

## Modelo de mapeo de enfermedades para los datos oficiales de COVID19

LUIS E. NIETO BARAJAS  
Departamento de Estadística, ITAM  
lnieto@itam.mx

El objetivo de este modelo es estimar las tasas de incidencia o riesgos relativos de la enfermedad COVID19 para los 32 estados de la República Mexicana, para distintos grupos de edad y por sexo. Para lograr este objetivo se usarán los datos oficiales publicados diariamente por la Secretaría de Salud.

Las tasas se calcularán relativas a la población de cada cruce de categorías (entidad, edad y sexo) y para ello se usarán datos de CONAPO que pronostican la población en México para el año 2020 por entidad federativa, por grupo de edad y por sexo.

La información se representa por las siguientes variables:

- $Y_{i,j,k}$  = Casos confirmados por COVID19 para el grupo de edad  $i$ , entidad  $j$  y sexo  $k$
- $P_{i,j,k}$  = Población pronosticada para el 2020 para el grupo de edad  $i$ , entidad  $j$  y sexo  $k$  (en cien miles)

Las variables explicativas del modelo son factores que se forman con el cruce de tres variables categóricas edad, entidad y sexo, las cuales se definen a continuación:

- Edad: Se definen 9 grupos de edad,  $i=1,\dots,9$  donde "1"=(0,10), "2"=[10,20), "3"=[20,30), "4"=[30,40), "5"=[40,50), "6"=[50,60), "7"=[60,70), "8"=[70,80) y "9"=[80,∞)
- Entidad: Se consideran las 32 entidades federativas,  $j=1,\dots,32$  en orden alfabético
- Sexo: Se definen dos categorías,  $k=1,2$  en donde "1"=Hombre y "2"=Mujer

El modelo propuesto es el siguiente:

$$Y_{i,j,k} \sim Po(P_{i,j,k} \lambda_{i,j,k})$$
$$\log(\lambda_{i,j,k}) = \alpha + \beta_i + \gamma_j + \delta_k$$

Donde

- $\lambda_{i,j,k}$  = tasa de incidencia de COVID19 en el grupo de edad  $i$ , entidad  $j$  y sexo  $k$  por cada cien mil habitantes

Estas tasas están definidas a través de los siguientes parámetros

- $\alpha$  = ordenada que define el efecto medio global
- $\beta_i$  = efecto de grupo de edad  $i$ ,  $i=1,\dots,9$
- $\gamma_j$  = efecto de la entidad  $j$ ,  $j=1,\dots,32$

- $\delta_k$  = efecto del sexo k, k=1,2

Este modelo de factores está sobre parametrizado, por lo que se imponen las siguientes restricciones de estimabilidad:

$$\sum_{i=1}^9 \beta_i = \sum_{j=1}^{32} \gamma_j = \sum_{k=1}^2 \delta_k = 0$$

La estimación de los parámetros se realizará de manera bayesiana. Para ello establecemos nuestro conocimiento inicial para cada tipo de parámetro mediante el uso de distribuciones jerárquicas, las cuales permiten compartir información entre los parámetros asociados a la misma variable categórica. En particular las distribuciones iniciales son:

$$\alpha \sim N(0, 0.001)$$

$$\beta_i | \tau_\beta \sim N(0, \tau_\beta), i = 1, \dots, 9 \text{ y } \tau_\beta \sim Ga(0.01, 0.01)$$

$$\gamma_j | \tau_\gamma \sim N(0, \tau_\gamma), j = 1, \dots, 32 \text{ y } \tau_\gamma \sim Ga(0.01, 0.01)$$

$$\delta_k | \tau_\delta \sim N(0, \tau_\delta), k = 1, 2 \text{ y } \tau_\delta \sim Ga(0.01, 0.01)$$

Las distribuciones posteriores se caracterizan mediante una muestra obtenida a través de un algoritmo Monte Carlo vía Cadenas de Markov, el cual se implementa en BUGS a través del paquete R.

