


Estadística y modelación de sistemas socioecológicos en R



Laboratorio
Nacional
de Ciencias
de la Sostenibilidad

Dra. Yosune Miquelajauregui Graf

Plan del día

1. Modelos lineales generalizados (regresión Poisson y binomial negativa)
 2. Ejercicios
 3. Disusión sobre análisis de datos de tesis y seminario
- 

Modelos lineales generalizados

Regresión múltiple: hay una relación entre el valor predicho y la ecuación de regresión

$$\hat{y} = \beta_0 + \beta_{Dist} * Dist_i + \beta_{Temp} * Temp_i$$

Podemos descomponer este modelo en tres partes distintas:

1. **Componente aleatorio**: especifica la distribución de las observaciones y_i (y de los respectivos errores)
2. **Componente sistemático**: especifica la combinación de las variables explicativas o independientes que forman el predictor lineal.

$$\beta_0 + \beta_{Dist} * Dist_i + \beta_{Temp} * Temp_i$$

3. **Función de vínculo**: especifica la relación entre las observaciones originales y el predictor lineal, es decir, el predictor lineal arroja resultados en la misma escala

Modelos lineales generalizados

Los modelos lineales generalizados están constituidos por estas mismas partes

1. **Componente aleatorio**: especifica la distribución de las observaciones y_i (y de los respectivos errores): distribución normal, Poisson, binomial, gamma
2. **Componente sistemático**: especifica la combinación de las variables explicativas o independientes que forman el predictor lineal.

$$\beta_0 + \beta_{Dist} * Dist_i + \beta_{Temp} * Temp_i$$

3. **Función de vínculo**: especifica la relación entre las observaciones originales y el predictor lineal.

$$g(\mu_i) = \eta_i \text{ donde } \mu_i \text{ es la media esperada de las observaciones } y_i$$

Modelos lineales generalizados

Función de vínculo:

$$g(\mu_i) = \eta_i$$

Indica la transformación necesaria para convertir el predictor lineal en valores predichos en la misma escala que las observaciones originales.

La función de vínculo depende de la distribución de nuestros datos.

Vínculo canónico: función de vínculo “naturalmente” asociada a una distribución.

Cada modelo lineal generalizado posee una estructura particular de función de verosimilitud, la cual depende de la distribución.

Modelos lineales generalizados

Función de vínculo:

Cada distribución posee una función de vínculo

Distribución	Naturaleza de y_i	Vínculo canónico	Función de vínculo $g(\mu_i) = \eta_i$	Función inversa $\mu_i = g^{-1}(\eta_i)$
Normal	$(-\infty, \infty)$	identidad	μ_i	η_i
Poisson	0,1,2,3..	log	$\log_e \mu_i$	e^{η_i}
Binomial	0,1,2,3..n/n	logit	$\log_e \frac{\mu_i}{1 - \mu_i}$	$\frac{e^{\eta_i}}{1 + e^{\eta_i}}$

Modelos lineales generalizados

Grande familia de análisis agrupados según la distribución de las observaciones:

Distribución normal : **Regresión lineal simple**
Regresión lineal múltiple
ANOVA
Análisis de covarianza

Otras distribuciones : **Regresión de Poisson**
Regresión logística
Regresión binomial negativa
Regresión de gamma

Regresión de Poisson

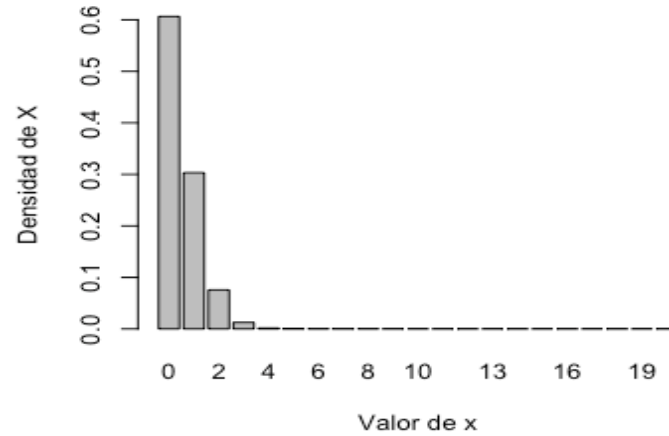
Distribución de Poisson

Características:

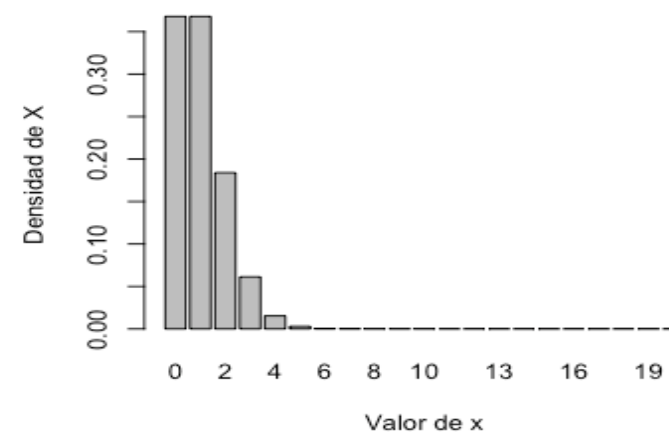
1. Distribución discreta (valores 0,1,2,3..)
2. Caracteriza los eventos raros (muchos ceros)
3. Distribución sesgada a la derecha (cola larga)
4. Se vuelve más simétrica con mientras λ aumenta

Regresión de Poisson

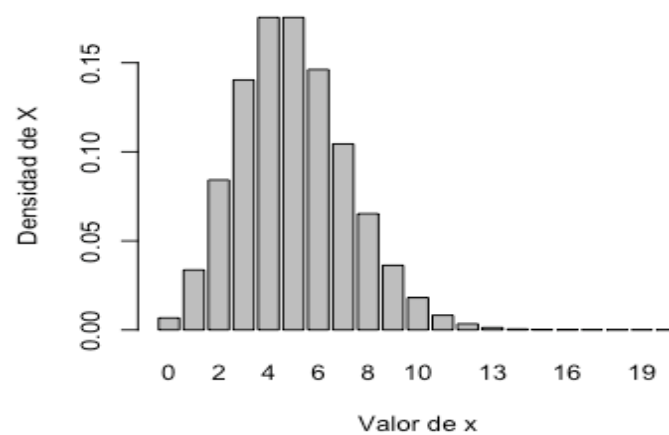
Distribución de Poisson con $\lambda = 0.5$



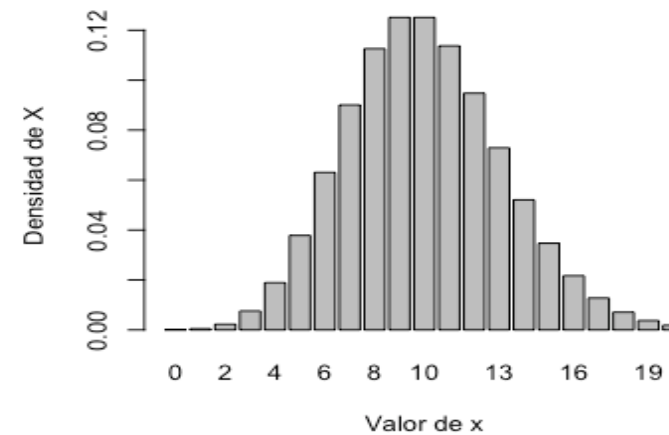
Distribución de Poisson con $\lambda = 1$



Distribución de Poisson con $\lambda = 5$



Distribución de Poisson con $\lambda = 10$



Regresión de Poisson

y_i es una variable discreta

Función de vínculo: log

Relación entre y_i y η_i (log)

Donde: $\log_e(\mu_i) = \beta_0 + \beta_1 * X1_i + \beta_2 * X2_i + \dots \beta_k * Xk_i$

y $\mu_i = e^{\beta_0 + \beta_1 * X1_i + \beta_2 * X2_i + \dots \beta_k * Xk_i}$

μ_i = número de eventos predichos

Regresión de Poisson

La función de $\log(L)$ para los modelos lineales generalizados con distribución Poisson:

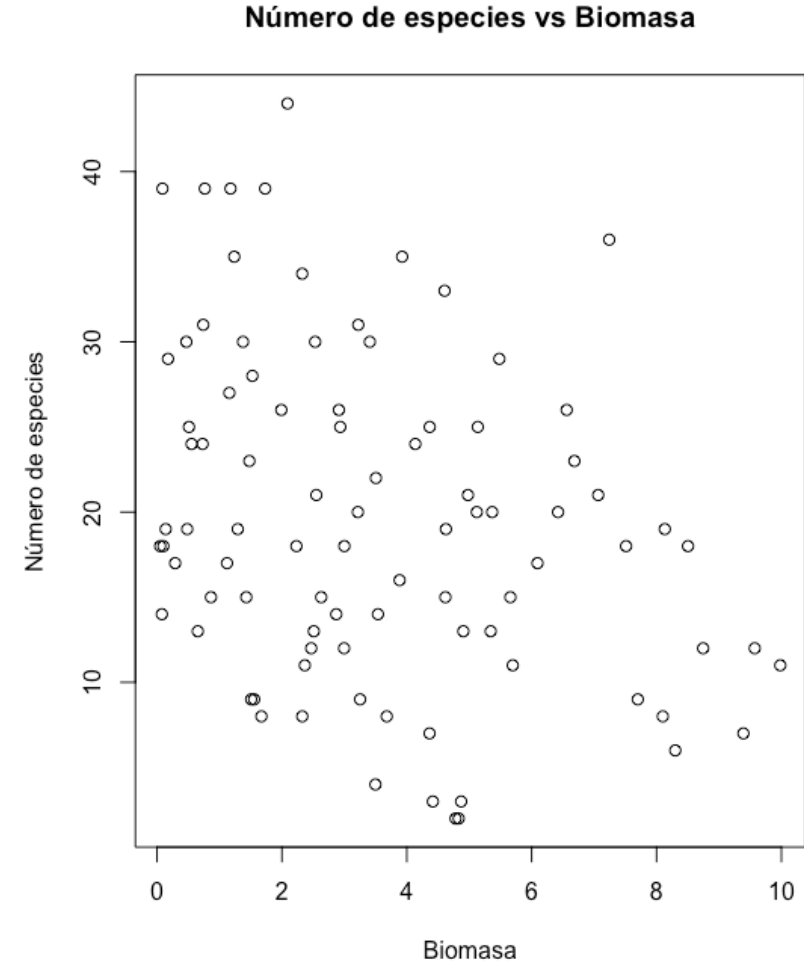
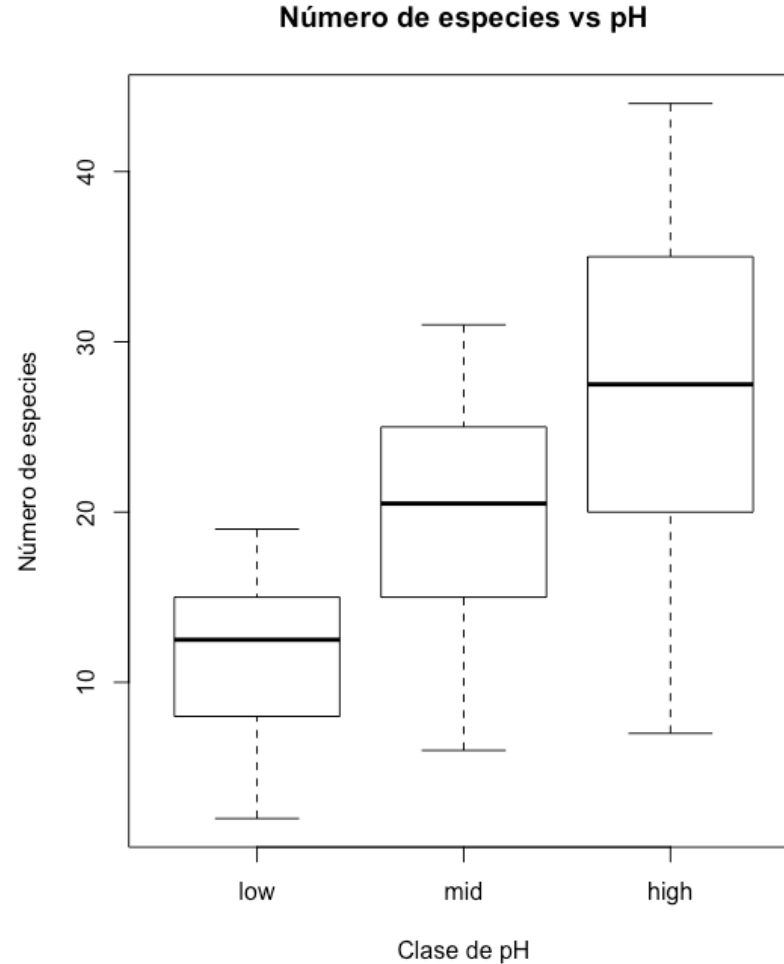
$$\log(L) = \sum_{i=1}^n (y_i \log \mu_i - \mu_i)$$

Es esta función la cual es maximizada para encontrar las estimaciones de los parámetros

Regresión de Poisson

Ejemplo: Número de especies vegetales en función de la biomasa total y del pH

Biomasa: continua
pH: continua
No. especies: discreta



Regresión de Poisson

Ajustar un modelo de regresión Poisson del número de especies en función del pH, la biomasa y una interacción Biomasa*pH

Función R: glm (family= ,)

```
mod1<-glm(Especie~Biomasa+pH+Biomasa:pH, family=poisson(link = log), data=especies)
```

family: especifica la distribución (Poisson, normal, binomial, etc.)

link: especifica la función de vínculo (log, id, logit)

Regresión de Poisson

```
names(mod1)
```

```
[1] "coefficients"    [2] "residuals"      [3] "fitted.values"
[4] "effects"         [5] "R"              [6] "rank"
[7] "qr"              [8] "family"         [9] "linear.predictors"
[10] "deviance"        [11] "aic"
[12] "null.deviance"   [13] "iter"           [14] "weights"
[15] "prior.weights"   [16] "df.residual"     [17] "df.null"
[18] "y"               [19] "converged"       [20] "boundary"
[21] "model"           [22] "call"           [23] "formula"         [24] "terms"
[25] "data"            [26] "offset"          [27] "control"          [28] "method"
[29] "contrasts"       [30] "xlevels"
```

Regresión de Poisson

```
> summary(mod1)
Call:
glm(formula = Especie ~ Biomasa + pH + Biomasa:pH, family = poisson(link = log),
data = especies)

Deviance Residuals:
    Min       1 Q   Median       3 Q      Max 
-2.4978  -0.7485  -0.0402   0.5575   3.2297 

Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  3.76812    0.06153   61.240 2e-16 ***
Biomasa      0.10713    0.01249   8.577 2e-16 ***
pHlow        0.81557    0.10284   7.931 2.18e-15 ***
pHmid        0.33146    0.09217   3.596 0.000323 ***
Biomasa:pHlow 0.15503    0.04003   3.873 0.000108 ***
Biomasa:pHmid 0.03189    0.02308   1.382 0.166954    
---
Signif. codes:  '0.001 ***' '0.01 **' '0.05 .' '0.1 ' ' '
(Dispersion parameter for Poisson family taken to be 1)

Null deviance: 452.346 on 89 degrees of freedom
Residual deviance: 13.201 on 84 degrees of freedom
AIC: 14.39
```

Regresión de Poisson

Podemos resumir la información con una tabla análoga de ANOVA que lidie con variables discretas (número de especies) utilizando una prueba Chi-cuadrada. **Función R `anova()`**

```
> anova(mod1, test="Chisq")
Analysis of Deviance Table

Model: poisson, link: log
Response: Especie
Terms added sequentially (first to last)

              Df Deviance   Resid. Df   Resid. Dev Pr(>Chi)
NULL              9      52.35
Biomasa           1      4.673    8      407.67 2.328e-11
pH                2      308.431   6      99.24 2.2e-16
Biomasa:pH        2      6.040    4      33.20 0.0003288
NULL              1
Biomasa          ***
pH               ***
Biomasa:pH       ***
```

- Da una idea de la importancia global de cada efecto.
- Compara el modelo añadiendo variables secuencialmente.

Regresión de Poisson

Podemos resumir la información con una tabla análoga de ANOVA que lidie con variables discretas (número de especies) utilizando una prueba Chi-cuadrada

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)	Modelo
NULL	89	452.35				Bo
Biomasa	88	407.67			2.328e-11	Bo+Biomasa
pH	86	399.24			2.2e-16	Bo+Biomasa+pH
Biomasa:pH	84	383.20			0.0003288	Bo+Biomasa+pH+Biomasa*pH
NULL						
Biomasa	**					
pH	**					
Biomasa:pH	**					
—						

Regresión de Poisson: Análisis de devianza

Análisis de devianza

- Prueba basada en la diferencia de $\log(L)$ entre los modelos:

$$LR = -2 * (\log L_{\text{modelo simple}} - \log L_{\text{modelo complejo}})$$

```
mod1<-glm(Especie~Biomasa+pH+Biomasa:pH, family=poisson(link = log), data=especies)
```

```
mod2<-glm(Especie~Biomasa+pH, family=poisson, data=especie)
```

Regresión de Poisson: Análisis de devianza

Podemos cuantificar el efecto de la inclusión de la interacción Biomasa*pH
Correr los dos modelos:

	Modelo complejo	Modelo sin interacción	LR	GL
logL	-251.1	-259.2	$-2 * (-259.2 + 251.1)$	
Parámetros	6	4		2

```
LR<--2*(logLik(mod2)[1]-logLik(mod1)[1])
```

	Modelo complejo	Modelo sin interacción	LR	GL
logL	-251.1	-259.2	16.04	
Parámetros	6	4		2

Regresión de Poisson: Análisis de devianza

De manera general podemos utilizar la función anova ()

```
> anova(mod2, mod1, test="Chi")
Analysis of Deviance Table

Model 1: Especie ~ Biomasa + pH
Model 2: Especie ~ Biomasa + pH + Biomasa:pH
Df Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1 6 9.242 
2 4 3.201 2 16.04 0.0003288 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

El efecto de la inclusión de la interacción Biomasa*pH es significativo

Regresión de Poisson: Devianza residual

Es una medida del ajuste del modelo (valor pequeño = mejor modelo)

$$DR = 2 \sum_{i=1}^n y_i * \log(y_i/\mu_i) - (y_i - \mu_i)$$

