Práctica 7

Maximiliano Vaca Montejano
364897
maximiliano.vaca@uabc.edu.mx

Profesor: Andrés García Medina
andres.garcia.medina@uabc.edu.mx
Fecha de entrega: viernes 17 de mayo, 2024 (23:59).

Nota técnica: Los datos se adjuntan en el archivo $AIDS.csv$. Considere los valores iniciales $\beta = [2, 1, 0]$. Una tolerancia de $tol = 1 \times 10 - 6$ (norma minima del paso Δ). Un número máximo de iteraciones $m = 100$. Recuerde convertir a la variable independiente (year) a incrementos

---

```
data <- read.csv('datos/AIDS.csv')
#colnames(data) <- c("edad", "CHD")

#head(data, 10)
#data
```

```
#convertir a la variable independiente (year) a incrementos
data$year <- data$year - 1980

data
```

A data.frame: 13 × 2

year	cases
<dbl>	<int>
1	12
2	14
3	33
4	50
5	67
6	74
7	123
8	141
9	165
10	204
11	253
12	246
13	240

Ejercicio 1

El siguiente conjunto de datos muestra el número de casos nuevos de SIDA en Bélgica (en los 80's)1. La pregunta cientificamente interesante es si los datos proporcionan alguna evidencia de que el aumento en la tasa subyacente de generación de nuevos casos se está desacelerando.

No description has been provided for this image

- Ajuste un modelo de regresión Poisson con ayuda de la función glm asumiendo que las observaciones $(y_1, x_1), \ldots, (y_n, x_n)$, son independientes, y considerando términos cuadráticos para los regresores $x_i = (1, t_i, t_i^2)^T$. (30 pts)

```
#regresión Poisson con glm
model <- glm(cases ~ year + I(year^2), data = data, family = poisson)

summary(model)
```

```
Call:
glm(formula = cases ~ year + I(year^2), family = poisson, data = data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.45903  -0.64491   0.08927   0.67117   1.54596

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.901459   0.186877  10.175  < 2e-16 ***
year         0.556003   0.045780  12.145  < 2e-16 ***
I(year^2)   -0.021346   0.002659  -8.029 9.82e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 872.2058  on 12  degrees of freedom
Residual deviance:   9.2402  on 10  degrees of freedom
AIC: 96.924

Number of Fisher Scoring iterations: 4
```

```
summary(model)$coefficients
```

A matrix: 3 × 4 of type dbl

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.90145858	0.186877465	10.174895	2.566767e-24
year	0.55600327	0.045779750	12.145179	6.084915e-34
I(year^2)	-0.02134627	0.002658609	-8.029112	9.818083e-16

todos los estimadores son significativos y con un error estandar no muy grande

- Encuentre el valor esperado de las betas y su desviación estándar por medio de optimización directa (IRLWLS) y compare sus resultados con los que arroja la función glm. (40 pts)

- re: Considere los valores iniciales $\beta = [2, 1, 0]$. Una tolerancia de $tol = 1 \times 10 - 6$ (norma minima del paso Δ). Un número máximo de iteraciones $m = 100$.

```
#condiciones
b0 = c(2, 1, 0); b0 = t(t(b0))
tol = 1e-6
m = 100

#definimos X y Y
Y = data$cases
#X = cbind(1, data$year, data$year^2) # matriz de diseño
#matriz de diseño
X = model.matrix(model)

print("X:"); head(X, 5)

#X transpuesta
#Xt = t(X)

print("dim X:"); dim(X)
print("dim b0:"); dim(b0)

# u --> a = lambda en poisson
a = exp(X%*%b0)

print("a:"); dim(a); a#head(a, 5)

X
```

[1] "X:"

A matrix: 5 × 3 of type dbl

	(Intercept)	year	I(year^2)
1	1	1	1
2	1	2	4
3	1	3	9
4	1	4	16
5	1	5	25

[1] "dim X:"
13 · 3

[1] "dim b0:"
3 · 1

[1] "a:"
13 · 1

A matrix: 13 × 1 of type dbl

1	2.008554e+01
2	5.459815e+01
3	1.484132e+02
4	4.034288e+02
5	1.096633e+03
6	2.980958e+03
7	8.103084e+03
8	2.202647e+04
9	5.987414e+04
10	1.627548e+05
11	4.424134e+05
12	1.202604e+06
13	3.269017e+06