Finalmente se reportan las tasas estimadas a distintos niveles:

- Tasa de incidencia media nacional:  $e^\alpha$
- Tasa de incidencia por entidad j:  $e^{\alpha+\gamma_j}$
- Tasa de incidencia por entidad j y sexo k:  $e^{\alpha+\gamma_j+\delta_k}$
- Tasa de incidencia por entidad j y grupo de edad i:  $e^{\alpha+\beta_i+\gamma_j}$

Adicionalmente es posible reportar las tasas de incidencia por grupo de edad i, entidad j y sexo k  $\lambda_{i,j,k} = e^{\alpha+\beta_i+\gamma_j+\delta_k}$ ,  $i = 1, \dots, 9, j = 1, \dots, 32, k = 1, 2$ . Cualquiera de estas tasas estimadas se pueden reportar a través de mapas de riesgo.

Este tipo de modelos de mapeo de enfermedades suelen suavizar los picos, tanto altos como bajos, en los datos observados, produciendo mapas de riesgo más informativos, aún para las categorías en donde no se cuenta con suficientes datos observados.

La idea es estimar las tasas de incidencia por COVID19 día por día usando cada vez la información acumulada más actual. De esta manera se producen mapas de riesgo dinámicos diarios. La primera base de datos que se analiza es la del 16 de marzo, la cual contiene 82 casos. A continuación presentamos los estimadores obtenidos de los parámetros del modelo para cuatro cortes, 11 de abril, 10 de mayo, 11 de junio y 10 de julio.

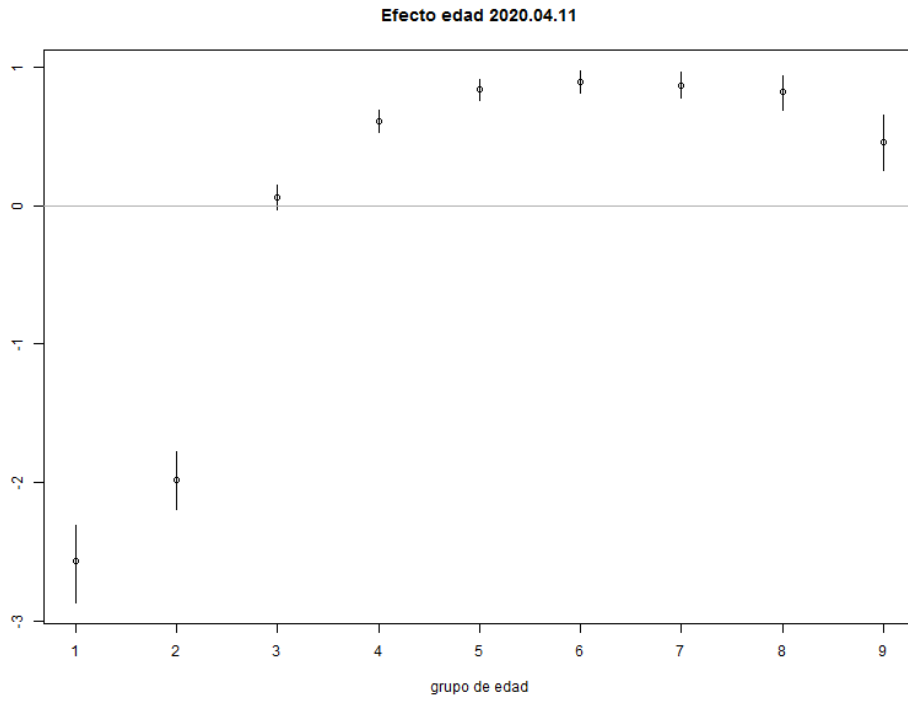


Figura 1. Estimadores de los efectos por grupo de edad  $\beta_i, i = 1, \dots, 9$  con datos al 11/04/2020.

