

Biometría II - 2024

TP N° 4

Modelos con más de una variable predictora

Objetivos

1. Ajustar modelos con más de una variable explicativa, sean de naturaleza continua, categórica o con ambos tipos
2. Interpretar en contexto el concepto de interacción entre dos variables categóricas y entre una variable categórica y una continua
3. Aplicar procesos de selección de modelos

Problema 1. Aclimatación al estrés hídrico en *Eucalyptus globulus*

En los sistemas forestales, luego de la plantación, múltiples factores de estrés abiótico y biótico afectan la supervivencia y el crecimiento de los árboles. El estrés hídrico es el fenómeno más frecuente. Es posible mejorar la eficiencia de una plantación, aplicando prácticas de aclimatación de las plantas en vivero. Una de ellas, es la técnica de restricción de riego que permite aclimatar a las plantas a las condiciones de sequía y se aplica al final del ciclo de producción de las plantas en vivero. Asimismo, se cree que la fertilización con potasio podría favorecer el proceso. El objetivo de este trabajo fue evaluar el comportamiento de plántulas *Eucalyptus globulus* sometidas a tres regímenes de riego y a dos niveles de fertilización potásica al final del ciclo de producción en vivero. Se trabajó con plantas de 8 meses de edad, en envases de 3 litros, utilizando como sustrato tierra orgánica. Las plantas fueron asignadas al azar a tres regímenes hídricos: Riego diario (sin estrés hídrico), Riego cada 3 días (estrés moderado) o Riego cada 6 días (estrés alto) y a dos niveles de fertilización potásica: con fertilización potásica (1,8 g/l) y sin fertilización potásica. Al cabo de tres meses de tratamiento en vivero, se evaluó el grado de aclimatación sometiendo a todas las plantas a un período de sequía de 9 días, al cabo de los cuales se midió el contenido relativo de agua (CRA %). El CRA es el contenido porcentual de agua de la planta en relación al contenido de agua a hidratación máxima y constituye una medida del estado hídrico de una planta (cuanto mayor, mejor el estado de la planta).

Los datos se encuentran en el archivo CRA.csv, contiene 30 observaciones (filas) y 3 variables (columnas).

Diccionario de variables:

“envase”: número de envase en el que se coloca la plántula, *integer*.

“riego”: frecuencia de riego 1: diario, 3: cada 3 días, 6: cada 6 días, *integer*.

“fertilización”: nivel de fertilización potásica, SI (1,8 g/l), NO (sin K). *character*.

“cra”: contenido relativo de agua (%), *integer*.

1. Identifique la unidad experimental, la variable respuesta, la o las variables explicativas (tipo y niveles), la cantidad de réplicas del ensayo y el diseño.

2. ¿En función de qué información debería haberse decidido el número de réplicas del experimento?

3. Especifique el modelo en parámetros y en términos del problema. Considere a riego como cualitativa y no olvide incluir los supuestos distribucionales. ¿Cuántos y cuáles parámetros tiene el modelo?

4. Describa estadísticamente los datos. Efectúe un gráfico de perfiles, concluya en relación a la potencial interacción entre los factores involucrados en el ensayo.

5. Evalúe los supuestos del modelo. Fundamente si tiene evidencia para rechazar alguno de ellos.

6. Ponga a prueba las hipótesis adecuadas en relación a los objetivos del ensayo. Informe la magnitud del efecto. ¿Con algún tratamiento cree que se lograron condiciones que favorezcan la adaptación al estrés hídrico? ¿Recomendaría una fertilización con K para mitigar los efectos del estrés hídrico?

7. Escriba el apartado de Metodología, Resultados y acompañe con una figura convenientemente rotulada.

8. Analice los datos considerando a la variable frecuencia de riego como cuantitativa (1, 3 y 6 días). Plantee el nuevo modelo. ¿Esperaría los mismos resultados que en el abordaje analítico anterior? Ajuste este nuevo modelo y concluya.

Problema 2. Regulación de la producción de aceites esenciales en cedrón (*Aloysia citriodora*)

El cedrón (*Aloysia citriodora*) ocupa un lugar trascendente en el mercado herbario debido a los atributos sensoriales y medicinales de su aceite esencial. La producción de aceite esencial está regulada por condiciones ambientales que generan estrés en la planta, como la competencia entre plantas de cedrón por recursos, disponibilidad de agua y radiación solar. Se cree que frente al estrés la planta destina mayor cantidad de fotoasimilados a la producción de compuestos secundarios como los aceites esenciales. Una investigadora evaluó el efecto de la competencia (sin competencia, es decir una planta por maceta, y con competencia, es decir 5 plantas por maceta) en tres genotipos correspondientes a tres localidades distintas (Buenos Aires, Mendoza y San Luis) sobre el rendimiento de aceite esencial. Para ello dispuso 30 macetas con sustrato uniforme, a las cuales les asignó al azar los tratamientos. Luego de 6 meses se midió la concentración de aceite en tallo y hojas (ml de aceite/100gr de MS). La investigadora desea responder las siguientes preguntas:

a) ¿El efecto de la densidad sobre la producción de aceites esenciales es el mismo en los tres genotipos?

