
Los artículos de Lindenmayer y sus colaboradores (en la bibliografía damos dos de estos artículos) proporcionan multitud de datos sobre las Montañas Centrales de Victoria en Australia. Aquí trabajaremos con datos sobre diferentes especies de marsupiales arborícolas de Bosques Montano tipo Ash (*Montane Ash Forest*).

En este estudio se analizaron 151 lugares diferentes de 3ha con vegetación uniforme, observándose en cada uno de éstos las siguientes 14 variables: La variable dependiente de respuesta, número de especies de marsupiales en el lugar (Diversidad), y las 13 covariables siguientes: el número de arbustos (Arbustos); si había, 1, o no, 0, tocones de pasadas operaciones forestales (Tocones) que es una variable cualitativa con dos niveles; el número de árboles de porte hueco (Stags); un índice de cortezas extraídas (Cortezas); un índice de habitabilidad para marsupiales (Habitat); el área de acacias (Acacias); el tipo de Eucalipto que es una variable cualitativa con tres niveles: Eucalipto regnans (Regnans), Eucalipto delegatensis (Delegatensis) y Eucaliptus nitens (Nitens); y, por último, el aspecto del lugar que es una variable de tipo cualitativo con cuatro niveles, (NWNE), (NWSE), (SESW) y (SWNW).

Se pretende ajustar un Modelo de Regresión Poisson a estos datos, primero clásico y, después, robusto.

Para los datos de Lindenmayer sobre marsupiales, que vienen recogidos en el fichero de datos `marsu` proporcionado entre el Material Didáctico del curso, se pretende ajustar un Modelo de Regresión Poisson (primero clásico) que tendrá 11 covariables, puesto que las cualitativas incorporan al modelo tantas covariables indicadoras como clases presentan menos una. Serán, 5 covariables cuantitativas, `Arbustos`, `Stags`, `Cortezas`, `Habitat` y `Acacias`, una indicador correspondiente a `Tocones`, dos covariables indicador correspondientes al tipo de Eucalipto, `Delegatensis` y `Nitens`, y tres covariables indicador correspondientes al aspecto del lugar, `NWSE`, `SESW` y `SWNW`, quedando el modelo de la forma

$$\begin{aligned} \log \text{Diversidad} = & \beta_0 + \beta_1 \text{Arbustos} + \beta_2 \text{Stags} + \beta_3 \text{Cortezas} + \beta_4 \text{Habitat} + \beta_5 \text{Acacias} \\ & + \beta_6 \text{Tocones} + \beta_7 \text{Delegatensis} + \beta_8 \text{Nitens} + \beta_9 \text{NWSE} + \beta_{10} \text{SESW} + \beta_{11} \text{SWNW} \end{aligned}$$

Como los datos a utilizar deben de estar en forma de *estructura de datos*, ejecutamos (1) para incluirlos en R^{mo} con este formato al utilizar la función `read.table`

```
> marsu<-read.table("a:\\marsu",header=T) (1)
```

Ahora, en (2), utilizamos la función glm apareciendo los resultados en (3), los cuales valoramos ejecutando (4).

```
> respu<-glm(Diversidad ~ Arbustos+Stags+Cortezas+Habitat+Acacias+ (2)
+ Tocones+Delegatensis+Nitens+NWSE+SESW+SNNW,
+ family=poisson,data=marsu)
```

```
> respu (3)
```

```
Call: glm(formula = Diversidad ~ Arbustos + Stags + Cortezas +
Habitat + Acacias + Tocones + Delegatensis + Nitens + NWSE + SESW +
SNNW, family = poisson, data = marsu)
```

Coefficients:

(Intercept)	Arbustos	Stags	Cortezas	Habitat
-0.94694	0.01192	0.04023	0.03989	0.07173
Acacias	Tocones	Delegatensis	Nitens	NWSE
0.01764	-0.27241	-0.01534	0.11492	0.06675
SESW	SNNW			
0.11695	-0.48890			

