

Trabajo Práctico Nº 3

A - Modelado de varianza

Objetivos

1. Fortalecer los conceptos relacionados al modelado de varianza
2. Utilizar R para ajustar modelos a partir de datos con heterocedasticidad
3. Identificar el tipo de modelado de varianza más adecuado
4. Comprender las salidas de los modelos e interpretar resultados

Problema 1

Durante los últimos años, se ha desarrollado una creciente conciencia social sobre el efecto adverso que tienen los pesticidas sobre la salud ambiental y humana. Es por ello que en lugares sensibles donde se desaconseja liberar insecticidas (en forma de polvos, sprays o líquidos en el piso), se recomienda la implementación de métodos de control de insectos minimizando el uso de insecticidas y evitando su dispersión. Particularmente para el control de hormigas se recomienda la utilización de cebos alimentarios adicionados con un compuesto activo letal. Luego de varios estudios de control de hormigas hospitalarias observaron diferencias en la eficacia de los cebos: en muchos casos los cebos comerciales no resultan efectivos, pero sí lo eran cebos azucarados diluidos con la misma dosis de tóxico. En base a estas observaciones, en el Laboratorio de Insectos Sociales se propuso estudiar aspectos comportamentales de la ingestión de cebos que expliquen la menor efectividad de los cebos comerciales en el control de hormigas con respecto a cebos alternativos preparados en laboratorio.

Para cumplir este objetivo, se analizó comparativamente el comportamiento alimentario de hormigas ante estos cebos alternativos y un cebo comercial. La experiencia fue realizada con colonias de hormigas *Camponotus mus* mantenidas en el laboratorio en nidos artificiales. En una arena experimental, se suministró una gota de cebo azucarado *ad libitum*. Se trabajó con 20 hormigas elegidas al azar para cada uno de los tres tratamientos aplicados: (1) Cebo alternativo Sacarosa 30 % p/p + tóxico, (2) Cebo alternativo Sacarosa 68 % p/p + tóxico y (3) Cebo comercial con concentración de sacarosa similar a 68 % p/p + tóxico. El experimento se llevó a cabo durante 5 días, midiendo 4 hormigas diferentes de cada tratamiento por día. Se las pesó antes de ingerir y luego de la ingesta, de donde se obtuvo el peso ingerido (mg). Sabiendo la concentración del tóxico, se obtuvo la masa ingerida de tóxico (μg) y, sabiendo las densidades de las soluciones, el volumen ingerido de solución (μl). También se registró el tiempo de ingestión, tiempo de pausa y tiempo total (como la sumatoria de ambas variables), en segundos. Estas dos mediciones resultan de interés ya que se plantea que soluciones más viscosas (aquellas con mayor concentración de azúcar) requieren mayores tiempos de ingesta, dado que la hormiga bombea la solución más lentamente y además genera que ésta haga pausas para poder llenar el buche con solución que luego repartirá en la colonia. Mayores tiempos en la fuente de alimentación provocarían un menor ingreso de tóxico en el nido por unidad de tiempo, haciendo menos eficiente a ese cebo para el control de la plaga. La

velocidad o tasa de ingestión se calcula a partir del cociente entre el volumen ingerido y el tiempo de ingestión.

Los datos se encuentran en el archivo Hormigas.txt, contiene 60 observaciones (filas) y 2 variables (columnas).

→ Puede utilizar el script Hormigas2024.R

Diccionario de variables:

“Tratamiento”: tratamientos aplicados, CC, SAC30 y SAC68, character

“Dia”: Día en que se realizó la experiencia, integer

“MTox_ing”: masa ingerida de tóxico (μg), integer

“VSol_ing”: volumen ingerido de solución (μl), numeric

“T_ing”: tiempo de ingestión (seg), numeric

“T_pau”: tiempo de pausa (seg), numeric

“T_tot”: tiempo total (seg), numeric

“Nro_pau”: número de pausas, integer

Abrir el data.frame, explorar y corroborar variables y clases.

Descriptiva

1. Defina el tipo de estudio, la unidad experimental y la cantidad de réplicas por tratamiento.
2. Realice un análisis exploratorio de los datos (tablas y gráficos). Caracterice a los tratamientos aplicados, en relación a la tendencia central y dispersión de cada variable estudiada. Estudie la simetría de las variables analizadas. ¿Qué concluye en relación a la dispersión de la variable “tiempo de ingesta” para cada tratamiento?

Parte A:

¿La masa de tóxico ingerida difiere entre los tratamientos?