b) Si el efecto de la densidad no es el mismo para los tres genotipos, para decidir qué tipo de comparación realizar (comparaciones múltiples de interacción o comparaciones de efectos simples) suponga dos escenarios distintos según los objetivos del ensayo:

i) se desea iniciar una producción comercial eligiendo la combinación de genotipo y densidad que rinda más ¿qué combinación de genotipo y densidad recomendaría? ¿Cuál es la máxima diferencia en el rendimiento (magnitud de efecto) observada entre combinaciones de genotipo y densidad?, Expresarla en términos absolutos y relativos

ii) se desea trabajar con baja densidad de plantas ¿qué genotipo le recomendaría para maximizar la producción de aceites esenciales? Informe e interprete el IC de la máxima magnitud de efecto.

Los datos se encuentran en el archivo CEDRON.cvs, contiene 30 observaciones (filas) y 4 variables (columnas).

Diccionario de variables:

“maceta”: número de maceta, *integer*.

“provincia”: provincia de procedencia, “ba” (Buenos Aires), “mza” (Mendoza), “sl” (San Luis), *character*.

“densidad”: cantidad de plantas por maceta, “1_planta”, “5_plantas”, *character*.

“aceite”: contenido de aceite en tallo y hojas (ml de aceite/100gr de MS), *numeric*.

1. Indique la unidad experimental, la variable respuesta y las variables explicativas, tipo y niveles. Indique la cantidad de réplicas del ensayo.
2. Especifique el modelo en parámetros y en términos del problema.
3. ¿Se verifican los principios de aleatoriedad, replicación y control del error? Fundamente su respuesta.
4. ¿Qué ventaja le aporta a la investigadora haber realizado un ensayo con los tratamientos en arreglo factorial en lugar de dos modelos con una sola predictora cada uno?
5. Identifique qué pruebas le permitirán responder a cada una de las preguntas de la investigadora.
6. Evalúe los supuestos del modelo. Calcule el residual para la maceta 5, especificando el cálculo. ¿Cuáles son sus unidades? ¿Cómo se interpreta?
7. ¿Cuánto vale el estimador de la varianza del modelo? ¿Cuáles son sus unidades?
8. Plantee las hipótesis y concluya en relación a las preguntas de la investigadora. Evalúe e informe la magnitud del efecto en valores absolutos y porcentuales.
9. Escriba un párrafo reportando materiales y métodos y otro reportando resultados. Acompañe con una figura convenientemente rotulada con las estimaciones del modelo.

→ Puede utilizar el script CEDRON2024.R

Problema 3. Tratamiento Electroquímico de tumores

El Tratamiento Electroquímico de tumores (TEQ) consiste en el pasaje de una corriente eléctrica continua a través de dos o más electrodos insertados localmente en el tejido tumoral. El proceso implica complejas reacciones bioquímicas que resultan en la necrosis del tejido expuesto. Se ha propuesto a los cambios extremos de pH inducidos alrededor de los electrodos (alta acidez en el ánodo y alta basicidad en el cátodo) como el principal mecanismo de destrucción tumoral (Miklavcic; 2017). Un grupo de investigadores e investigadoras de nuestro país realizan una experiencia para analizar dos configuraciones de colocación de electrodos (círculo o parábola) para frenar el desarrollo tumoral. Además se desea conocer si la efectividad de las configuraciones varía según la dosis del TEQ. Para ello, llevan a cabo un ensayo in vivo en el cual se inoculan tumores de la línea M2 (adenocarcinoma mamario murino) a 40 ratones BALB/c (1 tumor por ratón). Los ratones se dividen en forma aleatoria en 8 grupos balanceados y a cada grupo se le aplica una de las dos configuraciones de electrodos y una de las cuatro dosis de TEQ ensayadas (0, 30, 50 y

65 Coulombs). Al cabo de 20 días se determina por ecografía el volumen del tumor inoculado (en mm³).

Los datos se encuentran en el archivo TEQ.csv, contiene 40 observaciones (filas) y 4 variables (columnas).

Diccionario de variables:

“raton”: número de ratón, *integer*.

“dosis”: dosis de TEQ (Coulombs), *integer*.

“config”: configuración de colocación del electrodo, *character*.

“volumen”: volumen del tumor inoculado (mm³), *numeric*.

