

Ejercicio 3

Equipo

2024-03-26

3. Modelos lineales generalizados para datos binarios

La base de datos Preg3B.csv contiene información sobre 862 insectos que fueron expuestos a diferentes dosis (Deposit en mg) de tres insecticidas (Insecticide). La asignación a una dosis y al tipo de insecticida se realizó de forma aleatoria. Después de seis días se analizó si los insectos se habían muerto, de manera que la base de datos contiene también el número de insectos muertos (Killed) y el número total de insectos expuestos (Number) por cada dosis e insecticida. Dado que se asume que el costo de los insecticidas es el mismo, el objetivo del análisis es identificar para cada insecticida qué dosis es la mínima con la que se puede indicar que el 70% de los insectos se muere, así como si considerando la menor de esas tres dosis se puede afirmar que un insecticida es el mejor comparado con el resto.

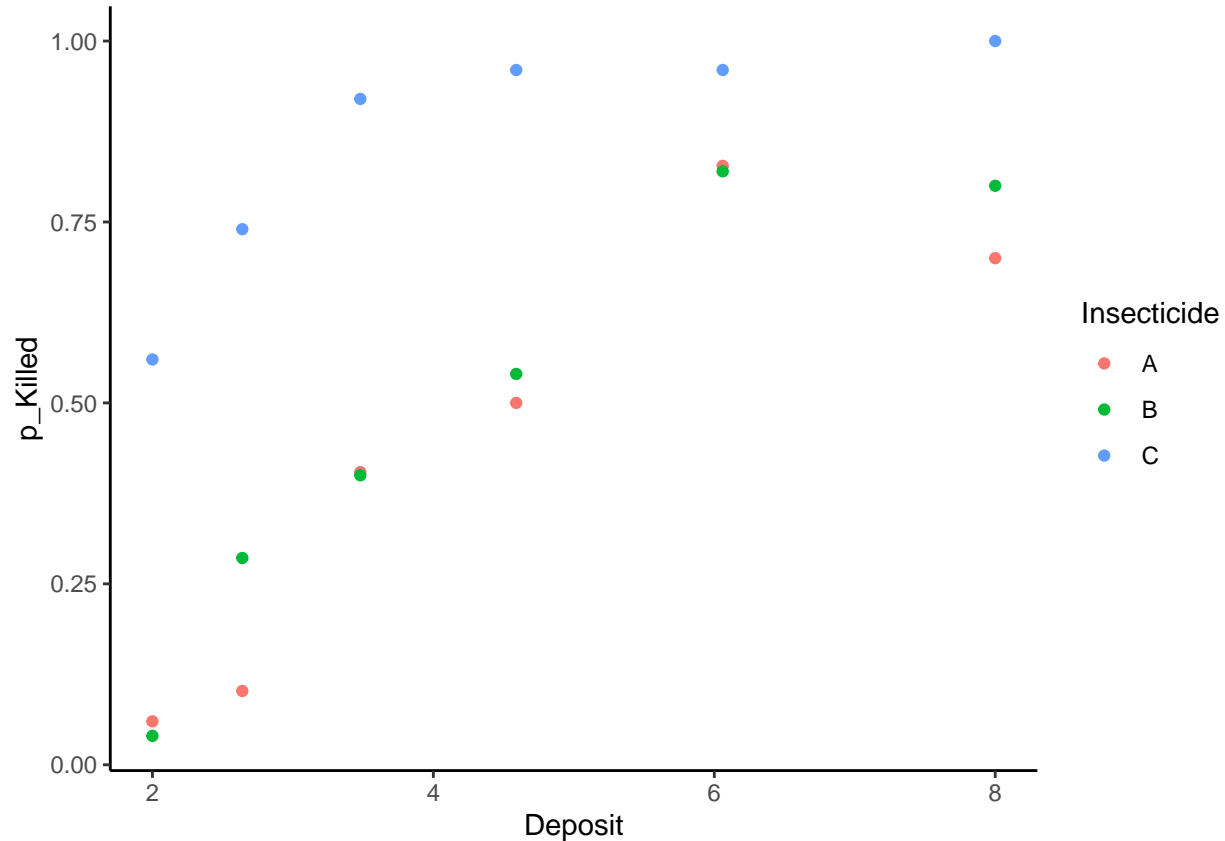
Notar que aquí el evento de interés es si el insecto muere o no. Además, dado que se tienen varios insectos para diferentes valores de dosis e insecticida, es posible realizar gráficas que ayudan a entender lo que se está modelando; de hecho la base de datos está en un formato agregado.

