|Bioestadística – Laboratorio

**MBIO 3203: Laboratorio (1 crédito)**

*Profesor del Curso:* Andrew J. Crawford, Profesor Asociado

*Asistente graduada:* Luisa A. Castellanos

*Monitora:* Valentina Muñoz

Departamento de Ciencias Biológicas, Universidad de los Andes

Semestre 2017-20

**Lab 10 – Modelos Lineales Generalizados**

**La asistencia es obligatoria. Cualquier estudiante que no se presente a la sesión de laboratorio será calificado con cero.**

**Modelos Generales Linealizados (GLM)**

Para **cada uno de los siguientes casos** debe entregar: (**1**) planteamiento de hipotesis (cientifica y estadisticas), (**2**) diagrama de causalidad, (**3**) exploracion grafica, (**4**) planteamiento del modelo estadistico, (**5**) evaluacion del modelo, (**6**) valores predichos para el modelo, (**7**) analisis de ODDS y (**8**) la interpretacion biologica de los resultados.

1. **[2,5 puntos]** La base de datos que encuentra en Sicua+ como Proteinas resume los resultados de un estudio donde se pretendia medir el desempeño de la proteina globina, encargada del transporte de oxigeno, y su desempeño a diferentes temperaturas. Realice en respectivo análisis estadístico y discuta las implicaciones biologicas de este estudio.
2. Planteamiento de hipótesis:

**Científica**

A medida que se aumenta la temperatura, la globina se denatura y no puede llevar a cabo sus funciones habituales. Así pues, a medida que aumenta la temperatura, se espera que haya una disminución en el desempeño de la proteina.

**Estadística**

Hay varias:

Primero, respecto a el poder exlicativo del modelo como tal,

Asumiendo una distribución binomial de la variable Prot\_Pref, se tiene que:

H0: β1= β2=0

Ha: β1≠0 ^ β1≠0

Esto es equivalente a decir que el modelo tiene un poder explicativo superior al de asumir que la temperatura no tiene nigun efecto sobre el desempeño de la proteina.

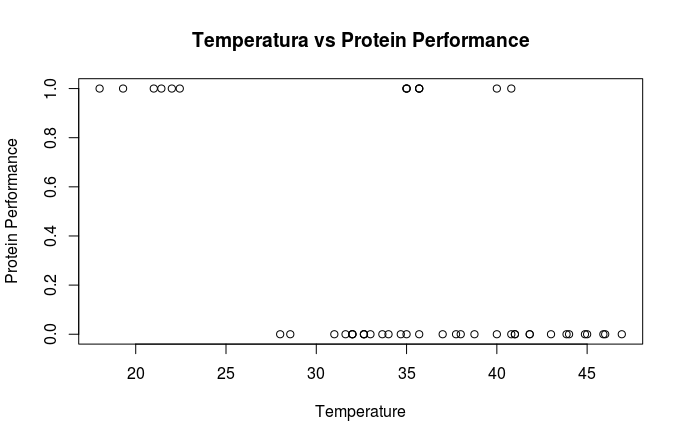
Segundo, respecto a la pendiente y el intercepto del modelo lineal, se puede evaluar el efecto que tiene cada variable sobre la explicación del modelo:

H0: βi= 0

Ha: βi≠ 0

2. Diagrama de causalidad

3. Exploración gráfica



Con la anterior gráfica, podemos ver que la variable Protein Performance tiene dos posibles estados y pareciera que a mayor temperatura, es más probable que la proteina no tenga un buen desempeño. Es decir, pareciera que siguira una distriución binomial.

4.Planteamiento del modelo estadístico

Podemos realizar un GML con función link logit pues los datos siguen una distribución binomial.

