



Universidad Nacional Autónoma de México

# Proyecto Final

Inferencia Bayesiana

¿NGEL

SHADANNA



**iimas**

DR. EDUARDO GUTIÉRREZ PEÑA

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

# Índice

<b>Introducción</b>	<b>2</b>
<b>Ejercicio 1: Modelo Binomial-Beta</b>	<b>2</b>
<b>Ejercicio 2: Modelo Lineal Generalizado</b>	<b>3</b>
<b>Ejercicio 3: Modelo Jerárquico</b>	<b>5</b>
<b>Ejercicio 4: Modelos</b>	<b>6</b>
Comparación de modelos . . . . .	6
Discusión de resultados . . . . .	6
<b>Comentarios finales</b>	<b>7</b>
<b>Referencias</b>	<b>8</b>

## Introducci3n

En la actualidad, una de las principales causas de muerte es por la enfermedad conocida como c3ncer. El c3ncer comienza en una c3lula normal que cambia a una c3lula neopl3stica<sup>1</sup> a trav3s de varias mutaciones en varios genes a lo largo de mucho tiempo, podr3an ser a3os, de estar expuesto a un agente carcinog3nico<sup>2</sup>. No obstante, las mutaciones inducidas por los carcin3genos no son la 3nica v3a que afecta a la c3lula, sino que a lo largo de cada divisi3n celular se producen errores espont3neos en cada duplicaci3n y los mismos se van acumulando constituyendo un factor intr3nseco de riesgo (Mart3n de Civetta y Civetta, 2011). Por lo cual, es de suma importancia estudiar la cura para esta enfermedad.

En este contexto, el presente trabajo analizar3 desde una perspectiva bayesiana de un experimento de un tipo de tumor en un grupo de ratas, dadas diferentes dosis de una droga. En otras palabras, estudiar la tasa a la que el riesgo de tumor crece o decrece como funci3n de la dosis.

Para ello, se examinar3n tres perspectivas de acuerdo al tipo de informaci3n inicial, despu3s se har3 la una comparaci3n entre modelos y por 3ltimo unos comentarios finales.

## Ejercicio 1: Modelo Binomial-Beta

Con el prop3sito de estudiar la relaci3n entre la dosis y la respuesta, se tienen los datos del experimento en el cuadro 1, donde  $x$  representa el nivel de la dosis, mientras que  $n_x$  y  $y_x$  denotan respectivamente, el n3mero de ratas tratadas y el n3mero de ratas que presentan tumor en cada nivel ( $x = 0, 1, 2$ ).

$x$	$n_x$	$y_x$
0	14	4
1	34	4
2	34	2

Tabla 1: Datos

Sea  $\pi_x$  la probabilidad de que una rata en el grupo  $x$  desarrolle un tumor. Entonces, se considera el modelo

$$Y_x \sim \text{Bin}(\pi_x, n_x) \quad (x = 0, 1, 2).$$

Dado que las investigadoras est3n interesadas en la forma como var3a  $\pi_x$  en funci3n de la dosis  $x$ , propusieron el modelo

$$\text{logit}(\pi_x) = \alpha + \beta x \quad (x = 0, 1, 2).$$

El par3metro de inter3s para las investigadoras es la pendiente  $\beta$ , pero no cuentan con informaci3n inicial sobre su valor.

Entonces, se realizar3 un resumen de la distribuci3n final de  $\beta$  suponiendo una distribuci3n inicial no informativa en la que  $\alpha$  y  $\beta$  se asumen independientes, con  $\alpha \sim N(0, 1000)$  y  $\beta \sim N(0, 1000)$ ; esto es, con media 0 y varianza 1000.

---

<sup>1</sup>C3lula con una multiplicaci3n o crecimiento anormal en un tejido del organismo.

<sup>2</sup>Agente capaz de causar c3ncer.

## Ejercicio 2: Modelo Lineal Generalizado

Dado que el tamao de las muestras en el problema anterior es muy pequeo, y en vista de la falta de informacin inicial, las investigadoras se dieron a la tarea de buscar informacin relevante en la literatura. Como producto de esta labor, encontraron datos de 10 estudios similares con ratas de la misma cepa. Desafortunadamente, todos estos datos correspondan a *controles*; es decir, ratas a las que no se les aplic la droga. Los datos se presentan en la Tabla 2a. Aqu  $n_{0,i}$  y  $y_{0,i}$  denotan, respectivamente, el nmero total de ratas y el nmero de ratas que presentaron un tumor en el  $i$ -simo estudio ( $i = 1, 2, \dots, 10$ ).