```
dev.residual1<-2*sum(especies$Especie*log(especies$Especie/mod1$fitted)-  
(especies$Especie-mod1$fitted))
```

```
dev.residual1[1]  
83.20114
```

O directamente extraerlo del objeto mod1

```
mod1$deviance[1] 83.2011
```

```
mod2$deviance[1] 99.2415
```

Regresión de Poisson: Verificación de supuestos

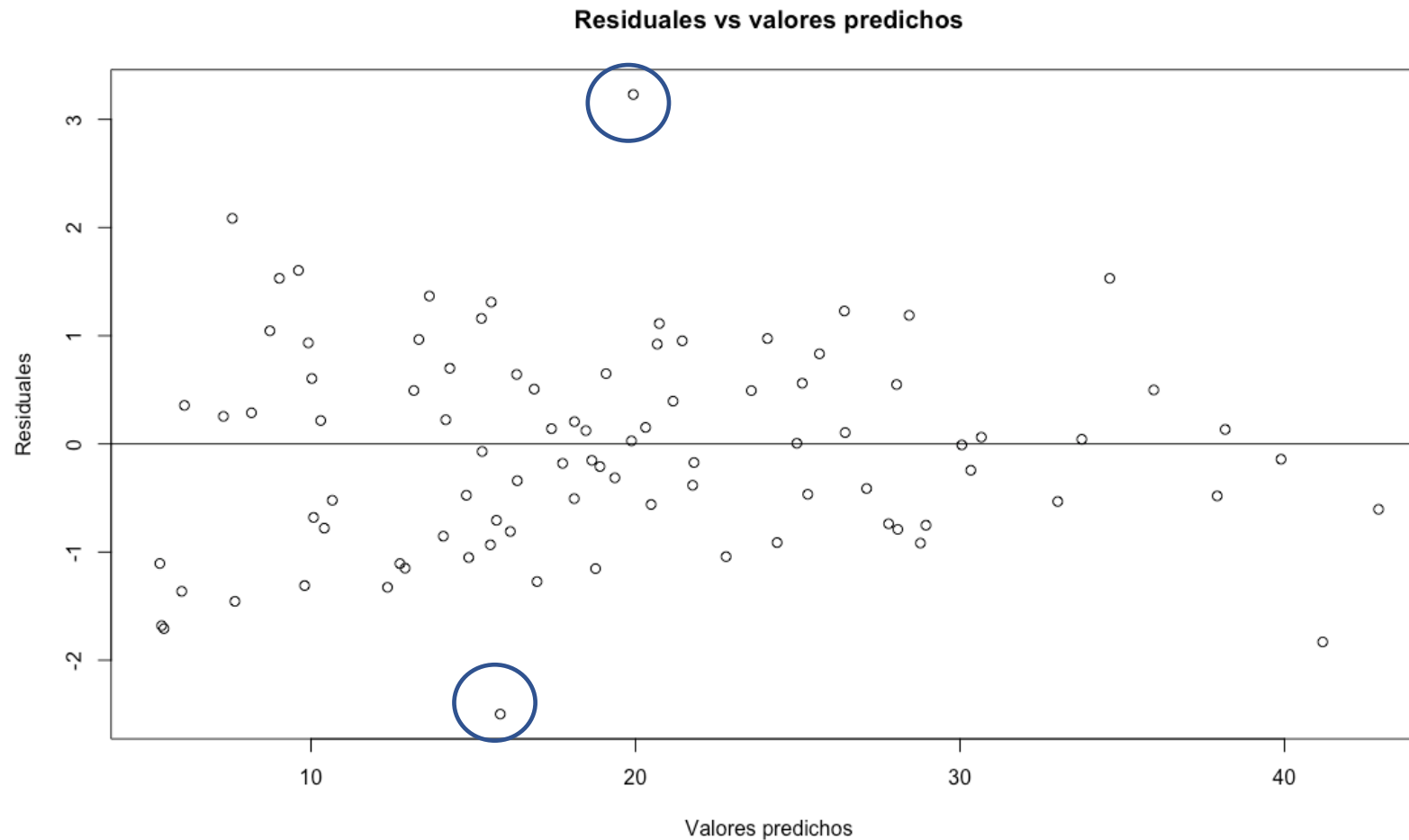
Puntos a verificar:

1. Gráfica de residuales y otros (efecto multiplicador “levier”, influencia)
2. Dispersión (varianza/media)

Regresión de Poisson: Verificación de supuestos

Residuales modelo: Especie~Biomasa+pH+Biomasa*pH

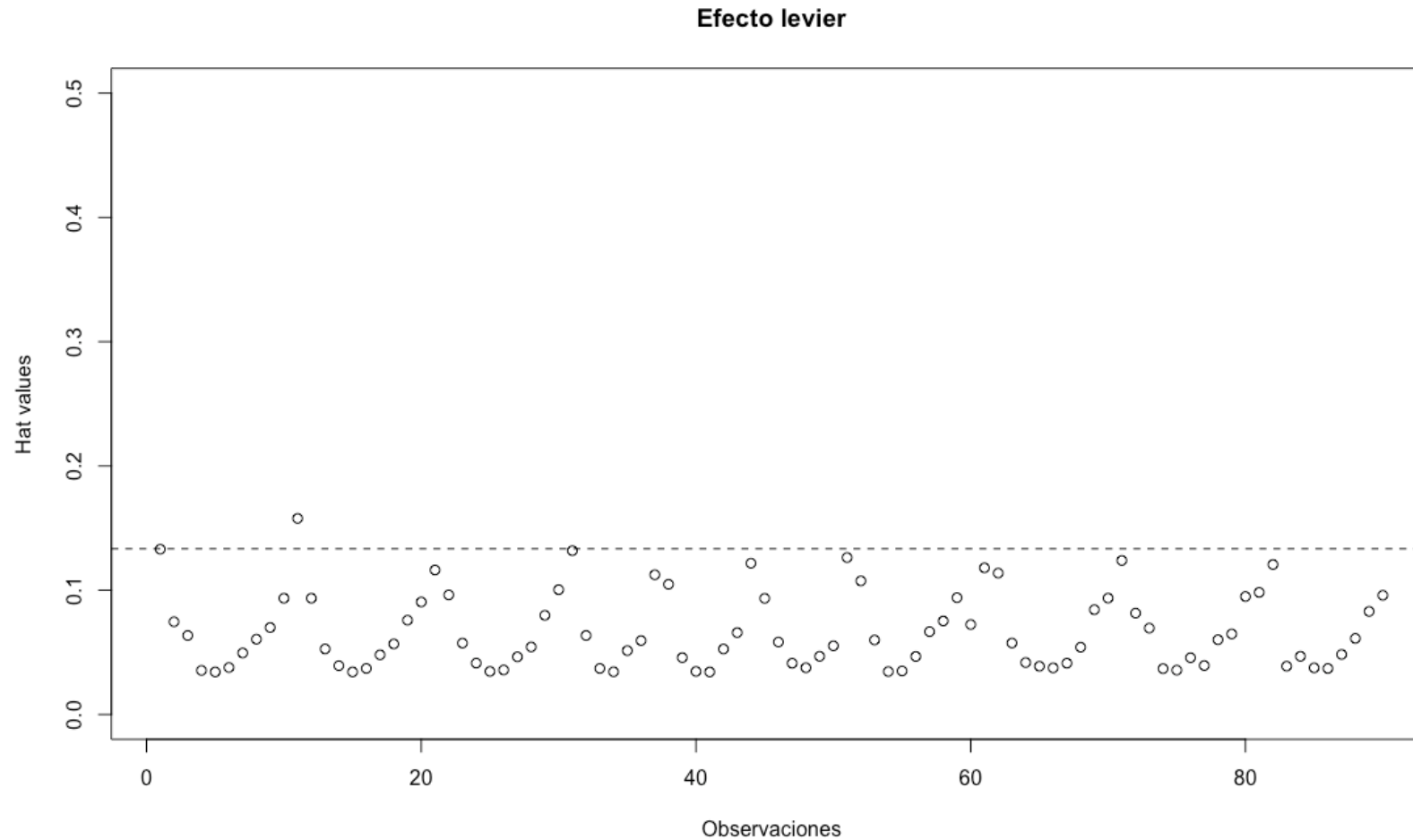
Función R : residuals (, type= “response”, “pearson”, “deviance”)



Regresión de Poisson: Verificación de supuestos

Efecto levier modelo: $\text{Especie} \sim \text{Biomasa} + \text{pH} + \text{Biomasa} * \text{pH}$

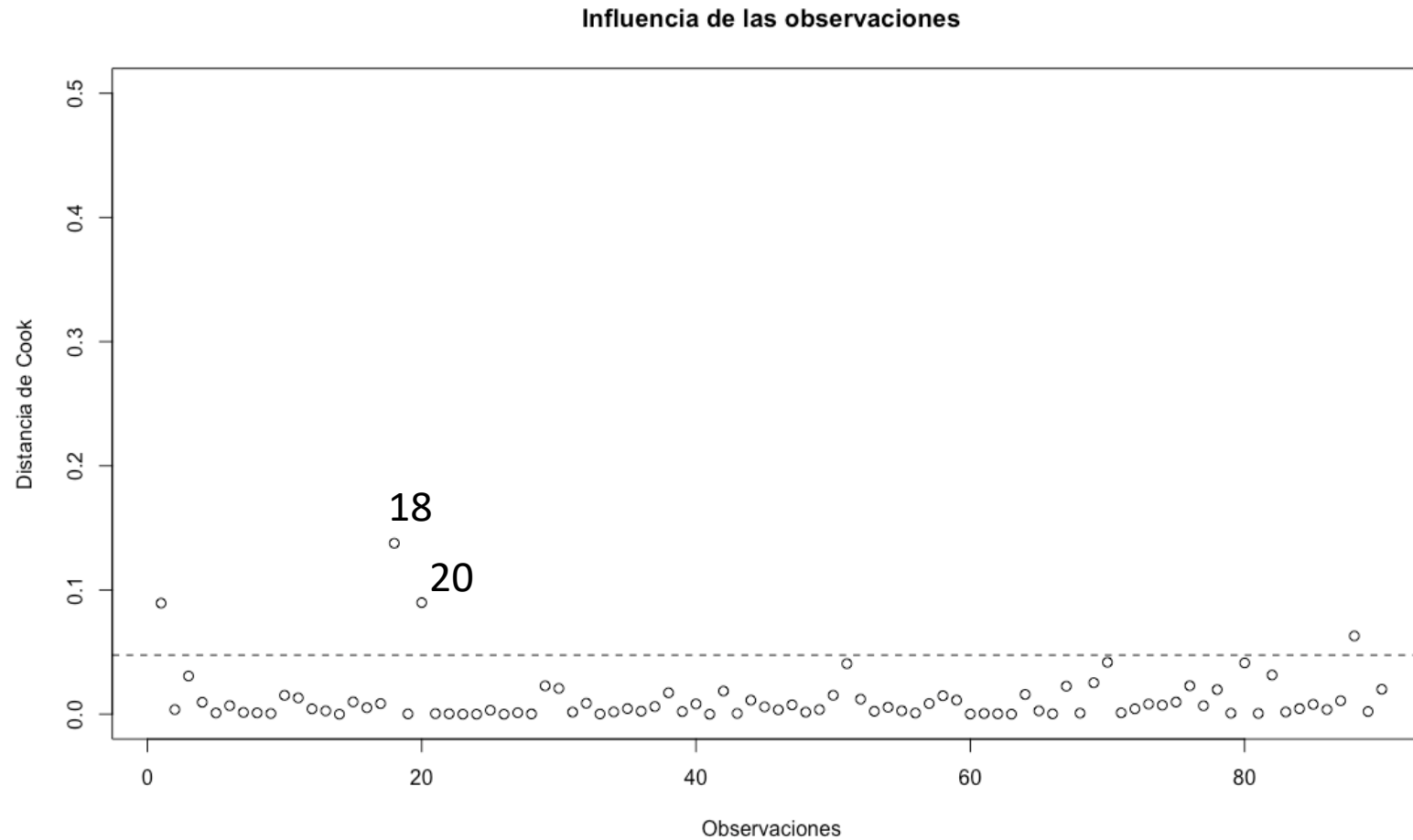
Función R : `hatvalues()`



Regresión de Poisson: Verificación de supuestos

Influencia modelo: Especie~Biomasa+pH+Biomasa*pH

Función R : `cooks.distance ()`



Regresión de Poisson: Verificación de supuestos

Modelo Especie~Biomasa+pH+Biomasa*pH

Dispersión: Varianza/media

En la distribución de Poisson la varianza=media

Por lo tanto varianza/media (\hat{c}) = 1

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for poisson family taken to be 1)

Regresión de Poisson: Verificación de supuestos

Modelo $\text{Especie} \sim \text{Biomasa} + \text{pH} + \text{Biomasa} * \text{pH}$

Cuando $(\hat{c}) > 1$ sobredispersión, por lo que los errores estándar deben ser ajustados
Sobredispersión: Mayor presencia de gran variabilidad (dispersión estadística) en los datos de lo que se esperaría dado el modelo estadístico.

Cuando $(\hat{c}) > 4$ la regresión Poisson no es una buena opción

$$\hat{c} = \frac{\chi^2 \text{ de Pearson}}{n - p}$$

donde

$$\chi^2 \text{ de Pearson} = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i}$$

Regresión de Poisson: Verificación de supuestos

Modelo Especie~Biomasa+pH+Biomasa*pH

