

Tarea 2 - Métodos Lineales Generalizados

Gerardo Rocha Benigno

2025-03-03

Problema

Un investigador desea evaluar la relación entre el salario anual de trabajadores de una compañía de nivel medio y alto (Y , en miles de dólares) y el índice de calidad de trabajo (X_1), número de años de experiencia (X_2) y el índice de éxito en publicaciones (X_3). La muestra consiste de 24 trabajadores. Realiza un análisis Bayesiano completo de los datos y obtén las predicciones de salarios para 3 nuevos empleados con variables explicativas:

$$x'_{1F} = (5.4, 17, 6.0), x'_{2F} = (6.2, 12, 5.8), x'_{3F} = (6.4, 21, 6.1)$$

Estadísticos descriptivos de la muestra

Estadístico	Y	X1	X2	X3
Min	30.10	3.10	5.00	3.50
Q1	35.70	4.42	17.25	5.00
Mediana	38.85	5.40	25.00	6.00
Media	39.50	5.36	24.96	5.99
Q3	42.92	6.28	33.25	7.00
Max	52.90	8.00	47.00	8.30
SD	5.47	1.29	11.22	1.30

Gráficas de dispersión entre variables explicativas y variable a predecir



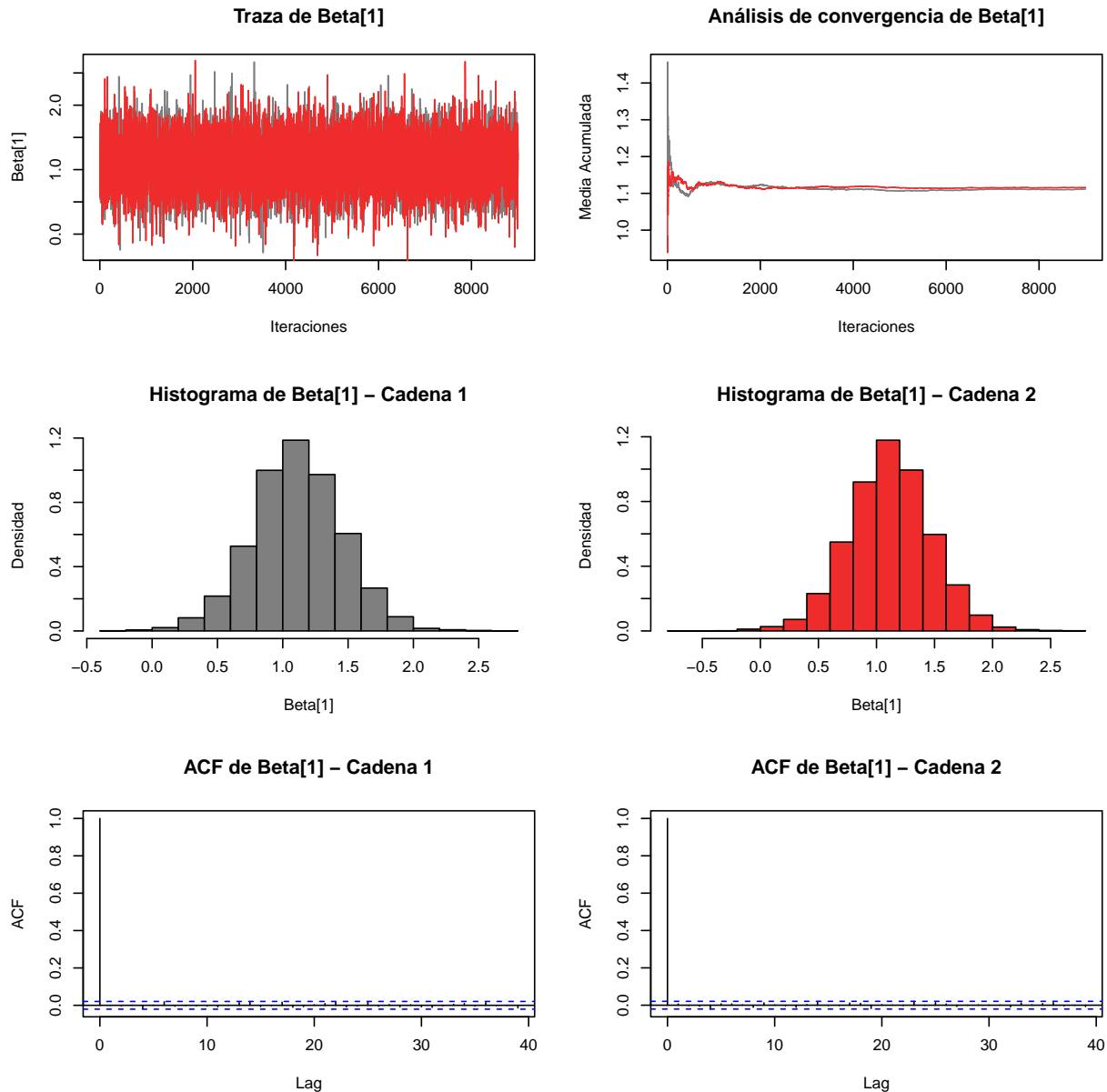
Análisis bayesiano

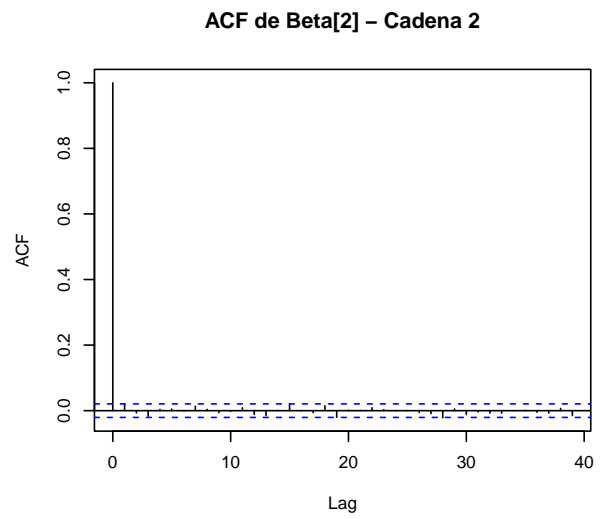
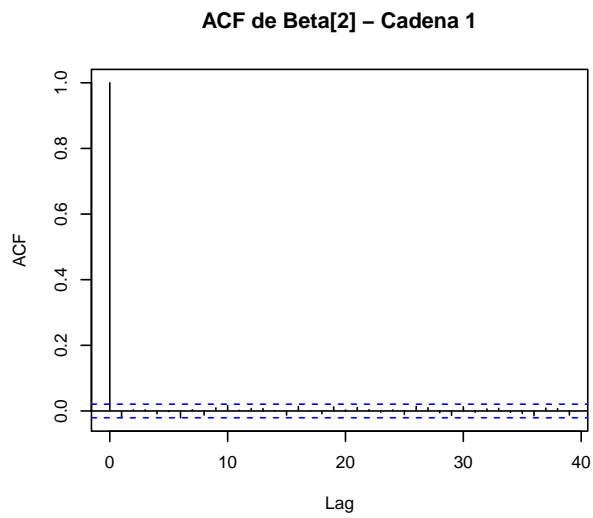
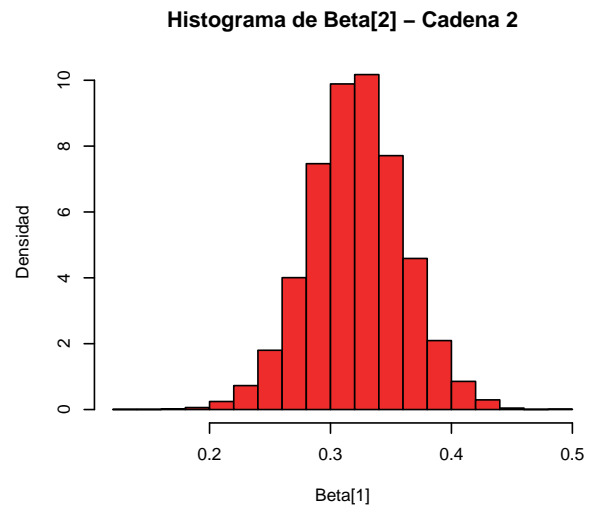
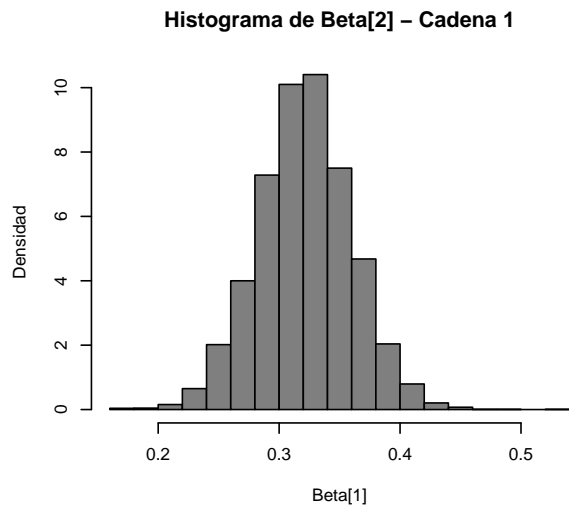
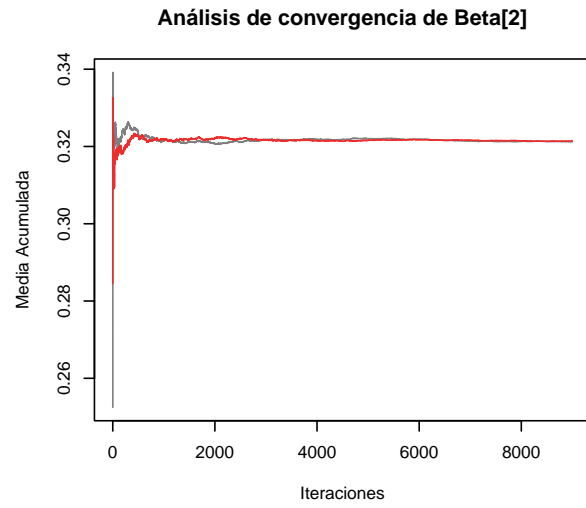