- i. Presente una gráfica de dispersión en donde en el eje x se incluye la dosis del insecticida y en el eje y la proporción de insectos muertos observados para cada combinación dosis-insecticida, distinguiendo con un color el insecticida asociado. Describa lo que se observa.
- ii. Ajuste modelos para datos binarios (ligas: logit, probit, cloglog) en donde incluya como covariables a Insecticide y $\ln D$ ($\ln D = \ln(\text{Deposit})$), así como su interacción. Describa las expresiones del componente lineal o sistemático para cada insecticida como función de $\ln D$. Indique si alguno de los modelos parece adecuado para realizar el análisis deseado.
- iii. Ajuste modelos para datos binarios (ligas: logit, probit, cloglog) en donde adicional a las covariables incluidas en ii), también incluya a la interacción de Insecticide con $\ln D^2$. Describa las expresiones del componente lineal o sistemático para cada insecticida como función de $\ln D$. Indique si alguno de los modelos parece adecuado para realizar el análisis deseado y si tiene alguna ventaja la inclusión de los términos cuadráticos en el modelo.
- iv. Sólo con el modelo que considere más adecuado entre los que se ajustaron en ii) y iii)
 - a. presente en la misma gráfica generada en i) los resultados de la estimación puntual para el valor esperado de la variable binaria (probabilidad de que un insecto muera).
 - b. calcule la dosis mínima para cada insecticida con la que se puede indicar que el 70% de los insectos se muere.
 - c. considerando la menor de las dosis encontradas en b), ¿se puede indicar que un insecticida es el mejor? Realice una prueba de hipótesis para argumentar en favor o en contra.
 - d. En general ¿se puede indicar que los insecticidas A y B tienen un desempeño similar? Realice una prueba de hipótesis para argumentar en favor o en contra.

i) Gráfica de dispersión de dosis del insecticida y la proporción de insectos muertos.

Se presenta una gráfica de dispersión en donde en el eje x se incluye la dosis del insecticida (Deposit) y en el eje y la proporción de insectos muertos observados (se generó la variable p_Killed) para cada combinación

dosis-insecticida (Deposit-Insecticide), distinguiendo con un color el insecticida asociado. Se puede observar que el insecticida C tiene una mayor tasa de mortalidad para todas las seis dosis consideradas (solamente la primera dosis es menor a 70%). Para el caso de los insecticidas A y B, los resultados son muy parecidos, aunque marginalmente parece que el insecticida A tiene menor tasa de mortalidad, al menos de manera evidente en tres dosis distintas.