```
chisq<-sum(((especies$Especie-mod1$fitted)^2)/mod1$fitted)  
chisq[1] 83.74812
```

```
c_hat<-chisq/mod1$df.residual  
c_hat[1] 0.9970014
```

**Un modelo lineal generalizado
con distribución Poisson es
apropiado**

$$\hat{c} = \frac{\chi^2 \text{ de Pearson}}{n - p}$$

$$\hat{c} = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i}$$

Regresión de Poisson: Interpretación

```
> anova(mod1, test="Chisq")
Analysis of Deviance Table

Model: poisson, link: log
Response: Especie
Terms added sequentially (first to last)

            Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL              9      452.35
Biomasa           1      4.673    8      407.67 2.328e-11
pH                2      308.431  6      399.24 2.2e-16
Biomasa:pH        2      6.040    4      333.20 0.003288
NULL
Biomasa ***
pH ***
Biomasa:pH ***
```

- Hay en efecto significativo de la interacción entre la biomasa y el pH sobre el número de especies.
- Por lo tanto, no podemos interpretar los efectos individuales de manera aislada.
- Conceptualmente, el análisis de una interacción equivale a un análisis separado por cada grupo (pH) de la variable de respuesta en función de la variable numérica (biomasa).

Regresión de Poisson : Interpretación

Coefficients:

	Estimate	Std. Error	Value	Pr(> z)
(Intercept)	3.76812	0.06153	61.240	2e-16***
Biomasa	0.10713	0.01249	8.577	2e-16***
pHlow	0.81557	0.10284	7.931	2.18e-15***
pHmid	0.33146	0.09217	3.596	0.000323***
Biomasa:pHlow	0.15503	0.04003	3.873	0.000108***
Biomasa:pHmid	0.03189	0.02308	1.382	0.166954

pH es una variable
categórica con nivel
de referencia “high”

β_0 = corresponde a la ordenada al origen el número de especies del grupo de pH “high” y la Biomasa

$\beta_{biomasa}$ = corresponde al efecto de la biomasa sobre el número de especies del grupo “high”

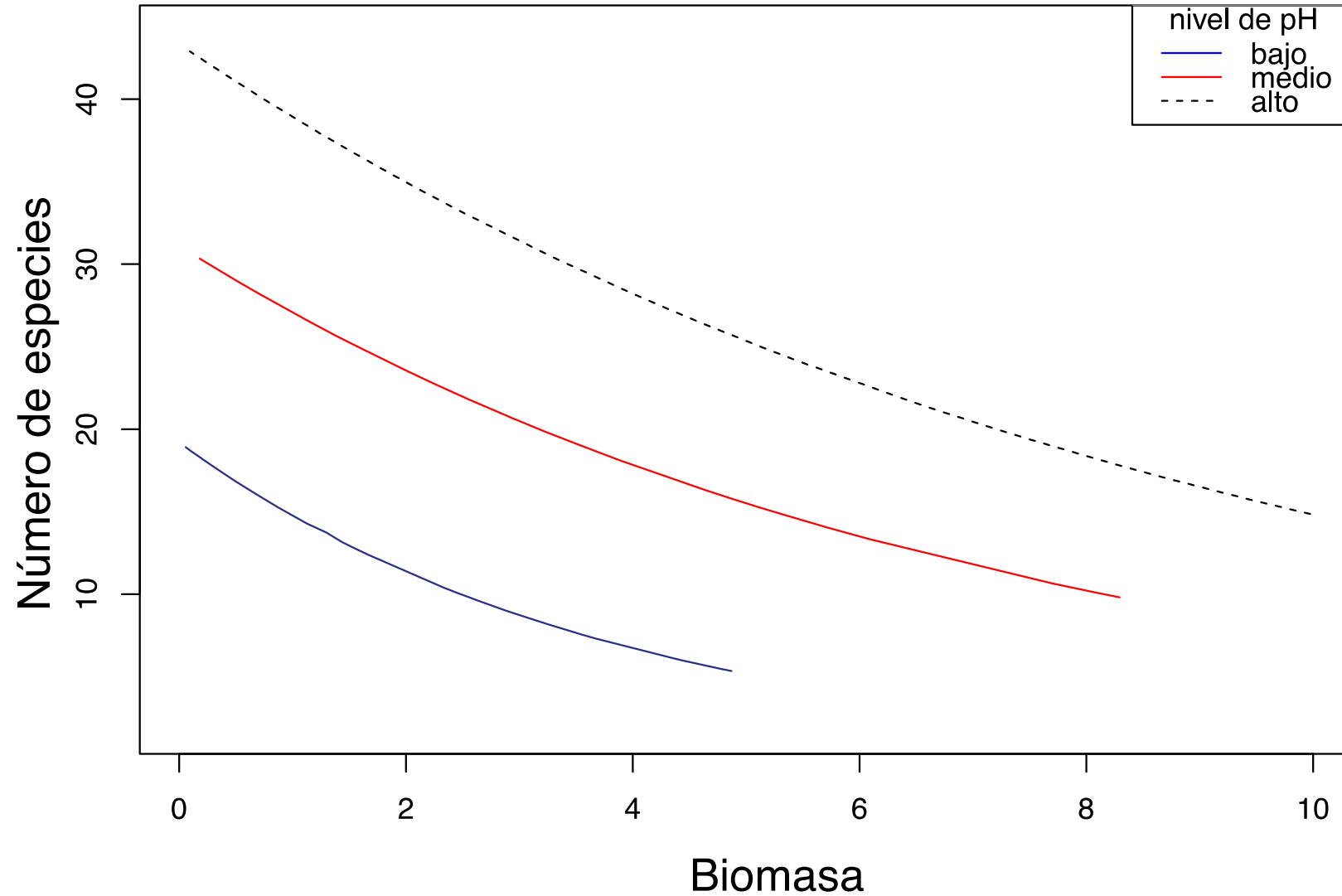
β_{pHlow} = corresponde a la diferencia entre la ordenada al origen del grupo “low” vs grupo “high”

β_{pHmid} = corresponde a la diferencia entre la ordenada al origen del grupo “mid” vs grupo “high”

$\beta_{biomasa:pHlow}$ = corresponde al efecto de la biomasa del grupo “low” vs grupo “high”

$\beta_{biomasa:pHmid}$ = corresponde al efecto de la biomasa del grupo “mid” vs grupo “high”

Regresión de Poisson : Interpretación



Diferencias de
pendiente entre los
tres niveles de pH

Regresión de Poisson : Interpretación

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.76812	0.06153	61.240	2e-16***
Biomasa	0.10713	0.01249	8.577	2e-16***
pHlow	0.81557	0.10284	7.931	2.18e-15***
pHmid	0.33146	0.09217	3.596	0.000323***
Biomasa:pHlow	0.15503	0.04003	3.873	0.000108***
Biomasa:pHmid	0.03189	0.02308	1.382	0.166954

—

$$\log_e(\mu_i) = \beta_0 + \beta_1 * X1_i + \beta_2 * X2_i + \dots \beta_k * Xk_i$$

Valor predicho para la primera observación

especies [1,]	pH	Biomasa	Especie
1 high	0.4692972	30	

Regresión de Poisson : Interpretación

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.76812	0.06153	61.240	2e-16 ***
Biomasa	-0.10713	0.01249	-8.577	2e-16 ***
pHlow	-0.81557	0.10284	-7.931	2.18e-15 ***
pHmid	-0.33146	0.09217	-3.596	0.000323 ***
Biomasa:pHlow	-0.15503	0.04003	-3.873	0.000108 ***
Biomasa:pHmid	-0.03189	0.02308	-1.382	0.166954

especies [1,] pH Biomasa Especie
 high 0.4692972 30

$$\log_e(\mu_i) = 3.76 - 0.10 * 0.46 - 0.08 * 0 - 0.33 * 0 - 0.15 * 0 - 0.03 * 0$$

$$\log_e(\mu_i) = 3.714$$

$$\mu_i = \exp(3.714) = 41.0$$

Regresión de Poisson : Interpretación

Podemos acceder directamente a los valores del predictor lineal y los valores predichos

