

**Modelos Lineales 1**

Código: 104860  
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503852 Estadística Aplicada	OB	2	1

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: Mercè Farré Cervelló

Correo electrónico: Merce.Farre@uab.cat

**Uso de idiomas**

Lengua vehicular mayoritaria: catalán (cat)

Algún grupo íntegramente en inglés: No

Algún grupo íntegramente en catalán: Sí

Algún grupo íntegramente en español: No

**Prerequisitos**

Fundamentos de estadística descriptiva e inferencial y de probabilidades, así como conocer los rudimentos de programación con el lenguaje R.

**Objetivos y contextualización**

El objetivo del curso es el estudio de la modelización y el análisis de datos mediante la teoría de los Modelos Lineales, así como las aplicaciones en diversos ámbitos (economía, salud, ingeniería, y ciencias en general). Los métodos y técnicas se introducen en base a ejemplos y se trabajan a partir de la resolución de los problemas propuestos y de prácticas de ordenador pensadas para ser ejecutadas con el lenguaje R. En primer lugar, se presenta el modelo de regresión simple porque tiene numerosas aplicaciones y porque es un buen prólogo para la comprensión del modelo múltiple. El modelo de regresión múltiple, expresado matricialmente e incluyendo algunas variantes (polinómica, con interacciones, utilizando variables regresoras ficticias, etc.), constituye la segunda parte del curso. En todos los procedimientos de modelización analizan el ajuste y la especificación correcta del modelo, la satisfacción de las hipótesis, la detección de datos "especiales" (anómalas e influyentes), y se estudian posibles soluciones cuando se detectan anomalías.

**Competencias**

- Analizar datos mediante la aplicación de métodos y técnicas estadísticas, trabajando con datos de diversas tipologías.
- Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otras personas.
- Diseñar un estudio estadístico o de investigación operativa para la resolución de un problema real.
- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- Formular hipótesis estadísticas y desarrollar estrategias para confirmarlas o refutarlas.
- 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.
- Resumir y descubrir patrones de comportamiento en la exploración de los datos.
- Seleccionar los modelos o técnicas estadísticas para aplicarlos a estudios y problemas reales, así como conocer las herramientas de validación de los mismos.
- Seleccionar y aplicar procedimientos más apropiados para la modelización estadística y el análisis de datos complejos.
- 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. Analizar datos mediante el modelo de análisis de la covarianza.
2. Analizar datos mediante el modelo de análisis de la varianza con bloques.
3. Analizar datos mediante el modelo de análisis de la varianza con factores anidados.
4. Analizar datos mediante el modelo de análisis de la varianza de uno o varios factores.
5. Analizar datos mediante técnicas de inferencia usando software estadístico.
6. Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otros.
7. Detectar y contemplar interacciones entre variables explicativas.
8. Detectar y tratar la colinealidad entre variables explicativas.
9. Emplear software estadístico para llevar a cabo el cálculo del tamaño muestral.
10. Establecer las hipótesis experimentales de la modelización.
11. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
12. Identificar fuentes de sesgo en la obtención de la información.
13. Identificar las etapas en los problemas de modelización.
14. Identificar las suposiciones estadísticas asociadas a cada procedimiento.
15. Modificar ligeramente el software existente si el modelo estadístico propuesto lo requiere.
16. Planificar estudios basados en series temporales.
17. Predecir respuestas, comparar grupos (valor causal) e identificar factores relevantes.
18. Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
19. 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.
20. 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.
21. Seleccionar las variables explicativas relevantes.
22. Utilizar diversidad de software estadístico para ajustar y validar modelos lineales y sus generalizaciones.
23. Utilizar eficazmente bibliografía y recursos electrónicos para obtener información.
24. Utilizar paquetes específicos para el diseño de experimentos.
25. Validar los modelos utilizados mediante técnicas de inferencia adecuadas.

## Contenido

### 1. El modelo de Regresión lineal simple

- Introducción a los modelos de regresión y pasos previos en la regresión simple: Exploración de los datos.
- La regresión lineal simple: Modelo, hipótesis, parámetros.
- Estimación puntual de los parámetros del modelo: Método de mínimos cuadrados. Los estimadores de máxima verosimilitud.

- Inferencia sobre los parámetros del modelo bajo las hipótesis de Gauss-Markov: Intervalos y tests.
- Intervalo de confianza para la respuesta media e intervalo de predicción de nuevas observaciones. Inferencias simultáneas en la regresión simple. Bandas de confianza y de predicción.
- Análisis de la varianza (ANOVA) del modelo de regresión simple.
- Diagnósticos del modelo: Evaluación gráfica de la linealidad y verificación de las hipótesis mediante el análisis de los residuos. Test de falta de ajuste lineal.
- Datos anómalas o influyentes.

## 2. El modelo de regresión lineal múltiple

- Pasos previos en la regresión múltiple: Exploración de los datos con herramientas de visualización multidimensional.
- Expresión matricial del modelo y los estimadores de los coeficientes. Interpretación de los coeficientes del modelo múltiple.
- Leyes de los estimadores de los coeficientes, de las predicciones y de los residuos: Aplicación de las propiedades de las matrices idempotentes.
- Inferencia en el modelo lineal múltiple. Anova del modelo.
- Test de "ligaduras" para resolver restricciones lineales sobre los coeficientes: El principio de la variabilidad incremental.
- Discusión de las hipótesis del modelo lineal: Análisis de los residuos. Transformaciones de Box-Cox.
- El problema de la multi-colinealidad entre variables regresoras: Detección y soluciones.
- Variables ficticias en regresión (*dummies*): Interpretación de los coeficientes y aplicaciones.
- Selección de variables en un modelo lineal: El estadístico Cp de Mallows, la validación cruzada de modelos y la selección automática por pasos.

