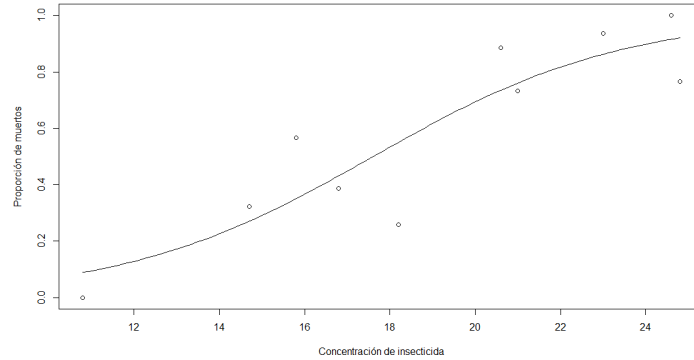


Análisis de datos categóricos

27 de febrero del 2020

La siguiente base de datos describe el efecto de diferentes concentraciones de insecticida sobre la mortalidad de los escarabajos. Se generará un Modelo Lineal Generalizado para estimar la probabilidad de morir en función de la concentración del insecticida.

	concentracion	afectados	expuestos	vivos
1	24.8	23	30	7
2	24.6	30	30	0
3	23.0	29	31	2
4	21.0	22	30	8
5	20.6	23	26	3
6	18.2	7	27	20
7	16.8	12	31	19
8	15.8	17	30	13
9	14.7	10	31	21
10	10.8	0	24	24



```
>insecto <- read_csv("C:/Users/Karla/Desktop/Categoricos/insecto.csv")
```

```
#Se estima la proporción de insectos muertos
```

```
>prop_muertos<-insecto$afectados/insecto$expuestos
```

```
>prop_muertos
```

```
[1] 0.7666667 1.0000000 0.9354839 0.7333333 0.8846154 0.2592593 0.3870968 0.5666667 0.3225806 0.0000000
```

```
#La gráfica arriba se generó así:
```

```
> plot(insecto$conc,prop_muertos, xlab="Concentración de insecticida", ylab="Proporción de muertos")
```

```
# Se va a evaluar el modelo  $\text{logit} = \log(p/(1-p)) \sim \beta_0 + \beta_1 x$ ,  $y_i \sim \text{Binomial}(p, n)$  para estimar la probabilidad de morir en función de la concentración de insecticida.
```

```
#El componente aleatorio de este modelo (la variable de respuesta) es una variable dicotómica (vivos/afectados). Por eso, vamos a modelarla con una distribución binomial.
```

```
#La función conectora en este caso es logit, ya que estamos modelando un odds. y se refiere a la manera en que va a estar relacionado este odds con la variable predictora "concentración".
```

```
>modelo1 <-glm(cbind(afectados,vivos)~concentracion,binomial(link="logit"), data=insecto)
```

```
> summary(modelo1)
```

```
Call:
```

```
glm(formula = cbind(afectados, vivos) ~ concentracion, family = binomial(link = "logit"),  
data = insecto)
```

```
Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max  
-3.0670 -1.7117  0.1495  1.7340  2.4150
```

```
Coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) -6.01047    0.77937  -7.712 1.24e-14 ***  
concentracion 0.34127    0.04153   8.217 < 2e-16 ***  
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

Null deviance: 138.001 on 9 degrees of freedom
Residual deviance: 37.697 on 8 degrees of freedom
AIC: 69.27

Number of Fisher Scoring iterations: 5

```
> b0.est<-coef(modelo1)[1]
> b1.est<-coef(modelo1)[2]
> plot(insecto$conc,prop_muertos, xlab="Concentración de insecticida", ylab="Proporción de muertos")
> curve(plogis(b0.est + b1.est*x),add=T)
#plogis lo que hace es : exp(x)/(1+ exp(x))
```

```
> plogis(b0.est + b1.est*1)
(Intercept)
0.003438757
> plogis(b0.est + b1.est*20)
(Intercept)
0.6931578
```

#El link probit también es adecuado cuando se trabaja con probabilidades. La \hat{y} está dada en z-scores

```
> modelo2<-glm(cbind(afectados,vivos)~concentracion,binomial(link="probit"), data=insecto)
> summary(modelo2)
```

Call:

```
glm(formula = cbind(afectados, vivos) ~ concentracion, family = binomial(link = "probit"),
    data = insecto)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9997	-1.6451	0.1878	1.8272	2.3810

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.57902	0.42991	-8.325	<2e-16 ***
concentracion	0.20265	0.02254	8.989	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 138.001 on 9 degrees of freedom
Residual deviance: 37.552 on 8 degrees of freedom
AIC: 69.126

Number of Fisher Scoring iterations: 5

```
> probabilidad_dosis20<-b0.est2+b1.est2*20
> probabilidad_dosis20
(Intercept)
0.4739657
> pnorm(0.47)
[1] 0.6808225
```

¿Cuál es la probabilidad de morir a una concentración de 3 mg/ml?

¿Cuánto vale la LD50? Utiliza los coeficientes estimados para los modelos 1 y 2