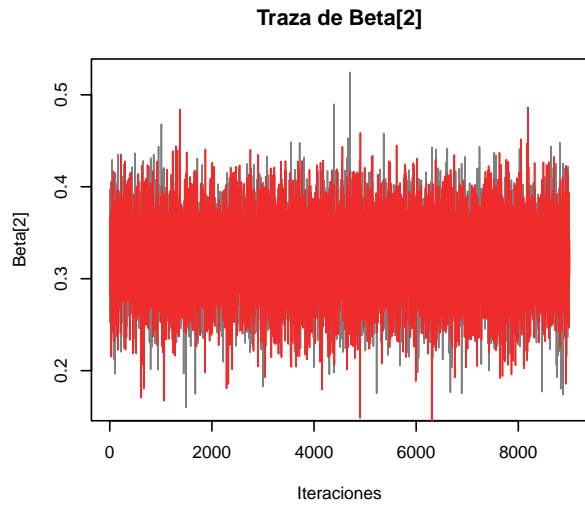
Para el análisis bayesiano suponemos lo siguiente: - Que $y_i \sim N(\mu_i, \tau)$, es decir, que los salarios siguen una distribución normal con media distinta para cada valor de y_i y precisión única. - Que $\mu_i = \alpha + \beta_1 * x_{i1} + \beta_{i2} * x_2 + \beta_3 * x_{i3}$, es decir, que la media condicional del salario es una combinación lineal de la calidad del trabajo, de los años de experiencia y del éxito de las publicaciones. - Las distribuciones iniciales son no informativas: $\alpha \sim N(0, 0.001)$, $\beta_i \sim N(0, 0.001)$ y $\tau \sim \text{Gamma}(0.001, 0.001)$.

