

Examen #3

Estudiante: Roberto Vásquez Martínez

NUA: 424662

Problema 1*(Solución)*

(a) El ajuste del modelo binomial con la liga logit a los datos de proporción de leucemia respecto a la variable dosis arroja los siguientes resultados

Call:

```
glm(formula = cbind(leucemia, total - leucemia) ~ dosis, family = binomial,
     data = datos)
```

Deviance Residuals:

1	2	3	4	5	6
1.7700	-0.8191	-1.5681	-1.0209	-0.8518	1.4651

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.91157	0.26402	-14.815	< 2e-16 ***
dosis	0.57307	0.08521	6.726	1.75e-11 ***

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

(Dispersion parameter for binomial family taken to be 1)

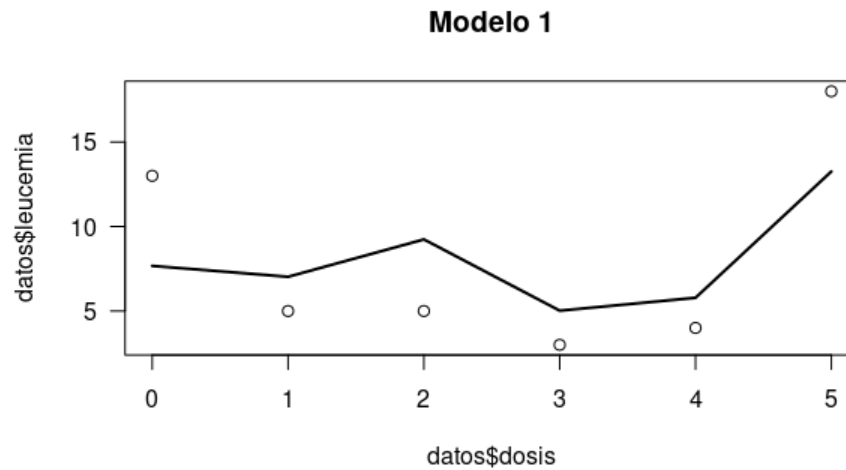
Null deviance: 54.351 on 5 degrees of freedom
 Residual deviance: 10.177 on 4 degrees of freedom
 AIC: 35.842

Number of Fisher Scoring iterations: 4

Usamos la prueba χ^2 para ver la bondad del ajuste, esto lo podemos hacer porque el parámetro de dispersión $\varphi = 1$, el resultado fue el siguiente

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			5	54.351	
dosis 1	44.174		4	10.177	3.005e-11 ***

La gráfica de la predicción se ve como sigue



(b) Ahora mostramos los resultados cuando utilizamos las variables dosis y dosis 2

```
glm(formula = cbind(leucemia, total - leucemia) ~ dosis + dosis2,
    family = binomial, data = datos)
```

Deviance Residuals:

1	2	3	4	5	6
0.04643	-0.18324	0.07499	0.22535	-0.19401	0.02917

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.38307	0.27259	-12.411	< 2e-16 ***
dosis	-0.41832	0.32872	-1.273	0.20318
dosis2	0.19440	0.06359	3.057	0.00223 **

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

(Dispersion parameter for binomial family taken to be 1)

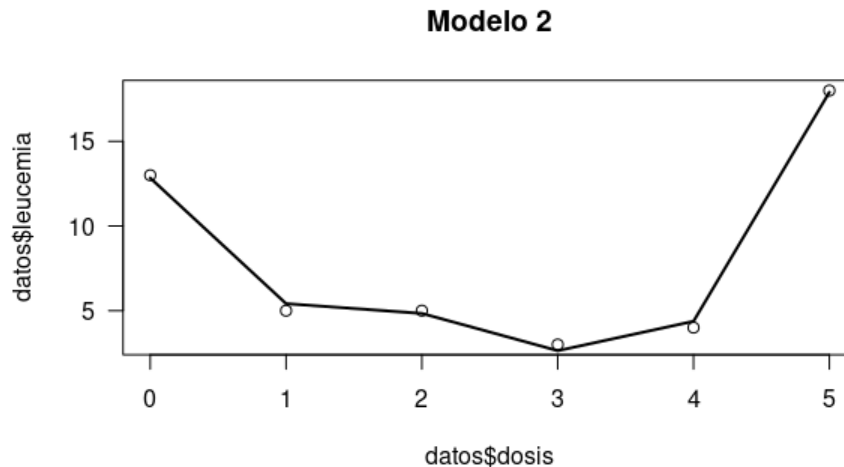
Null deviance: 54.35089 on 5 degrees of freedom
 Residual deviance: 0.13063 on 3 degrees of freedom
 AIC: 27.796