5.Evaluación del modelo

> mlr<- glm(Prot\_Perf~Temp,data = prot ,family = binomial)

> summary(mlr)

Call:

glm(formula = Prot\_Perf ~ Temp, family = binomial, data = prot)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.2342 -0.8360 -0.4709 0.6248 2.0798

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 4.88681 1.98638 2.460 0.01389 \*

Temp -0.16979 0.05848 -2.903 0.00369 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 57.949 on 47 degrees of freedom

Residual deviance: 46.135 on 46 degrees of freedom

AIC: 50.135

Number of Fisher Scoring iterations: 4

> anova(mlr,test="Chisq")

Analysis of Deviance Table

Model: binomial, link: logit

Response: Prot\_Perf

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL 47 57.949

Temp 1 11.814 46 46.135 0.0005878 \*\*\*

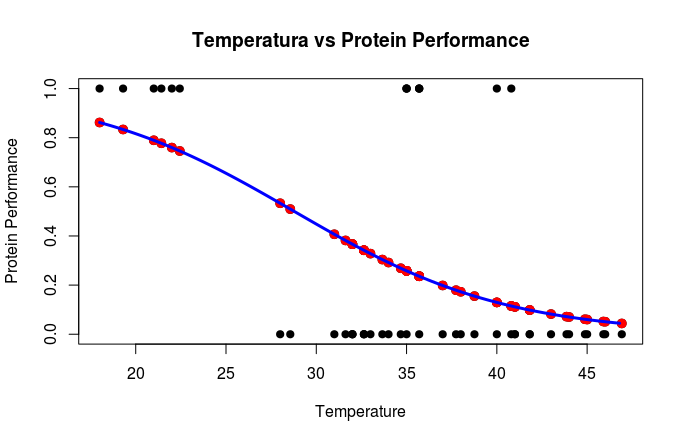
---

Signif. Codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

En base a la información anterior, podemos afirmar que a comparación de una dependencia lineal nula, (β1= β2=0) el modelo actual explica considerablemente mejor los datos dados (Pr(>Chi)=0.0005878) pues con un alpha = 0.05, claramente rechazamos H0.

Además, podemos afirmar que tanto el intercepto (Pr(>|z|) = 0.01389) como la pendiente ( Pr(>|z|) = 0.00369) explican significativamente mejor los datos a comparación de valores nulos (0) si se considera un alfa de 0.05.

6.Valores predichos para el modelo

En la gráfica anterior podemos ver en azul la función de probabilidad en función de la temperatura. Los puntos rojos son los valores predichos por el GML.

7.Análisis de ODDS

> exp(cbind(OR= coef(mlr), confint(mlr)))

Waiting for profiling to be done...

OR 2.5 % 97.5 %

(Intercept) 132.5301000 3.9989462 1.247849e+04

Temp 0.8438399 0.7378974 9.346772e-01

Aqui podemos ver que la temperatura determina en un 84% el desempeño de la proteina. Además, sabemos que determina el desempeño de la proteina en gran medida pues hay un intervalo de confianza del 95% entre 0.73-0.93.

8.Interpretación biológica

El desempeño de la globina respecto al transporte del oxígeno en la sangre sufre considerablemente cuando se incrementa la temperatura.



1. **[2,5 puntos]** La base de datos que encuentra en Sicua+ como Cocodrilos resume los resultados de un estudio donde se pretendia medir si existe alguna relacion entre el tamaño del huevo con el tamaño de la nidada en la especie *Caiman latirostris*. Realice en respectivo análisis estadístico y discuta las implicaciones biologicas de este estudio.

**1.** Planteamiento de hipotesis (cientifica y estadisticas)

**Científica**

Dado que en promedio cada hembra va a tener una cantidad fija de energía que le otorgara a sus huevos, al aumentar la nidada, dada una cantidad finita de energía, el tamaño de los huevos tiene que disminuir.

**Estadística**

Hay varias:

Primero, respecto a el poder exlicativo del modelo como tal, asumiendo una distribución de poisson del tamaño de los huevos, se tiene que:

H0: β1= β2=0

Ha: β1≠0 ^ β1≠0

Esto es equivalente a decir que el modelo tiene un poder explicativo superior al de asumir que la temperatura no tiene nigun efecto sobre el tamaño de los huevos.