A matrix: 13 × 3 of type dbl

	(Intercept)	year	I(year^2)
1	1	1	1
2	1	2	4
3	1	3	9
4	1	4	16
5	1	5	25
6	1	6	36
7	1	7	49
8	1	8	64
9	1	9	81
10	1	10	100
11	1	11	121
12	1	12	144
13	1	13	169

definimos Λ

```
#v = u*(1-u)

#lambda mayuscula en poisson
A = diag(as.vector(a))

dim(A)

#a
A
```

13 · 13

A matrix: 13 × 13 of type dbl

20.08554	0.00000	0.0000	0.0000	0.000	0.000	0.000	0.00	0.00	0.0	0.0	0	0
0.00000	54.59815	0.0000	0.0000	0.000	0.000	0.000	0.00	0.00	0.0	0.0	0	0
0.00000	0.00000	148.4132	0.0000	0.000	0.000	0.000	0.00	0.00	0.0	0.0	0	0
0.00000	0.00000	0.0000	403.4288	0.000	0.000	0.000	0.00	0.00	0.0	0.0	0	0
0.00000	0.00000	0.0000	0.0000	1096.633	0.000	0.000	0.00	0.00	0.0	0.0	0	0
0.00000	0.00000	0.0000	0.0000	0.000	2980.958	0.000	0.00	0.00	0.0	0.0	0	0
0.00000	0.00000	0.0000	0.0000	0.000	0.000	8103.084	0.00	0.00	0.0	0.0	0	0
0.00000	0.00000	0.0000	0.0000	0.000	0.000	0.000	22026.47	0.00	0.0	0.0	0	0
0.00000	0.00000	0.0000	0.0000	0.000	0.000	0.000	0.00	59874.14	0.0	0.0	0	0
0.00000	0.00000	0.0000	0.0000	0.000	0.000	0.000	0.00	0.00	162754.8	0.0	0	0
0.00000	0.00000	0.0000	0.0000	0.000	0.000	0.000	0.00	0.00	0.0	442413.4	0	0
0.00000	0.00000	0.0000	0.0000	0.000	0.000	0.000	0.00	0.00	0.0	0.0	1202604	0
0.00000	0.00000	0.0000	0.0000	0.000	0.000	0.000	0.00	0.00	0.0	0.0	0	3269017

para cada b

```
i = 0

#b <- c(b0) # no

b <- list(b0)
#as.vector(b)

while ((m > tol) & (i < 100)) {
    #print(i)
    i <- i + 1

    b1 <- b0 + solve(t(X)%*%A%*%X) %*% t(X) %*%(Y-a)

    m <- sqrt(sum((b1 - b0)^2))
    b0 <- b1

    #nueva μ
    #u = 1 / (1 + exp(- X%*%b0))
    #nueva W
    #v = u*(1-u)
    #W = diag(as.vector(v))

    #nueva a
    a = exp(X%*%b0)

    #nueva A
    A = diag(as.vector(a))


    b <- c(b, list(b0))
}

#b
b <- as.matrix(b)
```

```
head(b)
```

A matrix: 6 × 1

2, 1, 0
1.1211292668, 0.9795371949, 0.0008592407
0.44204553, 0.92534211, 0.00313429
0.254882683, 0.788217778, 0.008885712
1.06761280, 0.48361586, 0.02162036
2.86100041, 0.02436385, 0.04046591

utilizamos el ultimo resultado de la iteración: $\overline{\beta_0}$

```
b0
```

A matrix: 3 × 1 of type dbl

(Intercept)	1.90145858
year	0.55600327
I(year^2)	-0.02134627

Desviacion estandar

la calculamos similar a la logistica, cambiando W por Λ

```
#revisamos dimensiones
dim(t(X)); dim(A); dim(X)
```

3 · 13

13 · 13

13 · 3

$$

```
V = solve(t(X)%*%A%*%X)
print("V: "); V
print("diag(V)"); diag(V)
print("sqrt(diag(V))"); sqrt( diag( V ) )

print("glm model:")
summary(model)$coefficients
```