```
mod1$linear.predictors[1:5]
```

1	2	3	4	5
3.717848	3.582696	3.544246	3.347555	3.300310

```
mod1$fitted.values[1:5]
```

1	2	3	4	5
41.17568	35.97038	34.61358	28.43313	27.12105

Regresión de Poisson : Interpretación

```
especies [50,]  pH  Biomasa  Especie  
mid  5.120968    20
```

```
model.matrix(mod1)[50,]  
(Intercept)    Biomasa    pHlow    pHmid  
1.000000    5.120968    0.000000    1.000000  
Biomasa:pHlow  Biomasa:pHmid  
0.000000          5.120968
```

```
linpred<-model.matrix(mod1)[50,]%*%coef(mod1)
```

```
exp(linpred)  
15.25237
```

Regresión de Poisson : Selección de modelos

Corrección del AIC para la sobredispersión:

Incorporamos el \hat{c} en el cálculo de AIC (QAIC):

$$QAIC = \frac{-2 \log(L)}{\hat{c}} + 2K$$

Si $\hat{c} > 4$ el modelo de regresión de Poisson no es una buena opción.

El modelo de regresión binomial negativo permite modelar datos discretos con sobredispersión.

Regresión Binomial negativa

Distribución binomial negativa

Características:

- Distribución discreta (valores 0,1,2,3)
- Caracteriza los eventos raros (muchos ceros)
- Distribución sesgada sesgada a la derecha
- El parámetro de dispersión k permite incorporar la variabilidad cuando la distribución Poisson no es apropiada

Regresión Binomial negativa

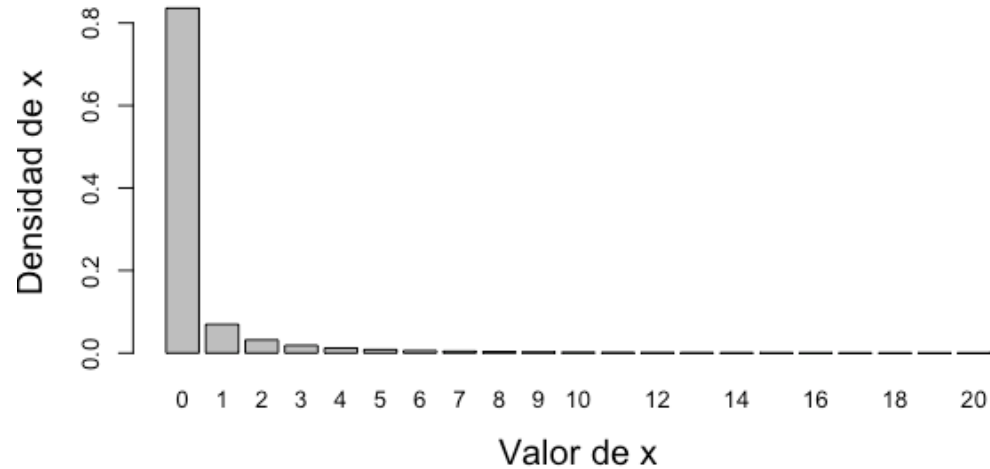
Distribución binomial negativa

Características:

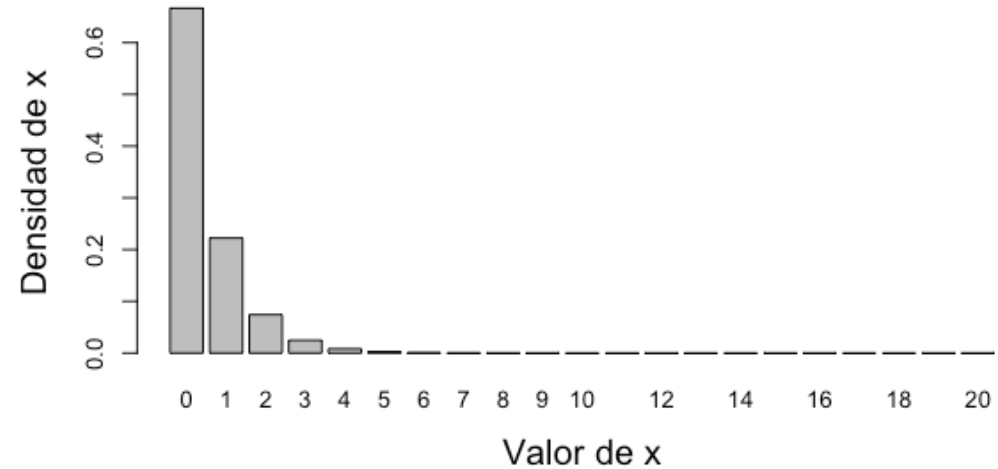
- El parámetro k es desconocido por lo debemos estimarlo a partir de los datos
- Se utiliza una función de vínculo log para la regresión binomial negativa (como en la Poisson).
- También se puede utilizar una función de vínculo de identidad o raíz cuadrada cuando sea necesario.