3. ¿Con cuál de las soluciones se ingiere mayor cantidad de tóxico? Analícelo a partir de un modelo estadístico lineal. Evalúe los supuestos. Concluya realizando recomendaciones de uso para el control de hormigas.

Parte B

Se desea conocer si el tiempo de ingesta de las hormigas difiere entre los tratamientos. Para ello:

1. ¿Cuál es la variable respuesta? ¿Cuál es la variable explicativa? ¿Cuál podría ser la potencial distribución de probabilidades de la variable respuesta?

2. Plantee el modelo en parámetros y en términos del problema. ¿Qué representan los residuos?
3. Plantee la/s hipótesis correspondientes.
4. Evalúe los supuestos.
5. Resuelva el modelo lineal utilizando la función `glm()`. Si corresponde, modele la varianza ¿Qué funciones de modelado de varianza podría utilizar?
6. Elija la/s funciones de varianza que considere adecuadas, corra los modelos con todas ellas y seleccione el mejor modelo en base al AIC.
7. Si corresponde, realice comparaciones de los tiempos de ingesta entre tratamientos.
8. Escriba los materiales y métodos del análisis de datos efectuado.
9. Concluya con respecto a las diferencias entre medias y varianzas de los tratamientos. Informe los resultados presentando un gráfico con epígrafe.

Problema 2. Efecto de la fracción lipoproteica HDL sobre la progresión de las lesiones ateroscleróticas en conejos alimentados con dieta rica en colesterol

Badimon y col llevaron a cabo un [estudio](#) a fin de determinar el efecto de la fracción lipoproteica HDL-VHDL sobre lesiones ateroscleróticas en conejos Nueva Zelanda. La aterosclerosis fue inducida por la alimentación de los animales con una dieta rica en colesterol 0,5% durante 60 días (Dieta 1). Otro grupo de animales fueron mantenidos con la misma dieta durante 90 días (Dieta 2). Un tercer grupo se alimentó con la misma dieta durante 90 días, pero recibió 50 mg de proteína HDL-VHDL por semana durante los últimos 30 días (Dieta 3). Al finalizar el ensayo los animales fueron sacrificados. En todos los casos se comprobaron lesiones aterogénicas en aorta. Se midió el contenido de colesterol en aorta (en mg/g).

Los datos se encuentran en el data.frame *Colesterol.txt*, contiene 24 observaciones (filas) y 2 variables (columnas).

Diccionario de variables:

“Dieta”: dieta a la que fue sometido cada conejo, integer (*)

“colesterol”: contenido de colesterol en aorta (en mg/g), numeric

Abrir el data.frame, explorar y corroborar variables y clases.

() si considera que no tiene sentido tratar a la variable como integer, generar una variable que se llame “Dietaf” y que sea un factor. (sugerencia no modificar la variable original)*

1. Identifique la variable respuesta y explicativa. ¿Cuál es el diseño experimental? ¿Cuál es la cantidad de réplicas?
2. Escriba el modelo en parámetros y en términos del problema, indicando los supuestos distribucionales
3. Concluya en relación al efecto de la dieta rica en colesterol y la administración adicional de la proteína HDL-VHDL. Igual que antes, antes de efectuar las pruebas evalúe los supuestos del modelo.
4. Resuelva el problema realizando la estimación por mínimos cuadrados -enfoque “clásico”, función `lm()` - y compare los resultados obtenidos.

Problema 3. Crecimiento vegetal

Un grupo de investigadores se encuentra analizando el crecimiento anual de dos especies de palmeras con fines de ser utilizadas en producción. En un primer estudio, relevan el crecimiento a lo largo de un año de 20 ejemplares de cada una de estas especies en dos regiones con diferentes características climáticas (norte y sur).

Los datos se encuentran en el archivo `dataSim.csv` contiene 80 observaciones (filas) y 3 variables (columnas).

Diccionario de variables:

“crecimiento”: crecimiento de la palmera (en cm/año), numeric

“especie”: especie de palmera, Sp1 o Sp2, character

“ubicacion”: ubicacion de la palmera, character

Abrir el `data.frame`, explorar y corroborar variables y clases.

1. Identifique la variable respuesta, la/s predictoras (tipo y niveles si corresponde) y cantidad de réplicas.
2. Escriba el modelo en parámetros explicando cada término en el contexto del problema.
3. Concluya en relación al crecimiento de las palmeras para ambas especies.
4. Resuelva el problema realizando la estimación por mínimos cuadrados -enfoque “clásico”, función `lm()` - y compare los resultados obtenidos.