[1] "V: "

A matrix: 3 × 3 of type dbl

	(Intercept)	year	I(year^2)
(Intercept)	0.0349231870	-0.008158846	4.394231e-04
year	-0.0081588462	0.002095786	-1.194640e-04
I(year^2)	0.0004394231	-0.000119464	7.068204e-06

[1] "diag(V)"
(Intercept): 0.0349231869815749 **year:** 0.00209578551918789 **I(year^2):** 7.06820357237233e-06

[1] "sqrt(diag(V))"
(Intercept): 0.186877465151834 **year:** 0.0457797500996662 **I(year^2):** 0.00265860933052834

[1] "glm model:"

A matrix: 3 × 4 of type dbl

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.90145858	0.186877465	10.174895	2.566767e-24
year	0.55600327	0.045779750	12.145179	6.084915e-34
I(year^2)	-0.02134627	0.002658609	-8.029112	9.818083e-16

son muy similares

- También es posible encontrar un intervalo de confianza para las estimaciones (e incluso para las predicciones si tuviéramos nuevos datos). Grafique los valores originales junto con la estimación (solución) considerando dos desviaciones estándar como intervalo de confianza. La idea es que reproduzca la figura . (20 pts)

No description has been provided for this image

```
desv = sqrt( diag( V ) )
desv = as.matrix(desv)
dim(desv)
desv
```

3 · 1

A matrix: 3 × 1 of type dbl

(Intercept)	0.186877465
year	0.045779750
I(year^2)	0.002658609

```
b0
dim(b0)
```

A matrix: 3 × 1 of type dbl

(Intercept)	1.90145858
year	0.55600327
I(year^2)	-0.02134627

3 · 1

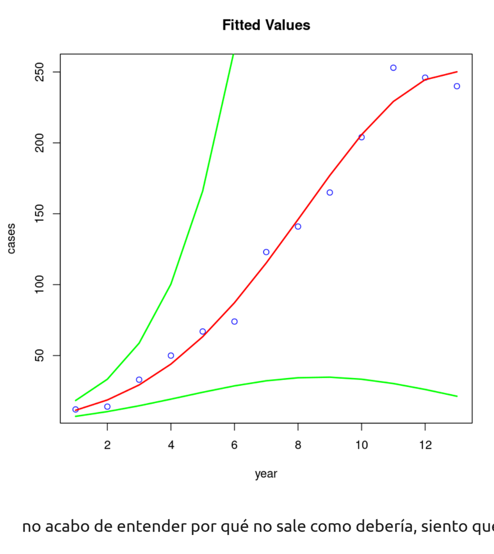
```
b0 + 2*desv
```

A matrix: 3 × 1 of type dbl

(Intercept)	2.27521351
year	0.64756277
I(year^2)	-0.01602905

```
#valores originales
plot(data$year, data$cases, col = "blue",  xlab = "year", ylab = "cases", main="Fitted Values")
#valores estimados
lines(data$year, exp(X%*%b0), col = "red", lwd = 2)

lines(data$year, exp(X%*%(b0 + 2*desv)), col = "green", lwd = 2)
lines(data$year, exp(X%*%(b0 - 2*desv)), col = "green", lwd = 2)
```



no acabo de entender por qué no sale como debería, siento que todo está bien

```
b0
```

A matrix: 3 × 1 of type dbl

(Intercept)	1.90145858
year	0.55600327
I(year^2)	-0.02134627

- Discuta las implicaciones de los resultados del caso de estudio particular ¿Como interpretaría el valor de β1? (10 pts)

Al interpretar el valor de $\beta_1 = 0.55600327$, este nos indica la velocidad del proceso, que los casos de sida por cada año que pasa, aumentan en un $e^{\beta_1} \approx 1.74368949888$.

Mientras tanto, el valor de $\beta_1 = -0.02134627$, se puede intepretar como la aceleración o tasa de aumento. Al tener un valor negativo, nos indica que el crecimiento indicado por β_1 no es indefinido, es moderado por β_2 y está desacelerando.