Number of Fisher Scoring iterations: 4

Y la correspondiente prueba χ^2

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			5	54.351	
dosis	1	44.174	4	10.177	3.005e-11 ***
dosis2	1	10.046	3	0.131	0.001527 **

Aquí tenemos un problema de significancia para la variable dosis, aunque resulta significativamente usar las variables dosis y dosis2, como lo muestra el siguiente gráfico



(c) Finalmente mostramos el modelo en el que incluimos a las 3 variables dosis, dosis2 y dosis3, el resultado fue el siguiente

Call:

```
glm(formula = cbind(leucemia, total - leucemia) ~ dosis + dosis2 +
    dosis3, family = binomial, data = datos)
```

Deviance Residuals:

1	2	3	4	5	6
0.03488	-0.15687	0.08481	0.21012	-0.22282	0.04073

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.379800	0.281190	-12.020	<2e-16 ***
dosis	-0.449365	0.746230	-0.602	0.547
dosis2	0.213066	0.407586	0.523	0.601
dosis3	-0.002545	0.054865	-0.046	0.963

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

(Dispersion parameter for binomial family taken to be 1)

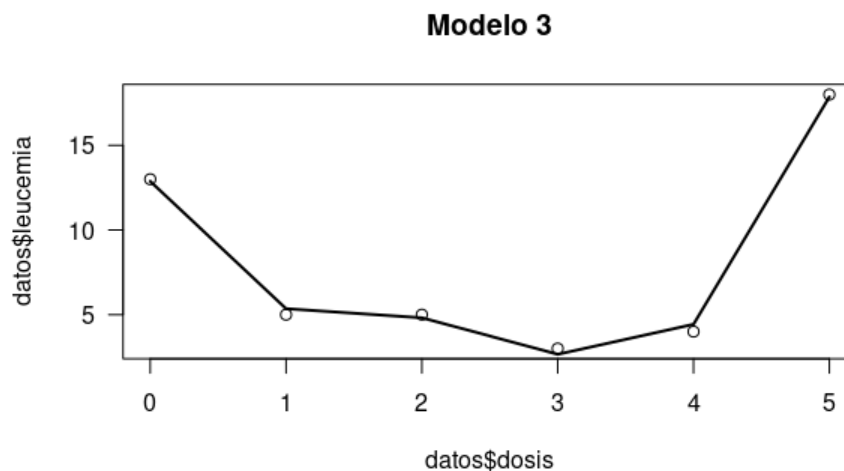
Null deviance: 54.35089 on 5 degrees of freedom
 Residual deviance: 0.12848 on 2 degrees of freedom
 AIC: 29.793

Number of Fisher Scoring iterations: 4

y las pruebas χ^2 para cada uno de los modelos es

	Df	Deviance	Resid.	Df	Resid.	Dev	Pr(>Chi)
NULL				5		54.351	
dosis	1	44.174		4	10.177	3.005e-11	***
dosis2	1	10.046		3	0.131	0.001527	**
dosis3	1	0.002		2	0.128	0.963014	

Aquí podemos notar que no resulta significativo tener un modelo con las 3 variables, puesto que tenemos problemas de significancia para cada parámetro y además la prueba χ^2 nos dice que no es significativo el modelo con las 3 variables respecto al modelo con 2, lo que se puede constatar con el siguiente resultado gráfico, muy parecido al del inciso anterior.



(d) A continuación mostramos los ajustes a los modelos faltantes, es decir, todas las posibles combinaciones restantes de las variables dosis, dosis2 y dosis3.

- Variables dosis y dosis3

```
Call:
glm(formula = cbind(leucemia, total - leucemia) ~ dosis + dosis3,
    family = binomial, data = datos)
```

Deviance Residuals:

1	2	3	4	5	6
0.20532	-0.47120	-0.05888	0.35393	0.08052	-0.05266

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.428390	0.270856	-12.658	< 2e-16 ***
dosis	-0.080007	0.228709	-0.350	0.72648
dosis3	0.025902	0.008626	3.003	0.00268 **

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 54.35089 on 5 degrees of freedom
 Residual deviance: 0.40218 on 3 degrees of freedom
 AIC: 28.067

Number of Fisher Scoring iterations: 4

```
> anova(model13, test = 'Chisq')
Analysis of Deviance Table
```

Model: binomial, link: logit

Response: cbind(leucemia, total - leucemia)

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			5	54.351	
dosis	1	44.174	4	10.177	3.005e-11 ***
dosis3	1	9.775	3	0.402	0.001769 **