<i>Estudio</i> $i$	$n_{0,i}$	$y_{0,i}$
1	10	1
2	13	2
3	48	10
4	19	5
5	20	0
6	18	0
7	25	2
8	49	5
9	48	9
10	19	4

Tabla 2: Tabla 2a

No satisfechas con estos datos, las investigadoras siguieron buscando trabajos recientes (no publicados). Finalmente encontraron dos reportes muy relevantes, de donde extrajeron los siguientes datos:

$x$	$n_{x,11}$	$y_{x,11}$
0	7	3
1	16	5
2	18	2

Tabla 3: Tabla 2b

$x$	$n_{x,12}$	$y_{x,12}$
0	5	2
1	11	1
2	9	0

Tabla 4: Tabla 2c

En vista de que para los datos de la Tabla 2a slo se recab informacin de controles, el nivel de la dosis es  $x = 0$  en todos esos casos. Por lo tanto el modelo que propusieron para esos datos es

$$Y_x \sim \text{Bin}(\pi_{0,1}, n_{0,i}), \quad i = 1, 2, \dots, 10,$$

donde

$$\text{logit}(\pi_{0,i}) = \alpha_i, \quad i = 1, 2, \dots, 10.$$

Por otra parte, para los datos de las Tablas 2b y 2c (Estudios 11 y 12), las investigadoras supusieron un modelo de la misma forma que el del problema 1, es decir:

$$Y_x \sim \text{Bin}(\pi_{x,i}, n_{x,i}) \quad (x = 0, 1, 2); \quad i = 11, 12,$$

con

$$\text{logit}(\pi_x) = \alpha_i + \beta_i x \quad (x = 0, 1, 2); \quad i = 11, 12.$$

Para simplificar el análisis en esta etapa, las investigadoras decidieron considerar todos estos estudios suficientemente similares como para suponer que los datos de las Tablas 1, 2a, 2b y 2c *proviene de un solo experimento*, de manera que  $\alpha_1 = \alpha_2 = \dots = \alpha_{12} = \alpha$  y  $\beta_{11} = \beta_{12} = \beta$ .

Utilizando la misma distribución inicial que en el Ejercicio 1, se proporciona un resumen de la distribución final de  $\beta$ .

### Ejercicio 3: Modelo Jerárquico

Poco tiempo después, una de las investigadoras tuvo la oportunidad de asistir a un curso de Análisis Bayesiano de Modelos Jerárquicos y convenció al resto del equipo de que esa es la manera más apropiada de analizar los datos con lo que contaban. Específicamente, dado que todos los estudios eran similares, consideraron que podían utilizar los 12 estudios que encontraron en la literatura para complementar la información de su experimento original (ver Tabla 1).

Las investigadoras supusieron entonces que los parámetros  $\alpha_1, \alpha_2, \dots, \alpha_{12}$  eran intercambiables, con distribución común  $N(\alpha^*, \sigma_\alpha^2)$ , y también que los parámetros  $\beta, \beta_{11}, \beta_{12}$  eran intercambiables con distribución común  $N(\beta^*, \sigma_\beta^2)$ . Finalmente, tanto para  $\alpha^*$  como para  $\beta^*$  supusieron una distribución  $N(0, 100)$ , mientras que para  $\tau_\alpha = 1/\sigma_\alpha^2$  y  $\tau_\beta = 1/\sigma_\beta^2$  consideraron una distribución  $\text{Gamma}(0.01, 0.01)$ .

Se propone un resumen de la distribución final de  $\beta$  (la correspondiente al Ejercicio 1) bajo estas condiciones.

## **Ejercicio 4: Modelos**

**Comparaci3n de modelos**

**Discusi3n de resultados**

## Comentarios finales



## Referencias

- Mart?n de Civetta MT y Civetta JD.(2011). Carcinog?nesis. Salud Publica Mex;53:405-414.