Ejercicio 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, 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.

3.1 Visualización de Datos

Para analizar y visualizar los datos de mejor manera, procederemos a realizar una gráfica de dispersión donde podamos ver la proporción de insectos muertos por cada combinación de dosis-insecticida.

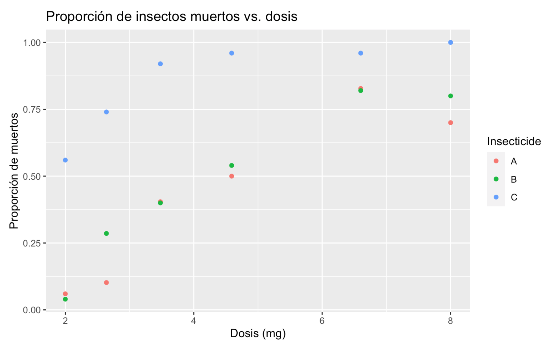

Figura 8: Proporción de insectos muertos por dosis de insecticida según su tipo (A, B o C)

Notemos que en la grafica 8, se puede observar que para todos los insecticidas (A,B y C), a mayor dosis, hay mayor proporción de insectos muertos. Sin embargo, es importante destacar cómo, con la misma cantidad de dosis para los tres tipos de insecticidas, hay una mayor proporción de insectos muertos con el insecticida tipo C; los insecticidas A y B parecen tener un comportamiento similar.

Ajuste de Modelos

3.2 Modelos con interacción

Ajustamos los tres modelos indicados en el inciso ii) con las siguientes ligas: *logit, probit y cloglog.*

Modelos:

- fit1 <- glm(cbind(Killed, Number-Killed) ~ Insecticide*log(Deposit), data = Preg3B, family = binomial(link="logit"))
- fit2 <- glm(cbind(Killed, Number-Killed) ~ Insecticide*log(Deposit), data = Preg3B, family = binomial(link="probit"))
- fit3 <- glm(cbind(Killed, Number-Killed) ~ Insecticide*log(Deposit), data = Preg3B, family = binomial(link="cloglog"))

Componente lineal según el tipo de insecticida: Notemos que ya que para todos estamos haciendo el modelo con interacciones y que tomaremos a el Insecticida A como el nivel de referencia, de forma general, el componente lineal se verá de la siguiente manera:

$$\eta = \eta(\beta, x_i) = \beta_0 + \beta_1 \mathbb{1}_{Inseticide=B} + \beta_2 \mathbb{1}_{Inseticide=C} + \beta_3 ln(D) + \beta_4 ln(D) \mathbb{1}_{Inseticide=B} + \beta_5 ln(D) \mathbb{1}_{Inseticide=C}$$

Ahora, ya que tenemos la expresión general para los tres modelos (tomando como nivel de referencia a el insecticida A), expresaremos el componente lineal para cada Insecticida en 3:

Cuadro 3: Tabla donde se muestra su componente lineal según el tipo de insecticida (A, B o C)

Insecticida	Componente Lineal
$\eta_{Insecticide=A}$	$\beta_0 + \beta_3 ln(D)$
$\eta_{Insecticide=B}$	$\beta_0 + \beta_1 + \beta_3 ln(D) + \beta_4 ln(D) = (\beta_0 + \beta_1) + ln(D)(\beta_3 + \beta_4)$
$\eta_{Insecticide=C}$	$\beta_0 + \beta_2 + \beta_3 ln(D) + \beta_5 ln(D) = (\beta_0 + \beta_2) + ln(D)(\beta_3 + \beta_5)$

Observemos que lo único que hay que hacer para obtener el componente lineal según el tipo de insecticida para cada uno de los **modelos**, es sustituir los correspondientes valores de β_i (los cuales se obtienen del *Estimate* de la función ***summary()***) en la expresión que deseen obtener en la tabla 3

Comparación de modelos: Compararemos los modelos por medio del criterio AIC y elegiremos el que tenga asociado el valor más pequeño:

Cuadro 4: Tabla comparativa de AIC

Modelo	AIC
Modelo 1: Liga Logit	96.53406
Modelo 2: Liga Probit	96.37190
Modelo 3: Liga cloglog	107.54953

Podemos ver en la tabla 4, que el menor AIC lo tiene el **modelo 2**, que es el **modelo con interacciones** y **liga *probit,*** por lo cual, pensamos que es el más adecuado.

Sin embargo, haremos el análisis de otros tres modelos para ver si encontramos uno que se adapte mejor a nuestros datos.