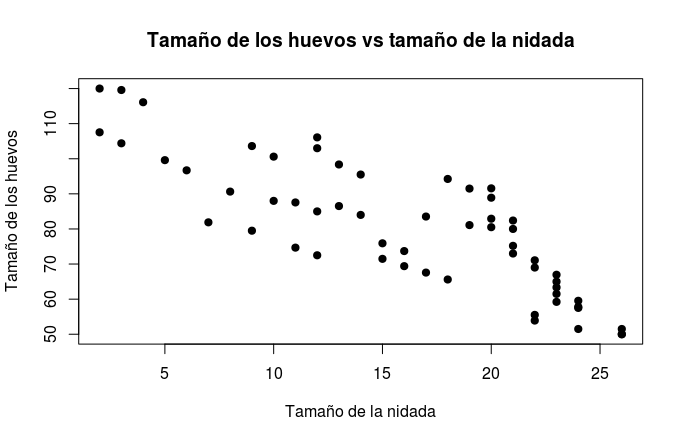
Segundo, respecto a la pendiente y el intercepto del modelo lineal, se puede evaluar el efecto que tiene cada variable sobre la explicación del modelo:

H0: βi= 0

Ha: βi≠ 0

**2.** Diagrama de causalidad

**3.** Exploracion grafica



Dada la forma de la distribución de los puntos y dado que el tamaño de la nidada es un conteo, es viable que los datos sigan una distribución de poisson.

**4.** Planteamiento del modelo estadistico

Podemos realizar un GML con función link log pues los datos siguen una distribución de poisson.

**5.** Evaluacion del modelo

> summary(mrp)

Call:

glm(formula = egg\_size ~ ClutchS, family = poisson, data = croc)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.8773 -0.8173 -0.2837 0.8348 2.3314

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 4.790381 0.034072 140.60 <2e-16 \*\*\*

ClutchS -0.026363 0.002084 -12.65 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 233.188 on 56 degrees of freedom

Residual deviance: 76.081 on 55 degrees of freedom

AIC: Inf

Number of Fisher Scoring iterations: 4

> anova(mrp,test="Chisq")

Analysis of Deviance Table

Model: poisson, link: log

Response: egg\_size

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL 56 233.188

ClutchS 1 157.11 55 76.081 < 2.2e-16 \*\*\*

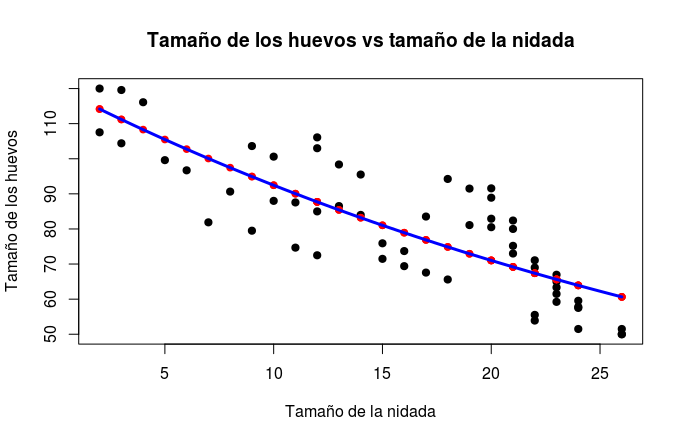
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Tenemos que el p-value asumiendo H0: β1= β2=0 es 2.2E-16 << alfa=0.05. Así pues, podemos afirmar que el modelo tiene un poder explicativo significativemente superior a comparación de un con parametros β1= β2=0.

Por otro lado, tenemos que para el intercepto y la pendiente, los p-values son de 2E-16. Así pues, podemos afirmar que tanto el intercepto como la pendiente explican significativamente mejor los datos a comparación de valores nulos (0) si se considera un alfa de 0.05.

**6.** Valores predichos para el modelo



En la gráfica anterior podemos ver en azul la función de probabilidad en función de l tamaño de la nidada. Los puntos rojos son los valores predichos por el GML.

**7.** Analisis de ODDS

> exp(cbind(OR= coef(mrp), confint(mrp)))

Waiting for profiling to be done...

OR 2.5 % 97.5 %

(Intercept) 120.3472321 112.5248730 128.604020

ClutchS 0.9739815 0.9700156 0.977973

El tamaño de la nidada explica en un 97% el tamaño de los huevos. Además, se sabe con confianza que explica en su mayoría el tamaño de los huevos pues tenemos un intervalo de confianza de un 95% con valores 0.970-0.977.

**8.** La interpretacion biologica de los resultados.

El tamaño de la nidada determina casi completamente el tamaño de los huevos dada una distribución de poisson descrita por la curva azul de la gráfica anterior.