## Metodología

La asignatura consta de teoría, problemas y prácticas. En la teoría se presentan y motivan las herramientas y métodos de modelos lineales y se hace un trabajo analítico. Se proporcionarán listas de problemas a lo largo del curso. Además, hay prácticas para analizar datos con el lenguaje de programación R. Se propondrán algunas tareas (ejercicios teóricos y con R) para entregar. Además de las entregas, el estudiante también realizará otro trabajo autónomo consistente en investigación bibliográfica y preparación de exámenes.

El material del curso (notas de teoría, listas de problemas y tareas de ordenador) estará disponible en el aula *moodle*.

La perspectiva de género va más allá de los contenidos de los cursos, ya que implica también una revisión de las metodologías y las interacciones entre los estudiantes y los profesores, tanto dentro como fuera del aula. En este sentido, las metodologías participativas de enseñanza que dan lugar a un entorno de igualdad, menos jerárquicas en el aula, evitando ejemplos estereotipados en el género y el vocabulario sexista, suelen ser más favorables a la plena integración y participación de las alumnas. Por ello, se hará esta aplicación efectiva durante el curso.

## Actividades

Título	Horas	ECTS	Resultados de aprendizaje
<b>Tipo: Dirigidas</b>			
Clases de teoría	26	1,04	1, 2, 3, 4, 6, 11, 10, 12, 13, 14, 21, 25
Prácticas tutorizadas	26	1,04	5, 14, 15, 21, 23, 22
<b>Tipo: Autónomas</b>			
Estudio y consultas	36	1,44	6, 11, 23
Resolución de ejercicios con R	32	1,28	1, 2, 3, 4, 5, 7, 8, 9, 12, 16, 17, 19, 20, 21, 24, 22, 25
Resolución de problemas	18	0,72	4, 10, 12, 13, 14, 18, 19, 21, 25

## Evaluación

PR: Entrega de los ejercicios teóricos y prácticos (con R). Valoración máxima de PR: 2 puntos. Esta parte no es recuperable.

P1: Prueba parcial de regresión simple (teoría, ejercicios, y prácticas). Valoración máxima de P1: 3 puntos.

P2: Prueba parcial de regresión múltiple (teoría, ejercicios y prácticas). Valoración máxima de P2: 5 puntos.

La nota de curso se calculará:  $NC = PR + P1 + P2$ . El aprobado por curso requiere que NC sea igual o mayor que 5 y que las notas de cada parcial sean mayores o iguales que 3.5 (sobre 10).

Al final del semestre se realizará un examen de recuperación que será una prueba de síntesis, PS, (teoría, ejercicios y prácticas) de los contenidos de todo el curso con una puntuación máxima de 8 puntos, por los alumnos que no hayan aprobado por curso o quieran mejorar la nota. Sólo se podrán presentar a la prueba de síntesis los estudiantes que hayan participado en 2/3 de las actividades de evaluación.

La nota final de los presentados a la prueba de síntesis se calculará:  $NF = PR + \max(PS, P1 + P2)$ .

Las matrículas de honor que eventualmente se concedan a partir de la NC no se retirarán incluso si otro estudiante obtiene una nota superior después de la PS.

Atención: "Sin perjuicio de otras medidas disciplinarias que se estimen oportunas, y de acuerdo con la normativa académica vigente, se calificarán con un cero las irregularidades cometidas por el estudiante que puedan conducir a una variación de la calificación de un acto de evaluación. Por lo tanto, plagiar, copiar o dejar copiar una práctica o cualquier otra actividad de evaluación implicará suspender con un cero y no se podrá recuperar en el mismo curso académico. Si esta actividad tiene una nota mínima asociada, entonces la asignatura quedará suspendida."

## Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entrega de tareas (problemas y prácticas resueltas)	20%	0	0	2, 3, 4, 5, 6, 11, 7, 8, 9, 15, 17, 18, 19, 20, 21, 23, 24, 22, 25
Examen final	80% (recupera los dos parciales)	4	0,16	4, 5, 10, 12, 13, 14, 15, 21, 22, 25
Primer parcial	30%	4	0,16	2, 4, 5, 10, 9, 14, 17, 22, 25

Segundo parcial	50%	4	0,16	1, 2, 3, 4, 5, 7, 8, 10, 12, 14, 15, 16, 17, 21, 22, 25
-----------------	-----	---	------	---

---

## Bibliografía

Montgomery, D. Peck, A. Vining, G.; *Introduction to Linear Regression Analysis*. Wiley, 2001.

Clarke, B.R.; *Linear Models: The Theory and Applications of Variance*. Wiley, 2008.

Christopher Hay-Jahans; *An R Companion to Linear Statistical Models*. Chapman and Hall, 2012.

Fox, J. and Weisberg, S.; *An R Companion to Applied Regression*. Sage Publications, 2nd edition, 2011.

Peña, D.; *Regresión y diseño de Experimentos*. Alianza Editorial (Manuales de Ciencias Sociales), 2002.

### Bibliografía complementaria:

Sen, A., Srivastava, M.; *Regression Analysis: Theory, Methods and Applications*. Springer, 1990.

Neter, M. H. Kutner, C. J. Nachtsheim, W. Wasserman; *Applied Linear Models*. Irwin (4th edition), 1996.

Faraway, J.; *Linear Models with R*. Chapman&Hall/CRC (2nd ed), 2014.

Rao, C. R., Toutenburg, H., Shalabh, Heumann, C; *Linear Models and generalizations*. Springer, 2008.