3.3 Modelos con la interacción anterior + Insecticide : I(ln(Deposit)^2)

Procedemos a ajustar los modelos presentados anteriormente, con la diferencia de que le agregaremos la interacción de Insecticide con (ln(Deposit))^2, es decir:

- fit4 <- glm(cbind(Killed, Number-Killed) ~ Insecticide*log(Deposit) + Insecticide:I(log(Deposit)^2), data=Preg3B, family=binomial(link="logit"))
- fit5 <- glm(cbind(Killed, Number-Killed) ~ Insecticide*log(Deposit) + Insecticide:I(log(Deposit)^2), data=Preg3B, family=binomial(link="probit"))
- fit6 <- glm(cbind(Killed, Number-Killed) ~ Insecticide*log(Deposit) + Insecticide:I(log(Deposit)^2), data=Preg3B, family=binomial(link="cloglog"))

Componente lineal según el tipo de insecticida: Sabemos que, la expresión general del componente lineal para los tres modelos (tomando como nivel de referencia a el insecticida A) es:

$$\eta = \eta(\beta, x_i) = \beta_0 + \beta_1 \mathbb{1}_{Inseticide=B} + \beta_2 \mathbb{1}_{Inseticide=C} + \beta_3 ln(D) + \beta_4 ln(D) \mathbb{1}_{Inseticide=B} + \beta_5 ln(D) \mathbb{1}_{Inseticide=C}$$

$$+\beta_6 ln(D)^2 \mathbb{1}_{Inseticide=A} + \beta_7 ln(D)^2 \mathbb{1}_{Inseticide=B} + \beta_8 ln(D)^2 \mathbb{1}_{Inseticide=C}$$

Ya que tenemos la expresión general para los tres modelos (tomando como nivel de referencia a el insecticida A), expresaremos el componente lineal para cada Insecticida en 5:

Cuadro 5: Tabla donde se muestra su componente lineal según el tipo de insecticida (A, B o C)

Insecticida	Componente Lineal
$\eta_{Insecticide=A}$	$\beta_0 + \beta_3 ln(D) + \beta_6 ln(D)^2$
$\eta_{Insecticide=B}$	$\beta_0 + \beta_1 + \beta_3 ln(D) + \beta_4 ln(D) + \beta_7 ln(D)^2 = (\beta_0 + \beta_1) + ln(D)(\beta_3 + \beta_4) + \beta_7 ln(D)^2$
$\eta_{Insecticide=C}$	$\beta_0 + \beta_2 + \beta_3 ln(D) + \beta_5 ln(D) + \beta_8 ln(D)^2 = (\beta_0 + \beta_2) + ln(D)(\beta_3 + \beta_5) + \beta_8 ln(D)^2$

Comparación y selección de modelo: Tomaremos el AIC más pequeño de los primeros 3 modelos, el cual fue el del **modelo 2**, y lo compararemos con los AIC de los **modelos 4, 5 y 6**, para ver si tuvo un efecto el haber agregado la interacción Insecticide : I(ln(Depositˆ2)) a los primeros modelos que se realizaron.

Cuadro 6: Tabla comparativa de AIC

Modelo	AIC
Modelo 2	96.37190
Modelo 4	93.70478
Modelo 5	94.01057
Modelo 6	93.14655

Vemos en el cuadro 6, que sí hubo una mejora en el AIC agregando $Insecticide : I(ln(Deposit)^2)$ a los modelos originales, por lo tanto, el mejor AIC de los 6 modelos realizados fue el **modelo 6,** y es el que ocuparemos de ahora en adelante.

Veamos si cumple los supuestos:

Verificación de Supuestos

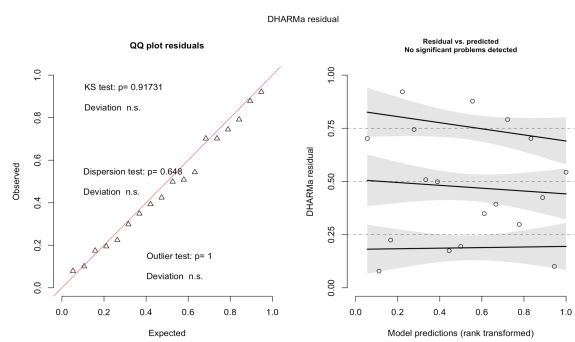

Figura 9: Graficas para analizar los supuestos del modelo 6 (fit 6)

A partir de la figura 9, parece no haber evidencia en contra de los supuestos, ya que en la primer gráfica no se encuentran problemas con el componente aleatorio y la función liga y, lo mismo para la segunda, tampoco se presentan problemas en el componente lineal.