Problema 4. Exposición prenatal a mercurio y actividad de la bomba de calcio

Si bien los efectos neurotóxicos de la exposición prenatal a altas dosis de mercurio están bien establecidos, poco se sabe sobre los efectos de la exposición prenatal a niveles por debajo de 10 ppm. Algunos autores han reportado deficiencias neuroconductuales sutiles, especialmente con respecto al lenguaje, atención y memoria. Se sospecha que estos efectos pueden estar mediados por la desregulación de la homeostasis del Ca a través de la inhibición de la bomba de Ca-ATPasa de membrana plasmática. Se llevó a cabo un [estudio](#) en una maternidad a fin de determinar si la exposición materna al mercurio medida por los depósitos de mercurio en el cabello (en ug/g) se relacionaba con la actividad de la bomba de Ca en el recién nacido. Para ello se seleccionaron 75 recién nacidos y sus madres. Se tomaron muestras de pelo de las madres de 3 cm desde la raíz y se determinó el contenido de mercurio (en ug/g). El cabello promedio crece aproximadamente 1 cm por mes, por lo que las muestras de cabello recogidas reflejaron las exposiciones durante los 3-4 meses previos al parto. La actividad de la bomba de Ca se midió en glóbulos rojos de sangre de cordón umbilical de los recién nacidos y se expresó en nmol/mg/hr.

Los datos se encuentran en el archivo `pelo.txt`, contiene 75 observaciones (filas) y 3 variables (columnas).

Diccionario de variables:

“X”: n° de dupla madre-hijx registrada, integer

“Hg_pelo”: contenido de mercurio en pelo de la madre (ppm), numeric

“act_bomba”: actividad de la bomba de calcio de los recién nacidos (nmol/mg/hr),
numeric

Abrir el data.frame, explorar y corroborar variables y clases.

- 1- Represente los datos en un diagrama de dispersión. ¿Existe una tendencia lineal en los datos? ¿Directa o inversa? ¿Débil o fuerte?
- 2- Estime la relación funcional. Evalúe los supuestos del modelo.
- 3- Interprete la pendiente en contexto. ¿Cuáles son sus unidades?
- 4- ¿Existe evidencia significativa de que la actividad basal de la bomba de Ca en el recién nacido disminuye linealmente con el nivel de mercurio en el pelo materno?
- 5- Sobre la base de su respuesta en (4) y el diseño de este estudio (experimental vs observacional), ¿qué puede decir con respecto a la pregunta de investigación sobre si la exposición materna al mercurio medida por depósitos de mercurio en cabello *afecta* la actividad de la bomba de Ca en recién nacidos?

Bibliografía recomendada (modelado de varianza):

- Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M. (2009). *Chapter 4: Dealing with Heterogeneity*. En: **Mixed Models Effects and Extension in Ecology with R**. Springer.
- Oddi, F. J., Miguez, F. E., Benedetti, G. G., & Garibaldi, L. A. (2020). [Cuando la variabilidad varía: Heterocedasticidad y funciones de varianza](https://doi.org/10.25260/EA.20.30.3.0.1131). *Ecología Austral*, 30(3), 438–453. <https://doi.org/10.25260/EA.20.30.3.0.1131>

B- Regresión polinomial

Problema 5. Variación de la capacidad pulmonar según la edad en personas sin enfermedad pulmonar

Se efectuó un estudio a fin de predecir la capacidad pulmonar en niños y jóvenes varones de 3 a 19 años. Se seleccionaron trescientos tres individuos abarcando el rango de edad mencionado, clínicamente sanos y sin antecedentes de enfermedad pulmonar, que fueron sometidos a una espirometría y se les determinó el volumen espiratorio forzado (VEF) en litros. Adicionalmente se registró la altura (en cm).

Los datos se encuentran en el archivo *VEF.txt* y contiene 306 observaciones (filas) y 4 variables (columnas).

Diccionario de variables:

“ID”: identificador del individuo, integer

“Edad”: edad del individuo (años), integer

“Altura”: altura del individuo (cm), integer

“VEF”: volumen espiratorio forzado (litros), numeric

Abrir el data.frame, explorar y corroborar variables y clases.

1. Realice diagramas de dispersión para analizar la relación entre las tres variables relevadas.
2. Plantee un modelo que permite explicar las variaciones de VEF en función de la edad en forma lineal. Analice el cumplimiento de los supuestos del modelo y estímelos.
3. Plantee un modelo que permite explicar las variaciones de VEF en función de la altura en forma lineal. Analice el cumplimiento de los supuestos del modelo y estímelos.
4. ¿Cuál de las dos variables predictoras analizadas le parece que sería un mejor predictor del VEF? Evalúe ajustar una regresión polinomial
5. Informe los resultados para el modelo que considere más adecuado