

# MÉTODOS ESTADÍSTICOS AVANZADOS EN ECOLOGÍA Y EVOLUCIÓN

## PROYECTO FINAL DE ANÁLISIS DE DATOS 2020.

Este examen consta de 3 preguntas. Resuelva cada pregunta del examen en este archivo. Copie y pegue en este archivo el output de R (tablas, gráficos, etc) que considere necesario para avalar sus respuestas a las preguntas. Comente todo output de R que incluya; output no comentado es lo mismo que output inexistente. Guarde este archivo con el formato (Nombre.Apellido.pdf). Envíe este archivo pdf y el script R (nombre.apellido.r) a [pablo.inchausti.f@gmail.com](mailto:pablo.inchausti.f@gmail.com). Buena Suerte.

---

### PREGUNTA 1:

Se desea investigar el efecto de la humedad del aire sobre la abundancia local de una especie de su invertebrado favorito en 4 sitios (a,b,c,d). Para ello se estimó la abundancia de esta especie empleando un método mágico que permite garantizar que la probabilidad de observación de un individuo presente en un micrositio es alta y constante durante el período de observación. ¿Qué se concluye?

### PREGUNTA 2:

Para evaluar el efecto de la fertilidad (nitrato añadido; 0,50, 100, 150 en unidades convenientes) y del genotipo (hay 3 genotipos) sobre el peso de las semillas (en mg) de su planta de cultivo favorita, se sembraron un número de plantas en pequeñas parcelas heptagonales idénticas. Se midió la variable de respuesta tras un período adecuado de crecimiento. El ensayo se realizó en 4 grandes áreas homogéneas en lo que respecta a la fertilidad edáfica. Dentro de cada una de estas grandes áreas, se definieron 12 pequeñas parcelas heptagonales en las que se sembraron plantas de los 3 genotipos. En cada una de estas parcelas pequeñas luego se aplicó uno de los niveles de nitrato. ¿Qué se concluye?

### PREGUNTA 3:

Datos (simulados) de 8525 pacientes oncológicos son empleados para establecer las variables determinantes en la Cura (var respuesta) de los mismos en un lapso de tiempo dado. Las variables a evaluar son el contenido de interleucina (IL-6) y de proteína reactiva P (CRP; ambos son bio-indicadores de inflamación), la edad y condición de fumador de cada paciente (olvide las otras variables del archivo pues ellas serán analizadas por su primo). Cada fila de la matriz de datos corresponde a un paciente, hay más de un paciente por médico (DoctorID; total de 407 médicos) y hospital (HospitalID; total de 35 hospitales), pero cada médico trabaja en un solo hospital. ¿Qué se concluye?

### Para cada pregunta:

- Identifique la variable respuesta. ¿De que tipo es? ¿Cuál sería su potencial distribución de probabilidades? Verifique la distribución de probabilidades de la variable de respuesta y explique por qué es importante realizar esta verificación. Identifique las variables explicativas e indique el tipo, número de niveles, condición de fijas o aleatorias, cruzadas o anidadas. Analice y discuta el diseño experimental empleado en relación a las variables a modelar como efecto aleatorio (si es que las hay...). Justifique brevemente su respuesta.
- Describa las grandes tendencias de variación de la variable de respuesta en relación con la(s) variable(s) explicativa(s) utilizando los gráficos y/o tablas que considere necesarios.
- Escriba el modelo estadístico que va a ajustar. Justifique brevemente su respuesta.

- d) ¿Qué distribuciones previas se van a emplear para cada parámetro? Justifique brevemente su respuesta.
- e) Ajuste el modelo que considere más adecuado. Comente el significado “biológico” de los parámetros estimados y sus intervalos de confianza. Utilice los gráficos y/o tablas que considere necesarios.
- f) Haga la validación exhaustiva del modelo ajustado. Nuevamente, utilice los gráficos y/o tablas que considere necesarios y coméntelos brevemente.
- g) Si ajustara (no es obligatorio hacerlo) más de un modelo estadístico para responder a la pregunta, explique cómo se realizó la selección de modelos.
- h) Interprete los resultados del modelo ajustado en términos de la pregunta formulada.