Por lo que, podemos utilizar nuestro modelo para resolver el ***inciso iv)***

a) Observemos la siguiente gráfica:

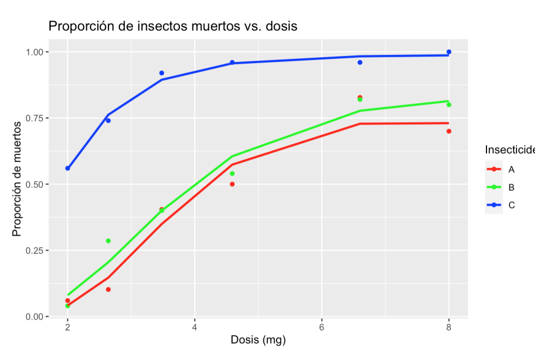

Figura 10: Estimaciones puntuales para cada insecticida (A, B y C)

Notemos que en 10, podemos ver como cómo para los tres tipos de insecticida, a mayor dosis, mayor es la proporcón de insectos muertos. También observemos que las estimaciones puntuales sugieren que el insecticida C es el más efectivo, seguido por B y luego A, donde la línea del insecticida C estima las mayores proporciones de insectos muertos en comparación con A y B.

b) Ahora, calculemos la dosis mínima para cada insecticida con la que se puede indicar que el 70 % de los insectos muere:

Recordemos que lo único que tenemos que hacer es sustituir en la liga, que en este caso es la *clolog*, $\mu_i = 0.7$, después, igualamos la liga con el componente lineal para cada insecticida que se encuentra en el cuadro 5, y finalmente, despejamos $D = Deposit$ de la ecuación.

A continuación, los resultados de la dosis mínima para cada insecticida:

Cuadro 7: Dosis mínimas para el 70 por ciento de mortalidad según el insecticida

Insecticidas	Dosis mínimas para que el 70 % de los insectos muera
Insecticida A	5.863675
Insecticida B	5.403546
Insecticida C	2.405983

Analizando la tabla 7, podemos ver como el Insecticida C es el que necesita una menor dosis para que mueran el 70 % de lo insectos, donde le sigue el Insecticida B, y finalmente el A, con la mayor dosis mínima.

c) Ya que, la menor dosis se obtuvo del Insecticida C, además que las gráficas presentadas en este problema nos muestran que, al parecer, el Insecticida C tiene un mejor desempeño sobre A y B, por lo que, realizaremos la siguiente prueba simultánea para ver si en efecto, C tiene un mejor desepeño que los otros dos insecticidas, por lo que, realizaremos la siguiente prueba:

```
         Simultaneous Tests for General Linear Hypotheses

Fit: glm(formula = cbind(Killed, Number - Killed) ~ Insecticide *
    log(Deposit) + Insecticide:I(log(Deposit)^2), family = binomial(link = "cloglog"),
    data = Preg3B)

Linear Hypotheses:
       Estimate Std. Error z value Pr(>z)
1 <= 0   2.4327     0.3307   7.357 <1e-10 ***
2 <= 0   1.9723     0.2684   7.348 <1e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```

Figura 11: Prueba de hipótesis simultáneas

Podemos ver en 11 que, en efecto, dado que en ambos casos se rechazan las hipótesis nulas (H_0), se puede concluir que el insecticida C es significativamente mejor (mata más insectos) que los insecticidas A y B, según los datos y el modelo utilizado, también los valores positivos de los estimados corroboran que el efecto de C es mayor.

d) Ya que sabemos que C tiene mejor desempaño que A y B, ¿podemos ver si A y B tiene un desempeño similar?

Realicemos la siguiente prueba de hipótesis:

```
         Simultaneous Tests for General Linear Hypotheses

Fit: glm(formula = cbind(Killed, Number - Killed) ~ Insecticide *
    log(Deposit) + Insecticide:I(log(Deposit)^2), family = binomial(link = "cloglog"),
    data = Preg3B)

Linear Hypotheses:
       Estimate Std. Error z value Pr(>|z|)
1 == 0  0.09418    0.18571   0.507    0.612
(Adjusted p values reported -- single-step method)
```

Figura 12: Prueba de hipótesis simultánea sobre contrastes definidos

Podemos ver en 12, que **no se rechaza** H_0, por lo que podemos decir que los insecticidas A y B presentan un desempeño estadísticamente similar, puesto que no hay evidencia suficiente de que los Insecticidas A y B tengan efectos diferentes (en cuanto a la variable respuesta de insectos muertos).