■ Variables dosis2 y dosis3

```
glm(formula = cbind(leucemia, total - leucemia) ~ dosis2 + dosis3,
    family = binomial, data = datos)
```

Deviance Residuals:

	1	2	3	4	5	6
	0.34118	-0.51986	-0.16046	0.28459	0.03657	-0.02934

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.46765	0.24964	-13.891	<2e-16 ***
dosis2	-0.02157	0.12552	-0.172	0.864
dosis3	0.02727	0.02464	1.107	0.268

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 54.3509 on 5 degrees of freedom
 Residual deviance: 0.4956 on 3 degrees of freedom
 AIC: 28.161

Number of Fisher Scoring iterations: 4

```
> anova(model23, test = 'Chisq')
Analysis of Deviance Table
```

Model: binomial, link: logit

Response: cbind(leucemia, total - leukemia)

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			5	54.351	
dosis2	1	52.568	4	1.783	4.157e-13 ***
dosis3	1	1.288	3	0.496	0.2565

■ Variable dosis2

```
Call:
glm(formula = cbind(leucemia, total - leukemia) ~ dosis2, family = binomial,
    data = datos)
```

Deviance Residuals:

	1	2	3	4	5	6
	0.8645	-0.4138	-0.5737	-0.3053	-0.5649	0.3507

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.62345	0.21881	-16.560	< 2e-16 ***
dosis2	0.11656	0.01512	7.711	1.25e-14 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 54.3509 on 5 degrees of freedom
 Residual deviance: 1.7831 on 4 degrees of freedom
 AIC: 27.448

Number of Fisher Scoring iterations: 4

```
> anova(model2, test = 'Chisq')
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(leucemia, total - leukemia)

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			5	54.351	
dosis2	1	52.568	4	1.783	4.157e-13 ***

■ Variable dosis3

Call:

```
glm(formula = cbind(leucemia, total - leukemia) ~ dosis3, family = binomial,
     data = datos)
```

Deviance Residuals:

1	2	3	4	5	6
0.42731	-0.50142	-0.22263	0.19528	-0.05952	0.00916

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.492788	0.204307	-17.096	< 2e-16 ***
dosis3	0.023072	0.002909	7.932	2.16e-15 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 54.35089 on 5 degrees of freedom
 Residual deviance: 0.52534 on 4 degrees of freedom
 AIC: 26.19

Number of Fisher Scoring iterations: 4

```
> anova(model3, test = 'Chisq')
Analysis of Deviance Table
```

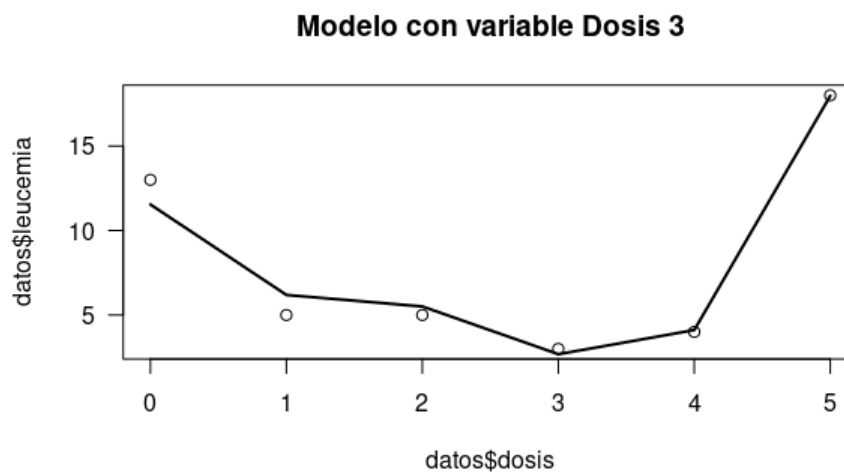
Model: binomial, link: logit

Response: cbind(leucemia, total - leucemia)

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			5	54.351	
dosis3 1	53.826	4	0.525	2.191e-13 ***	

De aquí se puede ver que los modelos con dos variables explicativas tienen problema de significancia en la variable dosis, por lo que esta podríamos descartar. Entonces nos queda comparar los modelos con dosis2, y dosis3, el modelo con dosis3 resulta tener un AIC relativamente mejor, y el ajuste es el siguiente solo considerando la variable dosis 3



que nos resulta mejor por simplicidad y al tener un buen ajuste.

□

Problema 2*(Solución)*

(a) Los resultados que tenemos respecto a la liga identidad son los siguientes

```
all:
glm(formula = Y ~ X, family = poisson(link = "identity"), data = datos)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.7019	-0.3377	-0.1105	0.2958	0.7184