ii) Ajuste modelos para datos binarios 1

Ajustaremos modelos para datos binarios (ligas: logit, probit, cloglog) en donde se incluyen como covariables a Insecticide y $\ln D$ ($\ln D = \ln(\text{Deposit})$), así como su interacción. Esto último implica que es el modelo completo y no uno reducido, tenemos el mismo AIC y Null Deviance, sin embargo el Residual deviance es distinto, por tener una covariable no categórica.

```
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##
## Call:
## glm(formula = mueren70 ~ Insecticide * lnD, family = binomial(link = "logit"),
##      data = datos)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.800e+02  6.173e+05      0      1
## InsecticideB    7.263e-09  8.730e+05      0      1
## InsecticideC    1.417e+02  6.748e+05      0      1
## lnD             1.684e+02  3.731e+05      0      1
```

```

## InsecticideB:lnD -4.766e-09  5.276e+05      0      1
## InsecticideC:lnD -2.529e+00  4.871e+05      0      1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2.4953e+01  on 17  degrees of freedom
## Residual deviance: 9.5953e-10  on 12  degrees of freedom
## AIC: 12
##
## Number of Fisher Scoring iterations: 25
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##
## Call:
## glm(formula = mueren70 ~ Insecticide * lnD, family = binomial(link = "probit"),
## data = datos)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.749e+01  1.144e+05 -0.001  0.999
## InsecticideB  3.330e-09  1.618e+05  0.000  1.000
## InsecticideC  3.904e+01  1.247e+05  0.000  1.000
## lnD  4.661e+01  6.911e+04  0.001  0.999
## InsecticideB:lnD -2.185e-09  9.773e+04  0.000  1.000
## InsecticideC:lnD -4.352e-01  8.957e+04  0.000  1.000
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2.4953e+01  on 17  degrees of freedom
## Residual deviance: 6.7916e-10  on 12  degrees of freedom
## AIC: 12
##
## Number of Fisher Scoring iterations: 25
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##
## Call:
## glm(formula = mueren70 ~ Insecticide * lnD, family = binomial(link = "cloglog"),
## data = datos)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.684e+02  4.388e+05  0      1
## InsecticideB -3.609e-09  6.205e+05  0      1
## InsecticideC  8.393e+01  4.552e+05  0      1
## lnD  9.524e+01  2.439e+05  0      1
## InsecticideB:lnD  2.368e-09  3.449e+05  0      1
## InsecticideC:lnD -5.028e+00  2.741e+05  0      1
##
## (Dispersion parameter for binomial family taken to be 1)

```

```

##
##      Null deviance: 2.4953e+01  on 17  degrees of freedom
## Residual deviance: 1.1755e-09  on 12  degrees of freedom
## AIC: 12
##
## Number of Fisher Scoring iterations: 25

##      Insecticide      lnD      logit      prob      cll
## 1      A 0.6931472 2.220446e-16 2.220446e-16 2.220446e-16
## 2      A 0.9707789 2.220446e-16 2.220446e-16 2.220446e-16
## 3      A 1.2470323 2.220446e-16 2.220446e-16 2.220446e-16
## 4      A 1.5238800 7.619972e-11 5.237229e-11 8.097854e-11
## 5      A 1.8017098 1.000000e+00 1.000000e+00 1.000000e+00
## 6      A 2.0794415 1.000000e+00 1.000000e+00 1.000000e+00
## 7      B 0.6931472 2.220446e-16 2.220446e-16 2.220446e-16
## 8      B 0.9707789 2.220446e-16 2.220446e-16 2.220446e-16
## 9      B 1.2470323 2.220446e-16 2.220446e-16 2.220446e-16
## 10     B 1.5238800 7.619972e-11 5.237229e-11 8.097854e-11
## 11     B 1.8017098 1.000000e+00 1.000000e+00 1.000000e+00
## 12     B 2.0794415 1.000000e+00 1.000000e+00 1.000000e+00
## 13     C 0.6931472 8.097656e-11 5.726081e-11 3.067200e-10
## 14     C 0.9707789 1.000000e+00 1.000000e+00 1.000000e+00
## 15     C 1.2470323 1.000000e+00 1.000000e+00 1.000000e+00
## 16     C 1.5238800 1.000000e+00 1.000000e+00 1.000000e+00
## 17     C 1.8017098 1.000000e+00 1.000000e+00 1.000000e+00
## 18     C 2.0794415 1.000000e+00 1.000000e+00 1.000000e+00

```

iii) Ajuste modelos para datos binarios 2

iv) Modelo adecuado. Comparaciones y prueba de hipótesis.