán para evaluar de forma continuada el aprendizaje del alumno. El resultado de estas prácticas supondrá el 60% de la nota.

### Examen:

Los alumnos realizarán un examen presencial tipo test para evaluar si ha adquirido los conceptos teóricos y prácticos explicados durante el curso. Esta prueba contendrá preguntas conceptuales sobre los modelos tratados en clase y sobre la interpretación de salidas de R similares a las obtenidas en los análisis que se han realizado durante el curso. El resultado de este test supondrá el 40% de la nota.

## Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Entrega de varias prácticas individuales	60%	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
Examen presencial tipo test	40%	0	0	2, 6, 7, 8, 9

## Bibliografia

### Material principal:

Todos los modelos y conceptos que se presentan en las sesiones teóricas se dispondrán en diapositivas con anterioridad al día de clase. El código de R para analizar los datos de estas sesiones teóricas también estará disponible para los alumnos así como los archivos de datos reales para la realización de las sesiones prácticas. Todo este material estará accesible en el Campus Virtual de la asignatura.

### Bibliografía complementaria:

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

González JR, Llorca F, Moreno V. Algunos aspectos metodológicos sobre los modelos edad-periodo-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*

GonzalezJR. 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