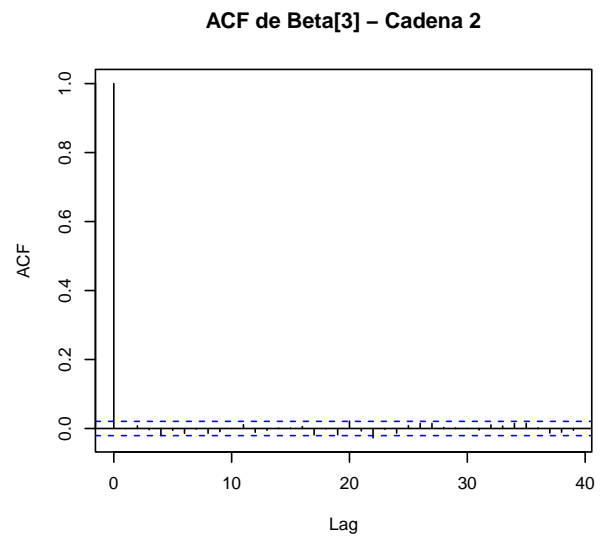
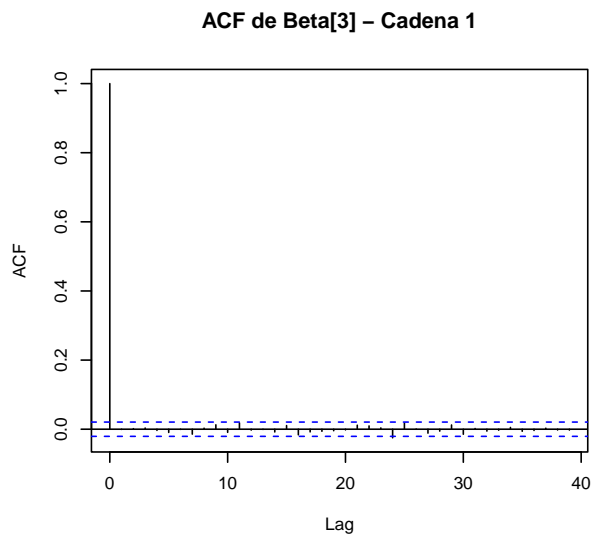
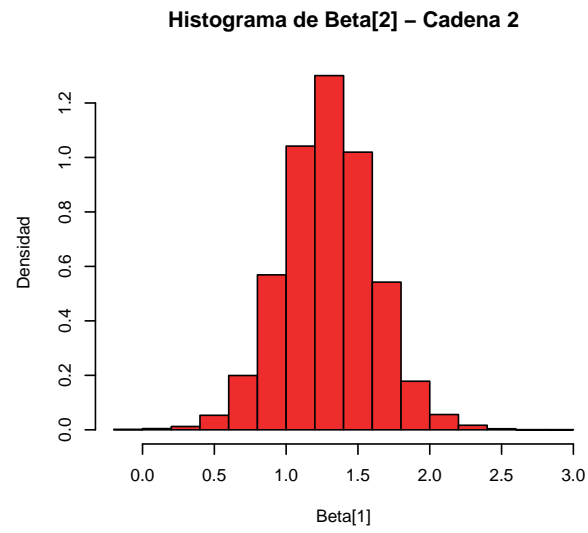
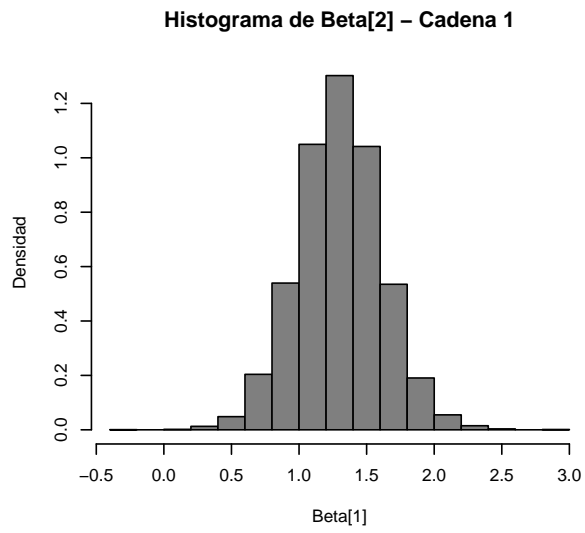
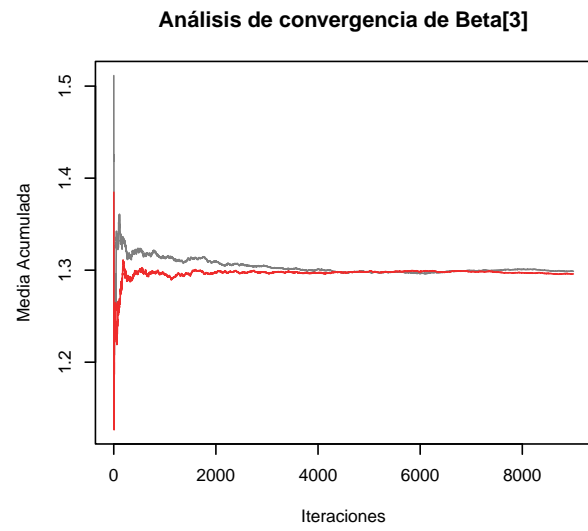
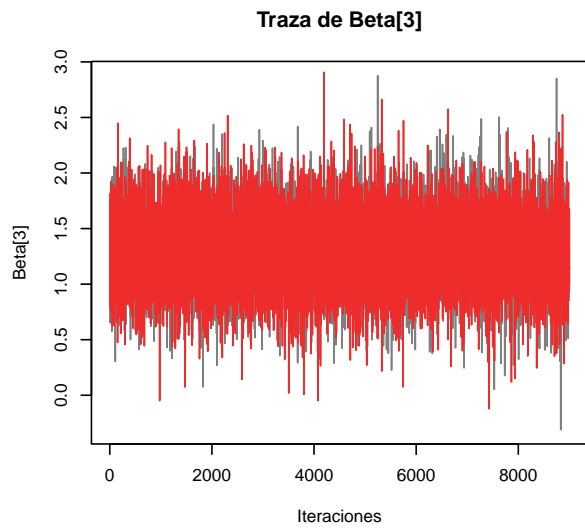
Con lo anterior, para realizar inferencia queremos encontrar las distribuciones posteriores de $\beta_1, \beta_2, \beta_3$.