1. Identifique las variables explicativas, el diseño y la cantidad de réplicas.
2. Escriba el modelo en parámetros, indicando el significado de cada término en contexto. Si alguna/s variables podrían ser incluida/s de diferente manera, elija aquella que resulte en un modelo con menor cantidad de parámetros.
3. Realice un gráfico descriptivo en donde se muestre la relación entre el volumen del tumor a los 20 días y la dosis de TEQ aplicada, discriminando por tratamiento.
4. Ajuste el modelo propuesto. ¿Cuáles y cuántos parámetros deben estimarse?
5. Realice un gráfico de residuos del modelo vs valores predichos. En base al patrón observado, ¿qué supuestos evalúa y qué concluye?
6. Evalúe el supuesto de normalidad. ¿Cuál es la hipótesis nula de la prueba de hipótesis correspondiente? ¿Cuáles y cuántos datos se utilizan? Efectúe la prueba y concluya.
7. Ajuste el modelo y concluya en relación al objetivo del estudio. Interprete el coeficiente estimado para configuración*dosis (con un IC_{95%})
8. A partir del modelo ajustado, ¿puede predecir el volumen medio que tendrá un tumor sometido a la configuración círculo con una dosis de 60 Coulombs? En caso que se pueda, estimar dicho valor, en caso que no sea posible, justifique indicando la razón
9. Informe e interprete el IC 95% para la magnitud del máximo efecto entre configuraciones.
10. Escriba un párrafo reportando materiales y métodos y otro reportando resultados. Acompañe con una figura convenientemente rotulada con las estimaciones del modelo.

Problema 4. Caracterización de aguas residuales

La demanda bioquímica de oxígeno (DBO) es considerada como un buen indicador para determinar el grado de contaminación del agua; a mayor DBO menor calidad. Esta variable mide la cantidad de O₂ consumido al degradar la materia orgánica de una muestra líquida y se expresa en mg O₂/l. En el marco de un estudio sobre diseño de técnicas de depuración de aguas residuales se llevó a cabo un relevamiento inicial con el objetivo de relacionar la DBO del agua residual respecto al pH de la misma en tres tipos de fuentes distintas (A, B y C). Para ello se recolectaron 18 muestras de aguas residuales a las que se le midió la DBO.

Los datos se encuentran en el archivo AGUAS.csv, 18 observaciones (filas) y 4 variables (columnas).

Diccionario de variables:

“muestra”: muestra de agua residual, *integer*.

“pH”: pH, *numeric*.

“fuente”: tipo de fuente de procedencia de la muestra (“A”, “B” o “C”), *character*.

“DBO”: demanda bioquímica de oxígeno (O_2/I), *integer*.

1. Indique cual es la variable respuesta y la/s variables explicativas (tipo y niveles si corresponde). Escriba el modelo en parámetros y en términos del problema, indicando los supuestos distribucionales. ¿Cuántos y cuáles parámetros tiene el modelo?

2. Evalúe si existe relación entre la DBO del agua residual respecto al pH. ¿Esa relación, es la misma para todas las fuentes?

3. Escribir las ecuaciones estimadas para las 3 fuentes. Interpretar las pendientes en contexto.

4. Comparar las pendientes de las ecuaciones para las 3 fuentes

5. Comparar las 3 fuentes en relación a la DBO media, para un valor de pH promedio.

6. Si se ajusta el mismo modelo que en 2), pero centrando el pH a su valor promedio, ¿cuál es la ventaja que presenta esta variante? Estime y compare la DBO media del agua de las tres fuentes, para un valor de pH promedio.

7. Estime con una confianza del 95% el valor de DBO esperado para aguas residuales provenientes de la fuente “A” a pH=7.

8. Escriba un párrafo reportando materiales y métodos y otro reportando resultados. Incluya un gráfico.

→ Puede utilizar el script AGUAS2024.R

Problema 5. Biomasa de comunidades de ratones granívoros

En una comunidad, la biomasa de un nivel trófico intermedio (consumidores) puede estar determinada más bien por la abundancia de sus recursos (un control “desde abajo” o bottom-up) o más bien por la de sus predadores (un control “desde arriba” o top-down). A su vez, todos ellos pueden estar afectados por condicionantes y modificadores de la productividad y la estructura del sistema (como la energía y recursos recibidos por los autótrofos, o la heterogeneidad espacial). Con el objetivo de estudiar algunas variables que podrían determinar la biomasa de las comunidades de ratones granívoros en 16 localidades ubicadas en ambientes desérticos, se realizó un estudio donde se registró: 1) la producción de semillas en esos ambientes, 2) la abundancia de predadores de la comunidad de ratones granívoros, 3) las precipitaciones medias anuales (un limitante de la productividad primaria en estos ambientes), y 4) la cobertura de la vegetación perenne (un factor estructural importante para los ratones). En esas localidades se realizaron muestreos con 10 trampas de captura continua durante 5 noches registrando la biomasa total de ratones capturados (en Kg).

