1. Tipo de actividad: Curso de Postgrado
2. Nombre de la actividad: “Estadística con R: modelos lineales y extensiones”
3. Docente: Dr. Javier SEOANE PINILLA (Grupo de Ecología Terrestre, Departamento de Ecología, Universidad Autónoma de Madrid).
4. Fundamentación: El modelo lineal general en estadística, que comprende las técnicas tan familiares como el ANOVA y la regresión, es como el sistema astronómico de Claudio Ptolomeo: en la actualidad está denostado porque sabemos que se basa en una concepción errónea, geocéntrica, del universo. Sin embargo, su modelo geométrico permitió explicar y predecir las posiciones de los planetas durante más de mil años. De forma similar, los modelos lineales son herramientas sencillas y flexibles que se pueden usar para explicar y predecir una amplia variedad de fenómenos, aunque se basen en presunciones erróneas. Y si esto no resultara satisfactorio estos modelos permiten construir otros más complejos a partir de ellos, que extienden las capacidades de los modelos lineales. La regresión examina la relación entre tres elementos: una variable respuesta y una función lineal de otras variables explicativas, más un término que da cuenta de los errores. La generalización de la variable respuesta ¬¬a distribuciones no normales nos lleva a los Modelos Lineales Generalizados. La consideración de datos agrupados conduce a los Modelos Mixtos. La descripción de relaciones no lineales entre la variable respuesta y las explicativas se acomete con Modelos Aditivos y Árboles de Regresión/Clasificación. En este curso abordaremos estas técnicas, considerando la regresión lineal como la piedra fundacional de todas ellas.
5. Objetivos: Este doble enfoque teórico-práctico persigue que los alumnos puedan desenvolverse con confianza con el programa R, afianzar conceptos estadísticos básicos y, en conjunto, facilitar el aprendizaje de las técnicas estadísticas particulares que les resulten necesarias una vez acabado el curso. Los objetivos del curso permitirán a los estudiantes:

1) familiarizarse con técnicas para describir las hipótesis de investigación en modelos estadísticos verificables con datos (elaboración e interpretación de modelos de regresión linear, modelos lineales generalizados y modelos mixtos), 2) facilitar el autoaprendizaje posterior de otras técnicas estadísticas más particulares y 3) entrenar el uso del lenguaje R para abordar los aspectos anteriores.

1. Arancel: $800 Docentes de la UNLPam

 $ 1000 Estudiantes de postgrado

 $ 1200 Profesionales

1. Modalidad: Presencial (40 hrs).
2. Programa:

Se impartirá clases en 6 jornadas consecutivas de lunes a sábado. Cada jornada consta de dos sesiones de 2,5 a 3,5 horas de duración. Cada sesión se dividirá en dos partes. En la primera parte (‘ejercicios guiados’) se impartirá la teoría a través de ejercicios guiados. En la segunda parte (‘trabajo en grupo’) se propondrán ejercicios a realizar en grupos pequeños con revisión por pares. Estas dos partes transcurrirán de la siguiente manera:

Primera parte (‘ejercicios guiados’):

* Presentación de conceptos teóricos de una forma muy práctica, insistiendo en explorarlos con código en R que se dispondrán en formato de acceso libre (.html o .pdf). Se reducirán al máximo las presentaciones con ‘diapositivas’ (tipo .pptx).
* Realización de ‘ejercicios guiados’, con los que los alumnos, de forma individual, irán siguiendo las instrucciones del profesor y escribiendo código (que podrán copiar de los documentos que se proporcionarán).

Segunda parte (‘trabajo en grupo’):

Se desarrollará de la siguiente forma:

* [5’] Se presenta una pregunta de investigación que requiere para resolverse alguno de los métodos explicados en la sesión de mañana.
* [30’] Trabajo en grupo pequeño (2 – 4 personas, pero preferiblemente en parejas) para resolver la pregunta. Los grupos deberán seleccionar el método adecuado, elegir una estrategia de resolución y desarrollarla con una sintaxis de R correcta. El profesor irá dando pistas haciendo referencia a los ejercicios guiados resueltos por la mañana.
* [20’] Discusión entre colegas y revisión. Cada grupo presenta y justifica al grupo vecino su respuesta. Después los grupos pueden corregir su respuesta de acuerdo a esta interacción.
* [25’] Debate abierto en grupo grande (para toda el aula) dirigido por el profesor. Se proponen, evalúan y critican las distintas respuestas.
* [10’] Descanso.

