

# Tarea Módulo 3

Julieta Rosell

Entregar 29 de abril de 2024

=====

Deberán entregar un archivo en word con las respuestas a las preguntas que se incluyen en este documento. LA TAREA ES POR EQUIPO Y SE ENTREGA EN EQUIPO. NO SE RECIBIRÁN TAREAS INDIVIDUALES. Deberán enviar su tarea por equipo a [julieta.rosell@iecologia.unam.mx](mailto:julieta.rosell@iecologia.unam.mx) Entreguen por favor un documento de word con toda la información necesaria. También envíen el script como respaldo pero no ubiquen ahí sus respuestas centrales a las preguntas de la tarea. Consultaré el script únicamente para revisar cómo hicieron las cosas, no para calificar directamente ese archivo.

Cuiden que su tarea no tenga faltas de ortografía (se tomarán en cuenta para la calificación) y que sus gráficas tengan nombres y unidades en los ejes. No olviden verificar los supuestos de las pruebas. Para los supuestos, reporten la Ho y Ha correspondientes y el p-value (no es necesario hacer la verificación de supuestos a mano, la pueden obtener de R).

1.En un ejido en Veracruz un grupo de familias ha montado un sistema de crianza de tilapias y desean conocer el efecto de la densidad de peces en el encierro y de la estación del año en el crecimiento de los individuos. Para probar el efecto de estos factores, realizan un experimento en encierros en donde colocan 10, 18 o 24 individuos (niveles dentro del factor densidad). Pesan 9 peces antes y después del experimento (marcándolos para recapturar al mismo pez) y registran el incremento en peso al cabo de dos semanas. Realizan estas mediciones en verano y en primavera (factor estación del año). ¿Cuál es el efecto de la época del año y de la densidad de peces en el encierro en el crecimiento de las tilapias? Utiliza todos los recursos vistos en clase para presentar los resultados de este ejercicio. NOTA: Las mediciones en cada estación vienen de peces distintos.

2.Analiza los mismos datos del ejercicio anterior pero probando el efecto de la estación y de la densidad por separado. ¿Qué diferencia encuentras en la interpretación de los resultados? Dado el diseño experimental y la pregunta central, ¿qué tipo de análisis es el más conveniente y por qué? No te olvides de revisar los supuestos.

3.Descarga los datos toluca2.csv de la página del curso. Estos datos provienen de un estudio del maíz en un área periurbana. Aunque hay más variables en la base, para este ejercicio queremos construir un modelo que permita predecir la producción de 2009 (variable `produccion_2009`) a partir de las hectáreas sembradas (`ha_maiz`). ¿Qué tan bueno es este modelo para realizar predicciones? Utiliza todos los recursos vistos en clase para contestar. NOTA. Elimina el dato con producción de 2009 igual a cero, pues el entrevistador ha reportado que se trata de un error de captura.

4.Utilizando de nuevo los datos del estudio de maíz periurbano, realiza una comparación entre la producción del 2009 entre las diferentes comunidades. Sabemos que la producción está asociada con el área sembrada, por lo que quisiéramos tomar esto en cuenta en nuestro análisis al comparar las comunidades, i.e. queremos incorporar la covariable hectáreas sembradas de maíz (`ha_maiz`). ¿Hay diferencias en producción entre los sitios tomando en cuenta las hectáreas sembradas de maíz? Aunque no se detecte una diferencia significativa entre comunidades, ¿cuál sería la ecuación de regresión para cada comunidad (la finalidad es ejercitar el uso de las variables dummy)? Genera la gráfica correspondiente con tres rectas, una para cada comunidad a partir de las ecuaciones que generaste para cada comunidad.

5. Entender qué aspectos del ambiente y de las actividades humanas afectan la abundancia de organismos es un aspecto muy importante para la conservación de áreas naturales. En 1987 se colectaron datos de aves en 56 parches de vegetación natural en Australia. En este estudio se registró la abundancia de aves (abundancia) y algunas variables que serán utilizadas como predictoras de esta abundancia, incluyendo el área del parche medido (area), el tiempo en años en que dicho parche ha quedado aislado del resto de la vegetación natural (anos.aislam), la distancia al parche de vegetación más cercano (dist), la distancia al parche más grande de vegetación en el área (dist.parche.grande), la cantidad de ganado presente en el parche (ganado, medido de 1 a 5 donde 1 es poco ganado y 5 es abundante ganado), y la altitud. Se tienen en total seis variables predictoras. OJO: Revisa que exista una relación lineal entre la variable de respuesta y las predictoras. Es posible que algunas requieran una transformación. Realiza lo siguiente:

5.1. Ajusta un modelo que permita predecir la abundancia de aves a partir de todas las seis variables predictoras. Interpreta los resultados (ajuste, pendientes, etc.)

5.2. Utiliza un procedimiento de eliminación backward comenzando con el modelo saturado (con las seis variables) para proponer un modelo múltiple que prediga la abundancia a partir de un subconjunto de las seis variables predictoras

5.3. Utiliza un procedimiento de selección forward comenzando con un modelo que solamente contenga al intercepto para proponer un modelo múltiple que prediga la abundancia a partir de un subconjunto de las seis variables predictoras

5.4. Utiliza el procedimiento de selección de modelos basado en el AIC para seleccionar un grupo de modelos que predigan la abundancia a partir de un subconjunto de las variables predictoras. Puedes acudir a Burnham et al. 2011 y a Symonds & Moussalli 2011 y al ejercicio que realizamos en clase para revisar de nuevo esta aproximación. Para hacer el ejercicio más sencillo (es decir, para reducir el número de modelos a explorar), utiliza solamente el siguiente conjunto de cuatro variables predictoras en lugar de las seis de los modelos de los incisos previos: área, ganado, distancia al parche más grande y altitud. Explora todos los modelos posibles para estas cuatro variables y realiza las comparaciones que sean necesarias a través del valor de AIC de los modelos. ¿Cuál sería el conjunto que propondrías como buenos modelos? ¿Cuál es la importancia relativa de las diferentes variables tomando en cuenta su aparición en los diferentes modelos candidatos?

5.5. Compara los resultados obtenidos con los diferentes métodos de selección de modelos. ¿A qué conclusión llegamos que conteste la pregunta principal de investigación?