Los datos se encuentran en el archivo BIOMASARATONES.csv (*), 16 observaciones (filas) y 6 variables (columnas).

Diccionario de variables:

“localidad”: localidad de muestreo, *integer*.

“ratones”: biomasa de ratones total por localidad (Kg), *integer*.

“lluvia”: precipitación media anual (mm), *integer*.

“predadores”: Abundancia de predadores (nº avistados / cámara trampa) , *integer*.

“cobertura”: cobertura de la vegetación perenne (% por parcela de 10 * 10 metros relevada), *integer*.

“semillas”: producción de semillas (Kg), *integer*.

(*) adaptado a partir de un ejemplo en Crawley 2002

1- Construya una matriz de gráficos de dispersión para hacer una primera inspección de la relación entre pares de variables. Calcule el coeficiente de correlación entre pares de variables y el factor de inflación de la varianza.

2- Ajuste distintos modelos de regresión múltiple con la biomasa de ratones como variable respuesta y las variables explicativas que considere apropiadas.

3- Compare los modelos alternativos en base a distintas métricas (R^2 , AIC, etc.)

4- Concluya sobre la importancia de los factores de control de la comunidad de ratones granívoros del desierto. ¿Diría que el control es bottom-up o top-down?

→ Puede utilizar el script Biomasa2024.R

Problema 6. Asociación entre rasgos paternos y tamaño de la bola de cría en el escarabajo estercolero *Sulcophanaeus* sp

Los rasgos de los padres suelen afectar el desarrollo de los rasgos de su cría. Un mecanismo por el cual los padres pueden influir en el fenotipo de la descendencia es a través del nivel de cuidado que proporcionan. *Sulcophanaeus* sp (Coleoptera, Scarabeidae) es un escarabajo estercolero con cuidados biparentales. El macho construye una bola de cría formada por estiércol, en la que la hembra deposita un huevo, que al eclosionar se alimenta de esa masa. Se sabe que cuanto mayor es el tamaño de la bola de cría, mayor es la disponibilidad de alimento y mayor el desarrollo de la cría, pero se desconocen los factores que regulan el tamaño de las bolas de cría. Los machos presentan además variaciones en el tamaño de sus cuernos, que utilizan para pelear por las hembras y les confiere éxito reproductivo.

Se desea probar la hipótesis de que ambas características morfológicas de los machos (tamaño corporal y la longitud de los cuernos) afectan en forma sinérgica el tamaño de las bolas de cría. Para demostrarlo se capturaron ejemplares adultos en la provincia de Buenos Aires y se les midió el largo corporal total (LT) y del cuerno (LC), ambos en mm. Se seleccionaron machos de manera de cubrir el rango de combinaciones de tamaños y se los cruzó con una hembra de tamaño promedio. Cada pareja fue mantenida en una cámara de cría individual con estiércol de vaca. Se obtuvieron 75 bolas de cría, a las que se les determinó el peso seco (PS), en gramos. Los resultados se encuentran en el archivo escarab.csv

1. Explore los datos.
2. Plantee el modelo que permita poner a prueba la hipótesis de investigación.
3. Evalúe los supuestos del modelo propuesto.

4. ¿Qué cambia si centra los valores de LT y LC con respecto a sus respectivas medias?
5. Ajuste el modelo y concluya.
6. Interprete el R^2 y estudie la relación entre los valores observados de PS en bolas de cría y los predichos por el modelo.
7. Grafique el PS esperado de las bolas de cría según el LC del macho para tres tamaños de LT (media \pm DE).
8. Otro investigador desea realizar el mismo ensayo, pero quiere diferenciar entre dos especies de *Sulcophanaeus*. Escriba la nueva ecuación del modelo en parámetros indicando qué significan en el contexto de esta experiencia.
9. ¿Cómo se modificaría el modelo si el PS de cada bola se hubiese determinado tres veces. ¿Y si la experiencia se realizaba en 5 “tandas” distintas?

Bibliografía

- Ellison, A. M., N. J. Gotelli, B. D. Inouye, and D. R. Strong. 2014. [P values, hypothesis testing, and model selection: it's déjà vu all over again](#).
- Spake, R., D. E. Bowler, C. T. Callaghan, S. A. Blowes, C. P. Doncaster, L. H. Antao, S. Nakagawa, R. McElreath, and J. M. Chase. 2023. [Understanding 'it depends' in ecology: a guide to hypothesising, visualising and interpreting statistical interactions](#). Biol. Rev.
- Zuur, A. F., and E. N. Ieno. 2016. [A protocol for conducting and presenting results of regression-type analyses](#). Methods Ecol. Evol. 7: 636–645.