- Mientras k aumenta, la función tiende a la distribución Poisson

Regresión Binomial negativa

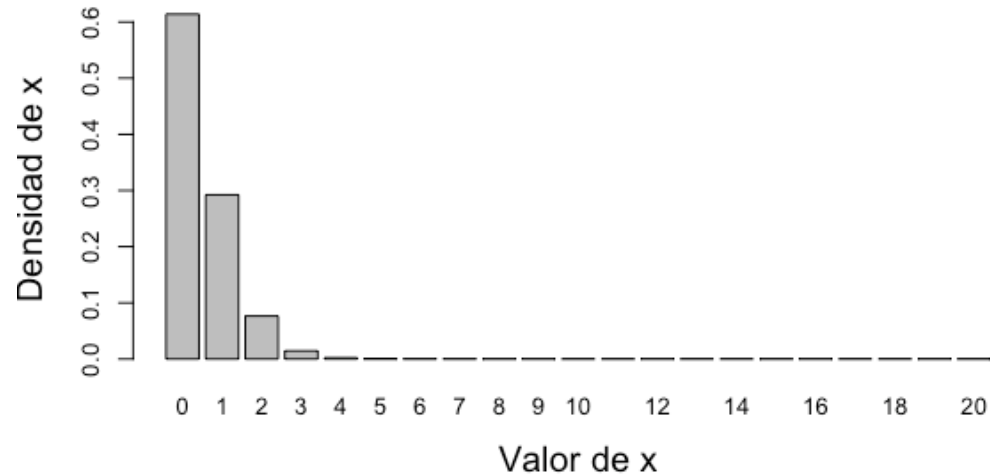
Distribución binomial negativa ($\mu = 0.5, k = 0.1$)



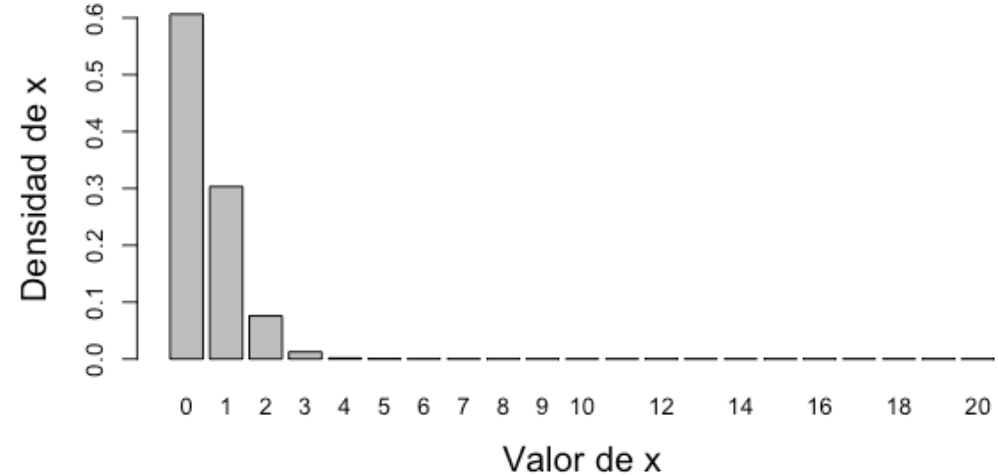
Distribución binomial negativa ($\mu = 0.5, k = 1$)



Distribución binomial negativa ($\mu = 0.5, k = 10$)



Distribución Poisson ($\mu = 0.5$)



Regresión Binomial negativa

Ejemplo: Diversidad de especies vegetales en las islas Galápagos. Relación entre el número de especies vegetales y variables geográficas

str(gala)'data.frame' 30 obs. of 7 variables:

\$ Species : num 58 31 3 25 2 18 24 10 8 2 (número de especies)

\$ Endemics : num 23 21 3 9 1 11 0 7 4 2 (número de especies endémicas)

\$ Area : num 25.09 1.24 0.21 0.1 0.05 (área de la isla km²)

\$ Elevation: num 346 109 114 46 77 119 93 168 71 112 (Elevación de la isla)

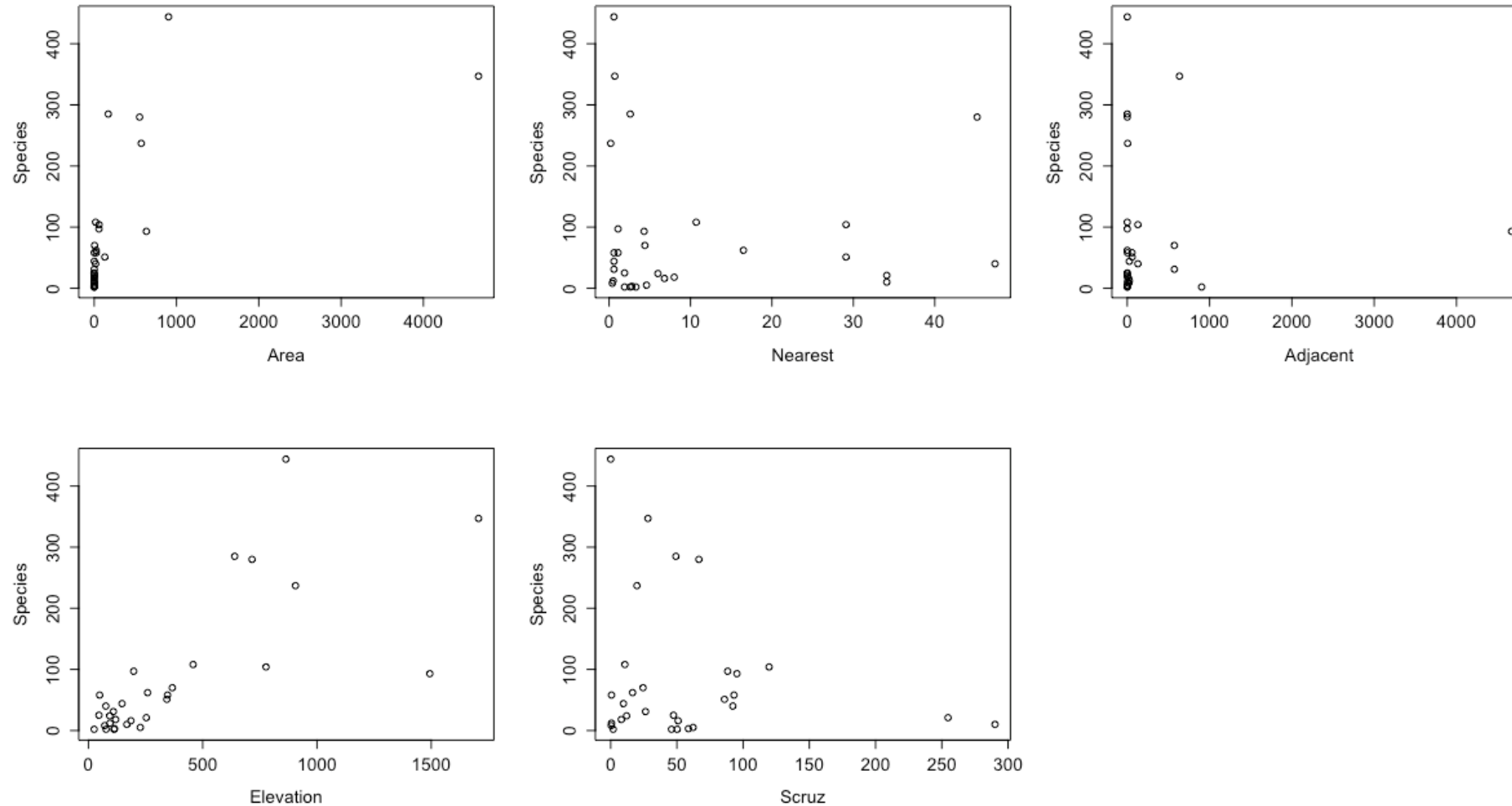
\$ Nearest : num 0.6 0.6 2.8 1.9 1.9 8 6 34.1 0.4 2.6 (la distancia a la isla más cercana)

\$ Scruz : num 0.6 26.3 58.7 47.4 1.9 (distancia a la isla Sta. Cruz)

\$ Adjacent : num 1.84 572.33 0.78 0.18 903.82 (el área de la isla adyacente)

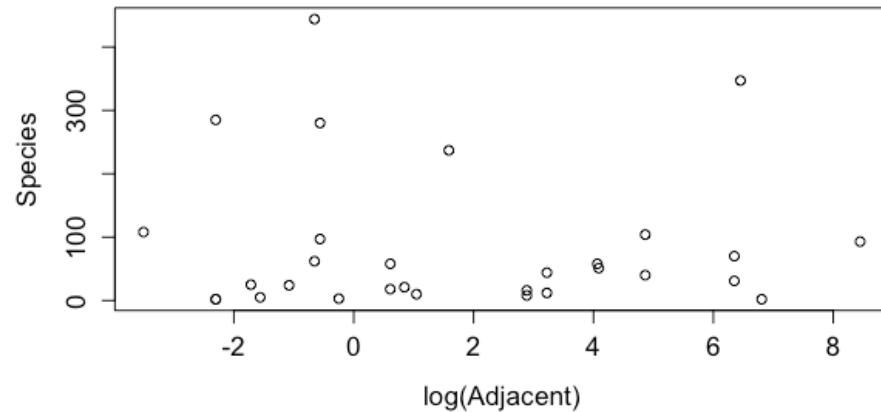
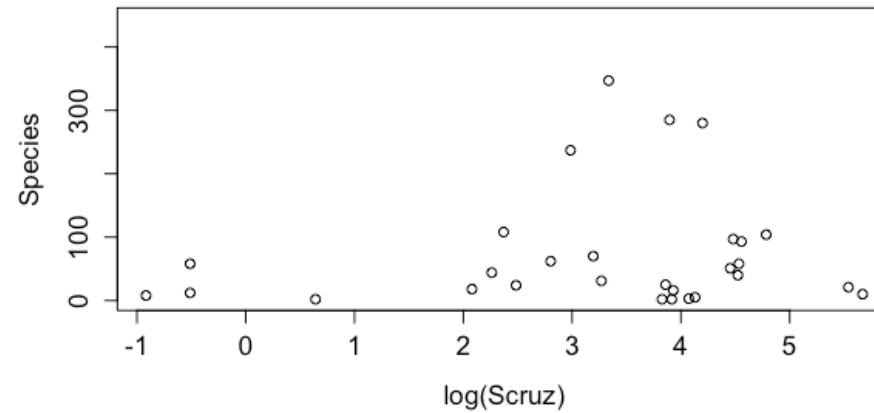
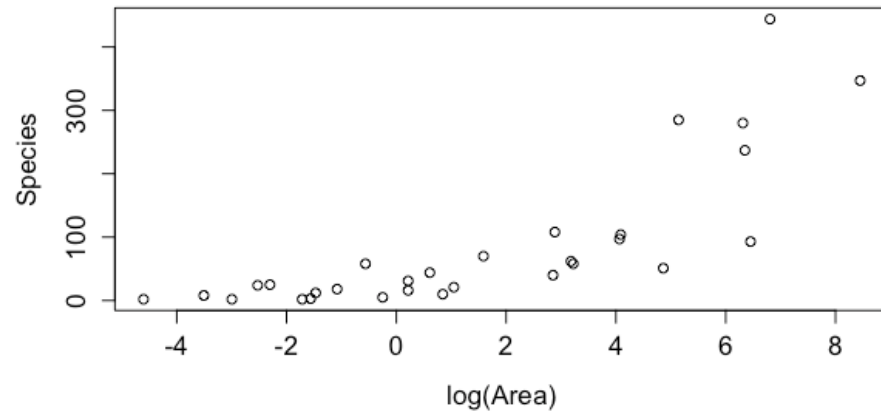
Regresión Binomial negativa

Ejemplo: Diversidad de especies vegetales en las islas Galápagos. Relación entre el número de especies vegetales y variables geográficas



Regresión Binomial negativa

Ejemplo: Diversidad de especies vegetales en las islas Galápagos. Relación entre el número de especies vegetales y variables geográficas



Transformar variables (log)
para mejorar la distribución
de las observaciones

Regresión Binomial negativa