Esta metodología pretende motivar a los alumnos y facilitar un aprendizaje activo que genere un conocimiento de larga duración. En suma, con esta metodología se conseguirá que este curso intensivo resulte lo más útil posible.

Contenidos:

**Jornada 1 (5 horas). Introducción al lenguaje R.**

* Primera sesión: Ejercicios guiados (3h) y Trabajo en grupo (2h)

• ¿Qué es el lenguaje/programa R? ¿Y Rstudio?

• Introducción de datos

• Operaciones matemáticas y de manejo de datos básicas.

• Manejo de datos (con la sintaxis del ‘tydiverse’)

• Bucles y funciones.

**Jornada 2 (7 horas). Modelos lineares (regresión).**

* Primera sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Regresión linear simple y múltiple.

• Análisis de la covarianza (ANCOVA)

• Análisis de la varianza (ANOVA)

* Segunda sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Selección de modelos mediante LRT y AIC.

• Diagnósticos numéricos y gráficos de la regresión.

• Regresión parcial y partición de la varianza.

**Jornada 3 (7 horas). Modelos lineares generalizados**

* Primera sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Introducción a los GLM.

• Regresión de Poisson (y binomial negativa).

* Segunda sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Regresión logística.

• Regresión binomial.

**Jornada 4 (7 horas). Modelos mixtos/jerárquicos/multinivel (I)**

* Primera sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Modelos mixtos para análisis encajados: introducción

* Segunda sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Estimas, p-valores, diagnóstico y predicción en modelos mixtos.

• Protocolo para la elaboración de modelos mixtos

**Jornada 5 (7 horas). Modelos mixtos/jerárquicos/multinivel (II)**

* Primera sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Modelos mixtos para medidas repetidas y diseños cruzados.

* Segunda sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Modelos generalizados mixtos (GLMM) y aditivos mixtos (GAMM).

**Jornada 6 (7 horas). Modelos para datos con sobreabundancia de ceros**

* Primera sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Modelos inflados por ceros (‘zero-inflated-‘).

* Segunda sesión: Ejercicios guiados (2h) y Trabajo en grupo (1,5h)

• Modelos alterados por ceros (‘zero-altered-‘).

Bibliografía (recomendada para profundizar en los aspectos teóricos y prácticos)

De introducción breve y estadística básica:

- Daalgard, P. 2008 (2nd ed.). Introductory Statistics with R. Springer.

- Zuur, A.F., Ieno, E. I., Meesters, E.H.W.G. 2009. A Beginner's Guide to R. Springer.

Conceptos estadísticos avanzados:

- Fox, J. 2011 (2nd ed.). An R Companion to Applied Regression. Sage.

- Zuur, A.F., Ieno, E. I., Smith, G. 2007. Analyzing Ecological Data. Springer.

- Zuur, A.F. 2012. Beginner's Guide to Generalized Additive Models with R. Highland Statistics.

- Zuur A.F., Hilbe J.M., Ieno E.N. 2013. Beginner's Guide to GLM and GLMM with R. Highland Statistics.

- Zuur A.F, Ieno E.N. 2016. Beginner's Guide to Zero-Inflated Models with R. Highland Statistics.

- Faraway, J. 2016 (2nd ed.). Extending the linear model with R. CRC press.

1. Fecha de inicio y (posible) finalización: 9 al 13 de Julio de 2018.
2. Carga horaria: 6 clases teórico-prácticas. Carga horaria total: 40 horas (una jornada de 5 horas y cinco jornadas de 7 horas). El horario tentativo es: jornada 1 (lunes) 9:00 –14:00, jornadas 2 a 6 (martes a sábado) 9:00 – 12:30 y 14:00 – 17:30
3. Destinatarios: (1) Este curso está dirigido a licenciados, ingenieros o postgraduados tanto en Biología como de Recursos Naturales, Ciencias Ambientales y Ciencias Exactas u otras ramas relacionadas con las ciencias naturales y en general, a profesionales (Agentes del medio ambiente, personal de organismos oficiales o ONG, etc.) con proyectos de ecología o biología ambiental, investigación y conservación de la naturaleza, que necesiten actualizar sus conocimientos en bioestadística y utilizar un programa estadístico (preferiblemente R) para desempeñar sus labores. (2) Para un buen aprovechamiento del curso los alumnos deberán tener experiencia anterior en estadística básica (estadística descriptiva y test de hipótesis), puesto que los conceptos teóricos se impartirán solo brevemente.
4. Cupo: 20 alumnos.
5. Requisitos de aprobación: Asistencia mínima de 80% de las clases.