1. Como primer paso, checamos si las cadenas de cada β convergen:

1.1 El análisis gráfico de cada β revela que las MCMC convergen y que no existe autocorrelación entre las muestras.



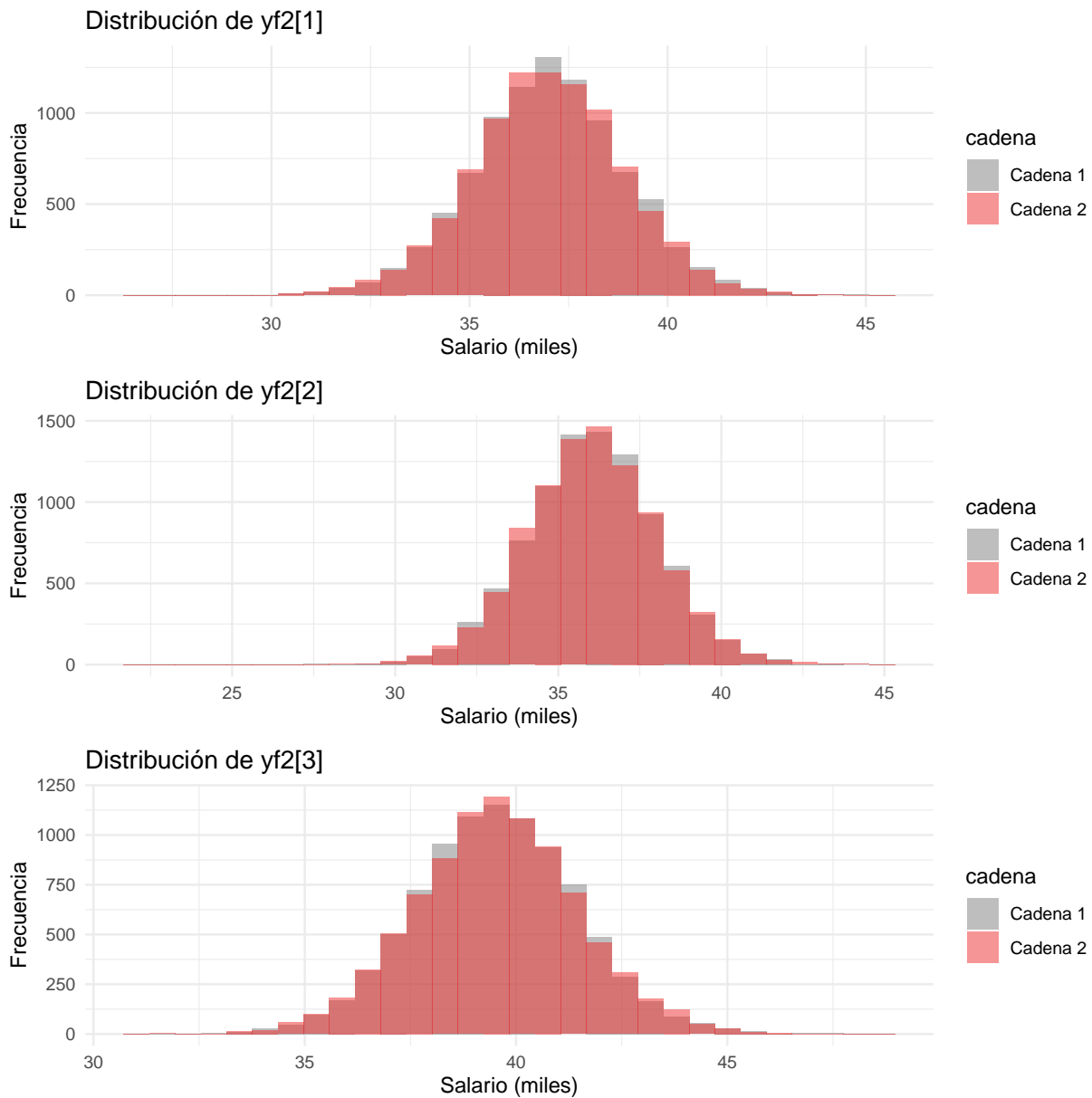




1.2 La siguiente tabla muestra la media posterior de cada β , los intervalos creíbles y el p-value asociado. La tabla muestra que los tres coeficientes son significativos y que la variable con mayor impacto en el salario es la éxito en las publicaciones.

Coef	mean	2.5%	97.5%	prob
beta[1]	1.1137118	0.4210466	1.7989900	0.0020555556
beta[2]	0.3213472	0.2438174	0.3982416	0.0000000000
beta[3]	1.2973076	0.6884508	1.9193268	0.0002222222

1.3 Por último, obtenemos las predicciones de $x'_{1F} = (5.4, 17, 6.0)$, $x'_{2F} = (6.2, 12, 5.8)$, $x'_{3F} = (6.4, 21, 6.1)$. En este caso solo graficamos las distribuciones de cada predicción.



Los valores predichos de las tres observaciones son: 36.99, 36.01 y 39.52 mil pesos.

Coef	mean	2.5%	97.5%	prob
yf2[1]	36.99651	33.16536	40.73141	0
yf2[2]	36.02925	32.10960	39.99119	0
yf2[3]	39.52997	35.72664	43.41718	0