```
summary(gala1)Call:glm(formula = Species ~ log_Area + Elevation + Nearest + log_Adjacent +  
log_Scruz, family = poisson, data = gala)
```

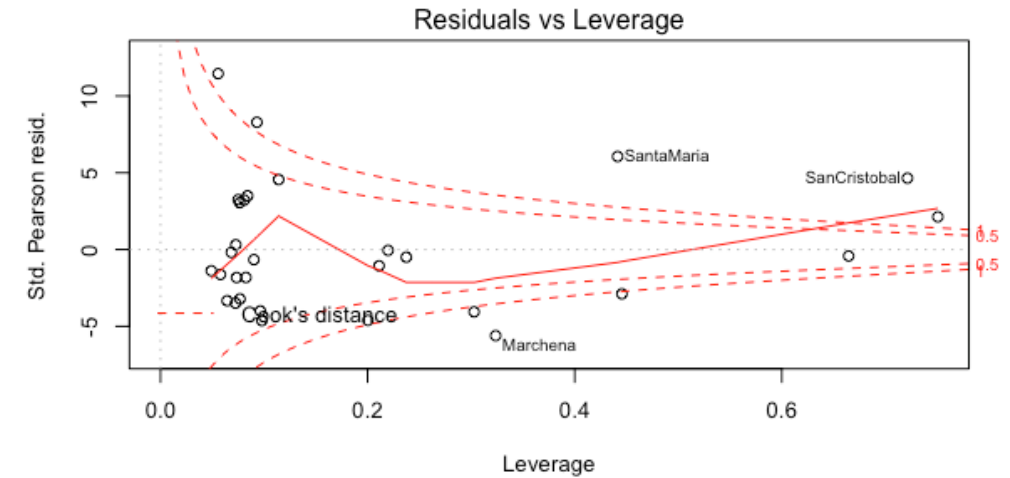
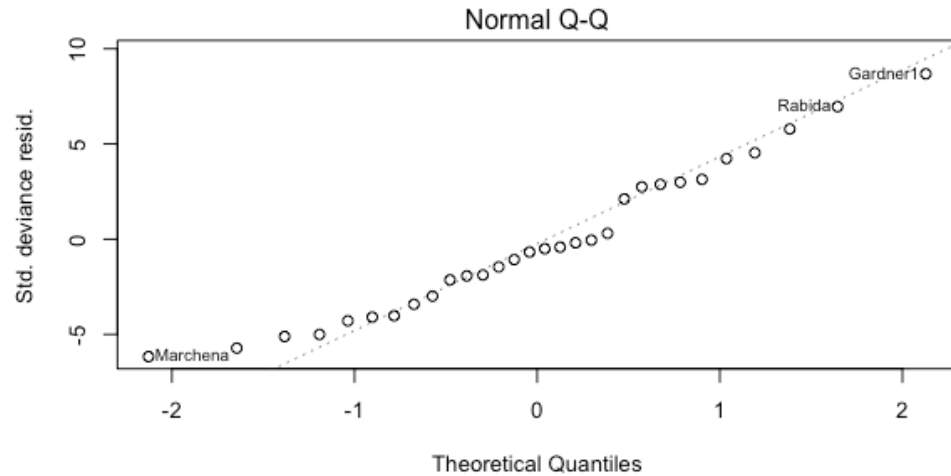
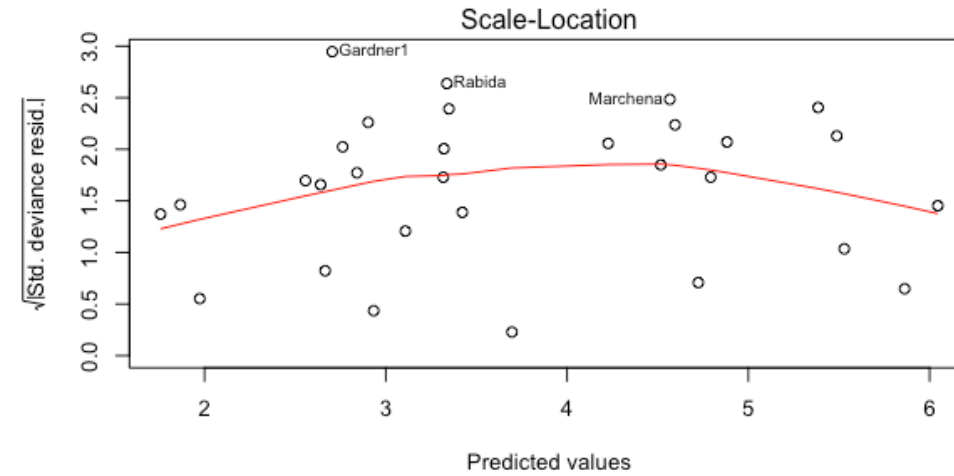
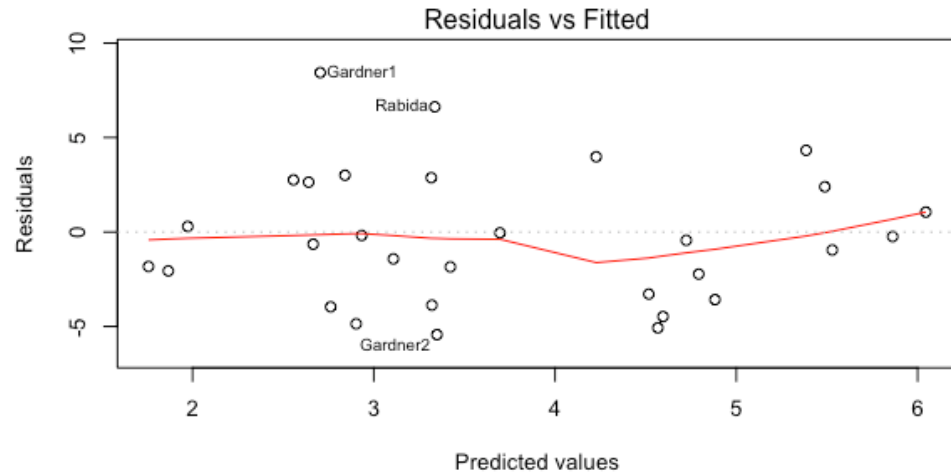
```
Deviance Residuals:   Min     1Q   Median     3Q      Max  
                   -5.4341  -3.0204  -0.5416   2.5770   8.4359
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.505e+00	6.368e-02	55.048	< 2e-16 ***
log_Area	3.575e-01	1.615e-02	22.135	< 2e-16 ***
Elevation	5.663e-05	1.044e-04	0.543	0.58732
Nearest	-3.431e-03	1.674e-03	-2.050	0.04038 *
log_Adjacent	-9.130e-02	8.760e-03	-10.423	< 2e-16 ***
log_Scruz	-5.003e-02	1.535e-02	-3.259	0.00112 **

```
---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for poisson family taken  
to be 1) Null deviance: 3510.73  on 29  degrees of freedomResidual deviance: 361.17  on 24  degrees of  
freedomAIC: 534Number of Fisher Scoring iterations: 5
```

Regresión Binomial negativa



Regresión Binomial negativa

```
chisq<-sum(((gala$Species-gala1$fitted)^2)/gala1$fitted)  
c.hat<-chisq/gala1$df.residual  
c.hat[1] 16.79366
```

Definitivamente no sigue una distribución Poisson.
Como $c.hat > 4$ es mejor intentar ajustar una distribución binomial negativa