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	7.4516	0.8841	8.428	< 2e-16 ***
X	4.9353	1.0892	4.531	5.86e-06 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 18.4206 on 8 degrees of freedom
 Residual deviance: 1.8947 on 7 degrees of freedom
 AIC: 40.008

Number of Fisher Scoring iterations: 3

```
> anova(model_id, test = 'Chisq')
```

Analysis of Deviance Table

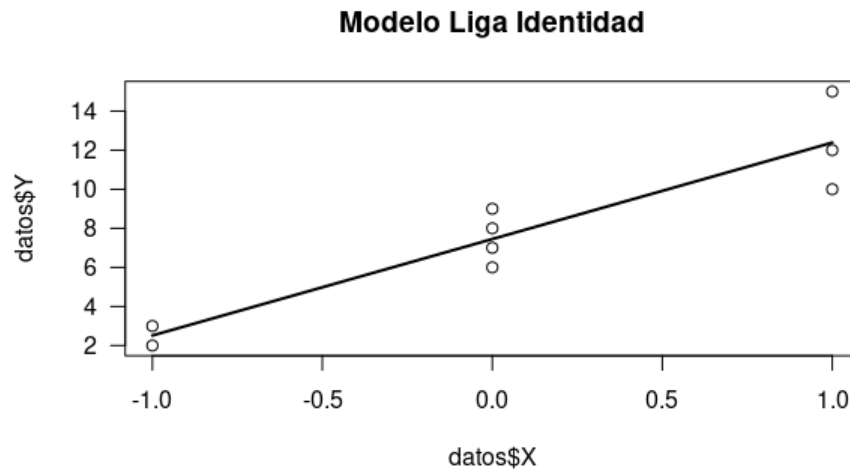
Model: poisson, link: identity

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			8	18.4206	
X	1	16.526	7	1.8947	4.799e-05 ***

Y el ajuste gráfico que tenemos es el siguiente



(b) Los resultados obtenidos con la liga logaritmo son

Call:

```
glm(formula = Y ~ X, family = poisson(link = "log"), data = datos)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.8472	-0.2601	-0.2137	0.5214	0.8788

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.8893	0.1421	13.294	< 2e-16 ***
X	0.6698	0.1787	3.748	0.000178 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 18.4206 on 8 degrees of freedom
 Residual deviance: 2.9387 on 7 degrees of freedom
 AIC: 41.052

Number of Fisher Scoring iterations: 4

```
> anova(model_log, test = 'Chisq')
```

Analysis of Deviance Table

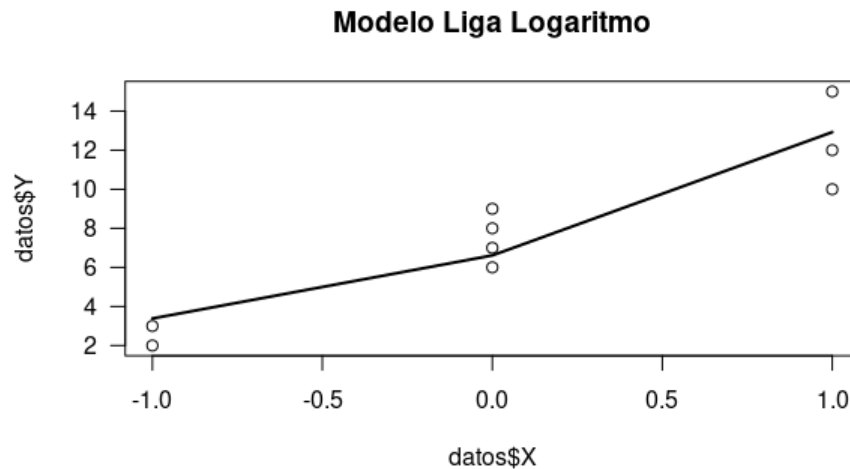
Model: poisson, link: log

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			8	18.4206	
X	1	15.482	7	2.9387	8.33e-05 ***

Y el ajuste gráfico resulto en lo siguiente



(c) Por un lado tenemos que la deviance del modelo con la liga identidad es 16.526 mientras que la deviance de la liga canónica es 15.428, por lo que a esto respecta resulta ser un mejor modelo el que tiene la liga logarítmica.

Sin embargo si nos fijamos el en AIC de ambos modelos el correspondiente a la liga identidad es 40.008 mientras que el correspondiente a la liga logarítmica es 41.052 por lo que en ese caso resultaría mejor modelo el que tiene la liga identidad.

(d) Podemos ver a partir del inciso anterior que no hay evidencia contundente a favor de uno de los modelos. □