Degrees of Freedom: 150 Total (i.e. Null); 139 Residual

Null Deviance: 187.5

Residual Deviance: 118.9 AIC: 423.7

```
> summary(respu) (4)
```

Call:

```
glm(formula = Diversidad ~ Arbustos + Stags + Cortezas + Habitat +
Acacias + Tocones + Delegatensis + Nitens + NWSE + SESW +
SNNW, family = poisson, data = marsu)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.04444	-0.97981	0.05173	0.44497	1.78911

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.94694	0.26524	-3.570	0.000357 ***
Arbustos	0.01192	0.02193	0.544	0.586722
Stags	0.04023	0.01120	3.592	0.000328 ***

Cortezas	0.03989	0.01438	2.774	0.005545	**	
Habitat	0.07173	0.03812	1.882	0.059845	.	
Acacias	0.01764	0.01059	1.665	0.095835	.	
Tocones	-0.27241	0.28572	-0.953	0.340385		
Delegatensis	-0.01534	0.19149	-0.080	0.936134		(7)
Nitens	0.11492	0.27214	0.422	0.672815		
NWSE	0.06675	0.19008	0.351	0.725442		
SESW	0.11695	0.19018	0.615	0.538598		
SWNW	-0.48890	0.24710	-1.979	0.047868	*	

(5) (6)

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 187.49 on 150 degrees of freedom
Residual deviance: 118.87 on 139 degrees of freedom
(8)

AIC: 423.67

Number of Fisher Scoring iterations: 4

Los estimadores de los coeficientes aparecen en (5) y sus errores estándar en (6) (iguales ambos a los que aparecen en la corrección al artículo de Cantoni y Ronchetti en la página web de la primera) y los p-valores de los contrastes de la hipótesis nula de ser éstos cero, aparecen en (7). Éstos parecen indicar que son significativas (es decir, que deberían de utilizarse) *Stags* y *Cortezas*; con dudas, el aspecto del lugar *SWNW* y, con muchas más dudas, *Habitat* y *Acacias*. Si nos quedáramos con estas cinco covariables, el modelo de Regresión Poisson clásico ajustado se obtendría ejecutando

```
> glm(Diversidad ~ Stags+Cortezas+Habitat+Acacias+SWNW,family=poisson,
+ data=marsu)$coeff
```

(Intercept)	Stags	Cortezas	Habitat	Acacias	SWNW
-0.82125317	0.04095897	0.04064307	0.07820446	0.01363301	-0.59674721

es decir, obtendríamos el modelo

$$\log \text{Diversidad} = -0'8213 + 0'0410 \text{Stags} + 0'0406 \text{Cortezas} + 0'0782 \text{Habitat} + 0'0136 \text{Acacias} - 0'5967 \text{SWNW} \quad (1)$$

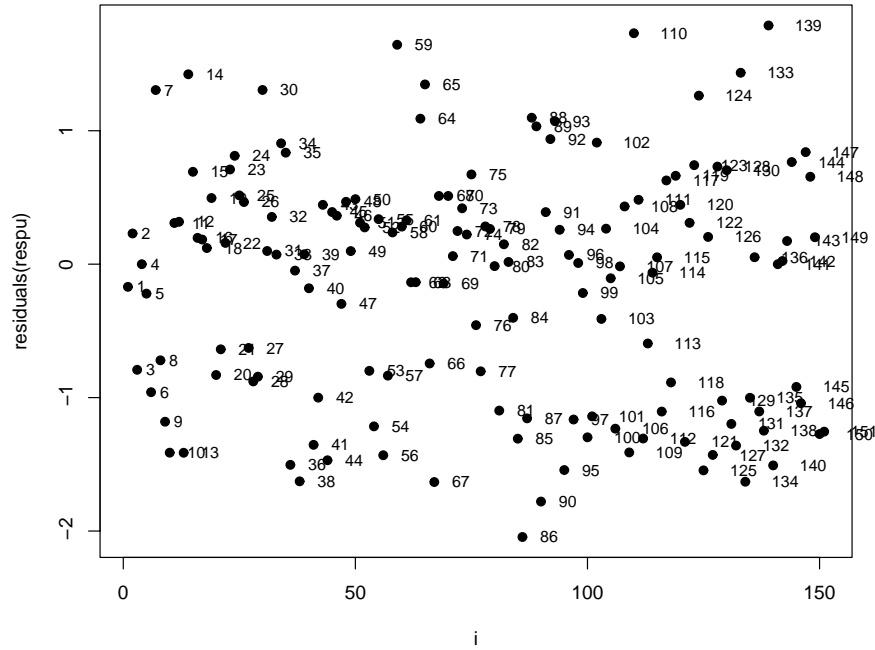


Figura 1: : Gráfico de los Residuos

el mismo (segunda columna de la tabla 5) de las correcciones al artículo de Cantoni y Ronchetti.

Obsérvese que si representamos los residuos del modelo ajustado en la Figura 1 mediante la siguiente secuencia,