Regresión Binomial negativa

```
library(MASS)
```

```
mod2<-glm.nb(Species~log_Area+Elevation+Nearest+log_Adjacent+log_Scruz,  
data=gala)
```

La función de vínculo por defecto es link=log

Regresión Binomial negativa

```
summary(mod2)
```

```
Call:glm.nb(formula = Species ~ log_Area + Elevation + Nearest + log_Adjacent + log_Scruz, data =  
gala, init.theta = 2.938490105, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1466	-0.8867	-0.2338	0.4581	1.9671

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.5621341	0.2971453	11.988	< 2e-16 ***
log_Area	0.4215363	0.0660078	6.386	1.7e-10 ***
Elevation	-0.0005481	0.0005335	-1.027	0.304
Nearest	-0.0137115	0.0102400	-1.339	0.181
log_Adjacent	-0.0250659	0.0376654	-0.665	0.506
log_Scruz	-0.0277083	0.0878747	-0.315	0.753

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

Regresión Binomial negativa

(Dispersion parameter for Negative Binomial(2.9385) family taken to be 1)

Null deviance: 149.161 on 29 degrees of freedom

Residual deviance: 32.782 on 24 degrees of freedom

AIC: 287.9 Number of Fisher Scoring iterations: 1

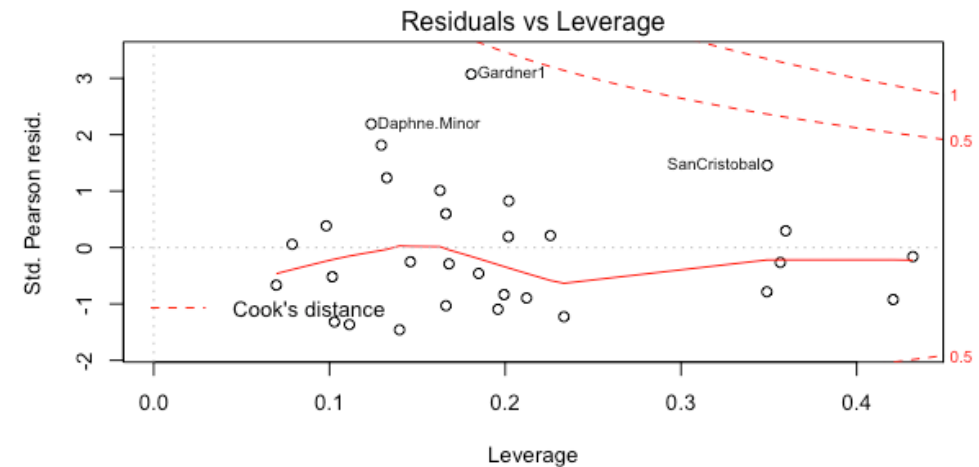
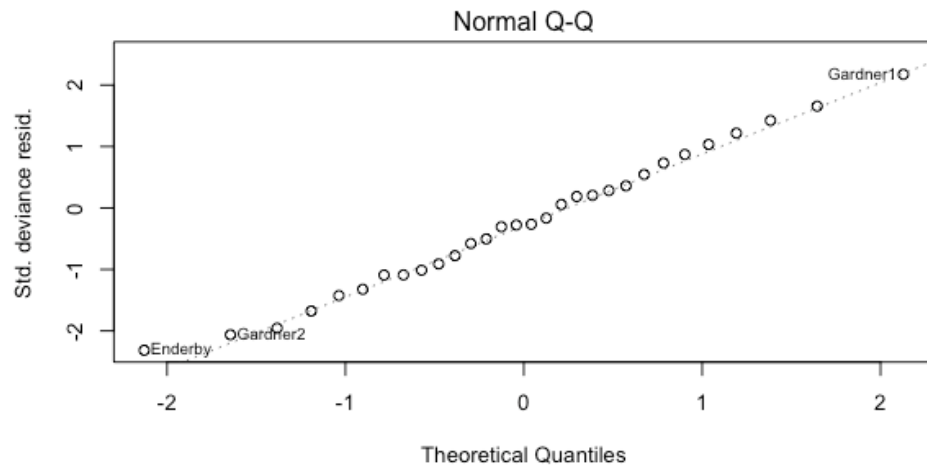
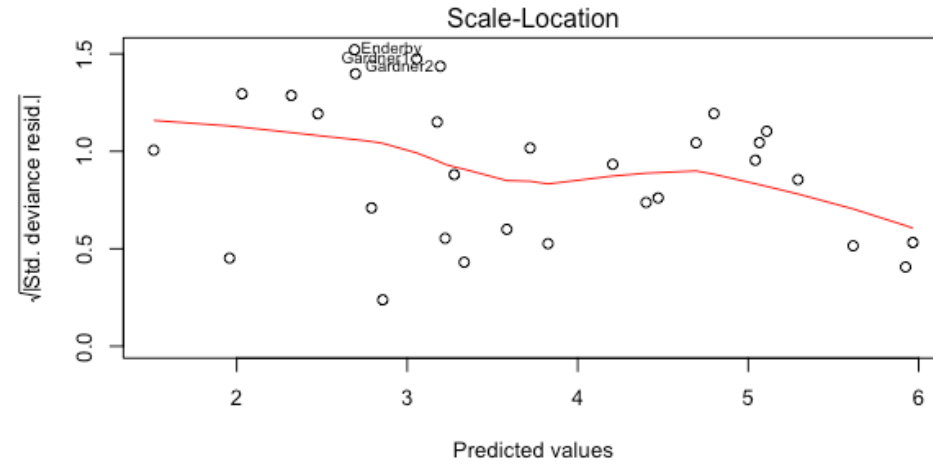
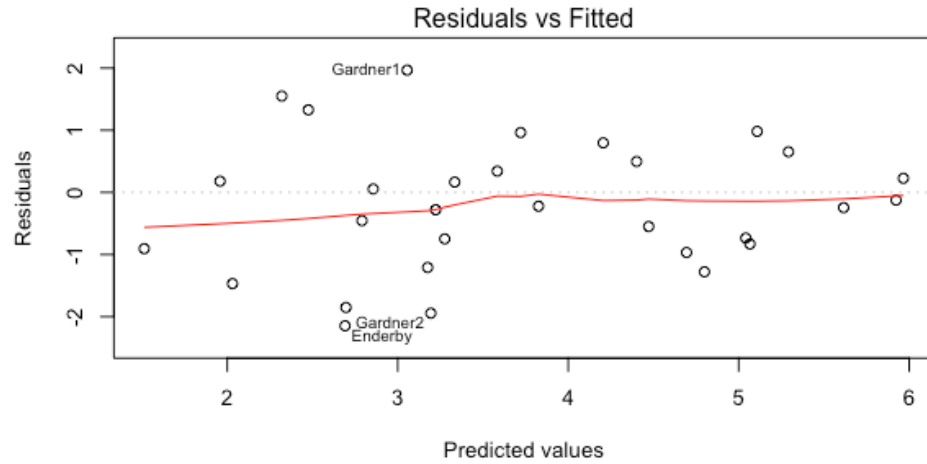
Theta: 2.938

Std. Err.: 0.866

2 x log-likelihood: -273.896

Theta: parámetro de dispersión estimado mediante máxima verosimilitud con su error estándar, que representa una medida de su precisión

Regresión Binomial negativa



Regresión Binomial negativa

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.5621341	0.2971453	11.988	< 2e-16 ***
log_Area	0.4215363	0.0660078	6.386	1.7e-10 ***
Elevation	-0.0005481	0.0005335	-1.027	0.304
Nearest	-0.0137115	0.0102400	-1.339	0.181
log_Adjacent	-0.0250659	0.0376654	-0.665	0.506
log_Scruz	-0.0277083	0.0878747	-0.315	0.753

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

Se concluye que el número de especies vegetales aumenta con el log de la superficie de la isla

Regresión Binomial negativa

```
mod2b<-glm(Species~log_Area+Elevation+Nearest+log_Adjacent+log_Scruz,  
data=gala, family=negative.binomial(theta=2.938))  
logLik(mod2) #theta es un parámetro adicional  
'log Lik.' -136.9479 (df=7)  
> logLik(mod2b)  
'log Lik.' -136.9479 (df=6)
```

Ejercicio

1. Importar la base de datos de insectos acuáticos (*Dystiscus sp.*) >>Prey_capt.txt>> que contiene información sobre el número de presas capturadas por individuo durante 3 horas en un acuario, en función del tipo de hábitat (abierto o cerrado), el grosor de el esqueleto (mm), el largo total(mm), y el género (macho o hembra).
2. Según la literatura, se presentan 6 modelos potenciales:
 - a) No. Presas = Género
 - b) No.Presas= Género+Hábitat
 - c) No. Presas= Género+Hábitat+Género:Hábitat
 - d) No. Presas= Grosor+Largo+Hábitat
 - e) No. Presas= Largo+Habitat
 - f) No. Presas= Grosor+Largo+Hábitat+Género+Género:Hábitat

Ejercicio

3. Seleccionar el tipo de regresión : Poisson o binomial negativa más apropiado para la base de datos. Justificar la respuesta
 4. Efectuar la selección de modelo AIC (o QAIC, en caso de Poisson) y hacer la tabla de selección modelo. Cuantificar la respuesta.
- 