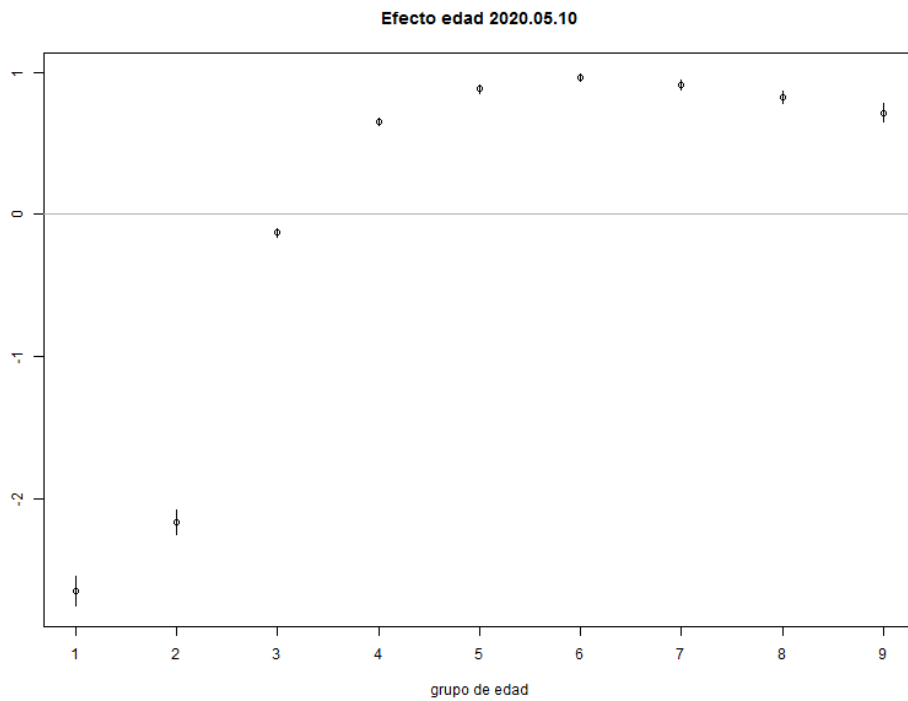


Figura 2. Estimadores de los efectos por grupo de edad  $\beta_i, i = 1, \dots, 9$  con datos al 10/05/2020.

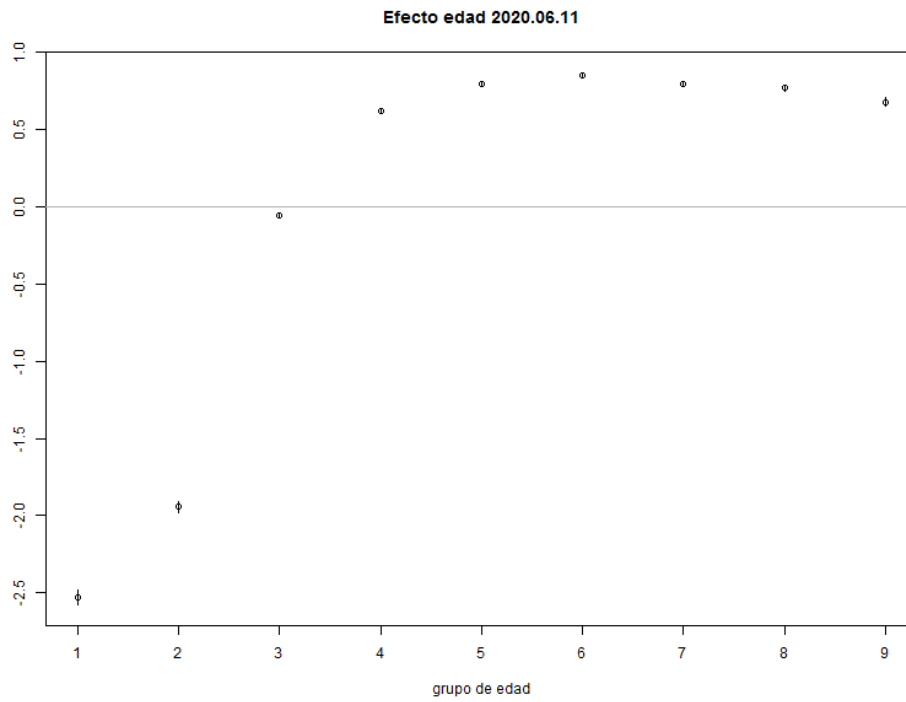


Figura 3. Estimadores de los efectos por grupo de edad  $\beta_i, i = 1, \dots, 9$  con datos al 11/06/2020.