```
> i<-seq(1,151)
> plot(i,residuals(respu),pch=16)
> text(i,residuals(respu),1:151,adj=-1,cex=0.8)
```

no vemos aparentemente casi ninguna observación influyente. Cantoni y Ronchetti dicen que lo son la 59, la 110, la 139 y la 133, pero esto es un tanto arriesgado. Lo que sí pone de manifiesto este ejemplo es que el método de observar, quitar las observaciones anómalas y utilizar métodos clásicos para las observaciones restantes, no es operativo. Más abajo aplicaremos a estos datos métodos robustos.

Observemos también, que hemos utilizado como variable dependiente de respuesta el número de especies marsupiales del lugar y no una tasa de éstas como hacemos habitualmente con los Modelos de Regresión Poisson. No debemos preocuparnos ya que el modelo sigue siendo válido al estar considerando, de hecho, una

tasa hipotética consistente en dividir el número observado por 10 ó 100, y hablar de número de especies de marsupiales de cada 10 ó, de cada 100. Lo importante es tenerlo en cuenta cuando si hiciéramos predicciones con el modelo ajustado.

Para determinar la Regresión Poisson Robusta, primero fijamos el valor de la constante de Huber en 1'6 que es el valor establecido en Cantoni y Ronchetti (2001). Después utilizamos la función `glm.rob` en la estimación robusta de los parámetros de la Regresión Poisson, los cuales obtenemos en (1).

```
> chuber<-1.6

> C<-glm.rob(as.matrix(marsu[,c(2,3,4,5,6,7,9,10,12,13,14)]),
+ as.matrix(marsu[,c(1)]),choice="poisson")

> C$coeff (1)
[1] -0.89780510  0.00994289 -0.25141328  0.04016733  0.03999019
[6]  0.07141413  0.01777746 -0.02022772  0.12693237  0.06009973
[11]  0.09492416 -0.50792232
```

Si aceptáramos este modelo de Regresión Poisson robusta, nos quedaría por tanto,

$$\begin{aligned} \log \text{Diversidad} = & -0'8978 + 0'0099 \text{Arbustos} + 0'0402 \text{Stags} + 0'04 \text{Cortezas} \\ & + 0'0714 \text{Habitat} + 0'0178 \text{Acacias} - 0'2514 \text{Tocones} \\ & - 0'0202 \text{Delegatensis} + 0'1269 \text{Nitens} + 0'0601 \text{NWSE} \\ & + 0'0949 \text{SESW} - 0'5079 \text{SNNW} \end{aligned}$$

que son los mismos valores que aparecen en la corrección del trabajo de Cantoni y Ronchetti.

Ahora deberíamos realizar tests condicionales para ver con qué modelo nos quedamos finalmente. Como hay muchas covariables y muchos datos, el programa da errores en algunos contrastes anidados. Si nos limitamos a ajustar el Modelo de Regresión Poisson robusto para las cinco covariables con las que nos quedamos en los métodos clásicos, ejecutaríamos

```
> glm.rob(as.matrix(marsu[,c(4,5,6,7,14)]),as.matrix(marsu[,c(1)]),
+ choice="poisson")$coeff

[1] -0.79811068  0.04057311  0.04099017  0.07762185  0.01429919 -0.60443908
```

con lo que nos quedaríamos con el Modelo de Regresión Poisson robusto,

$$\log \text{Diversidad} = -0'7981 + 0'0406 \text{Stags} + 0'0410 \text{Cortezas} + 0'0776 \text{Habitat}$$

$$+0'0143 \text{Acacias} - 0'6044 \text{SNNW}$$

el mismo obtenido en la corrección del artículo de Cantoni y Ronchetti y casi idéntico al clásico [1] como era de esperar, ya que allí comentamos que no veíamos observaciones influyentes.