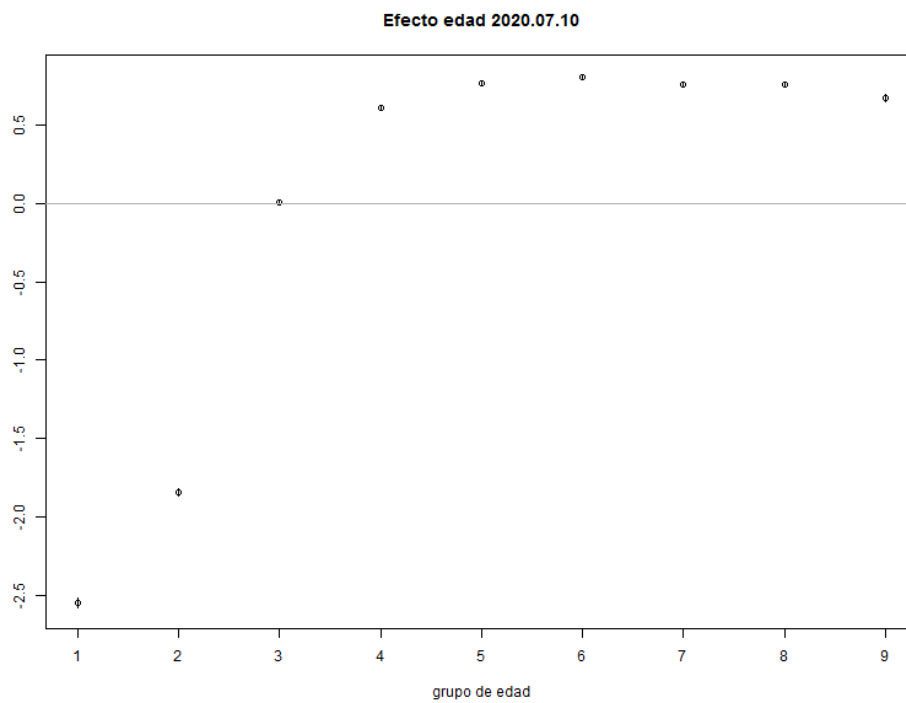


Figura 4. Estimadores de los efectos por grupo de edad  $\beta_i, i = 1, \dots, 9$  con datos al 10/07/2020.

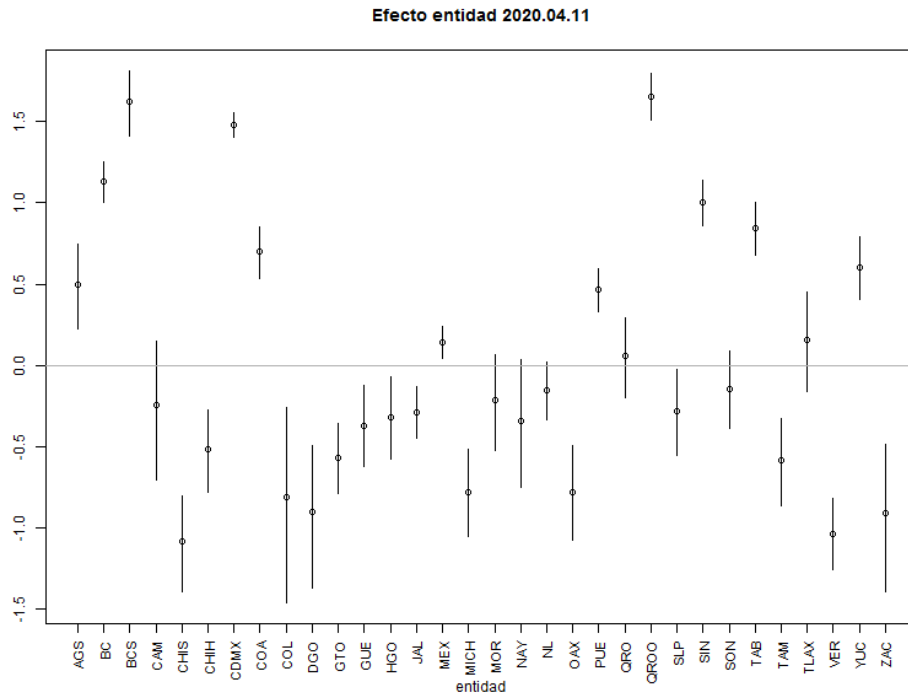


Figura 5. Estimadores de los efectos por entidad  $\gamma_j, j = 1, \dots, 32$  con datos al 11/04/2020.

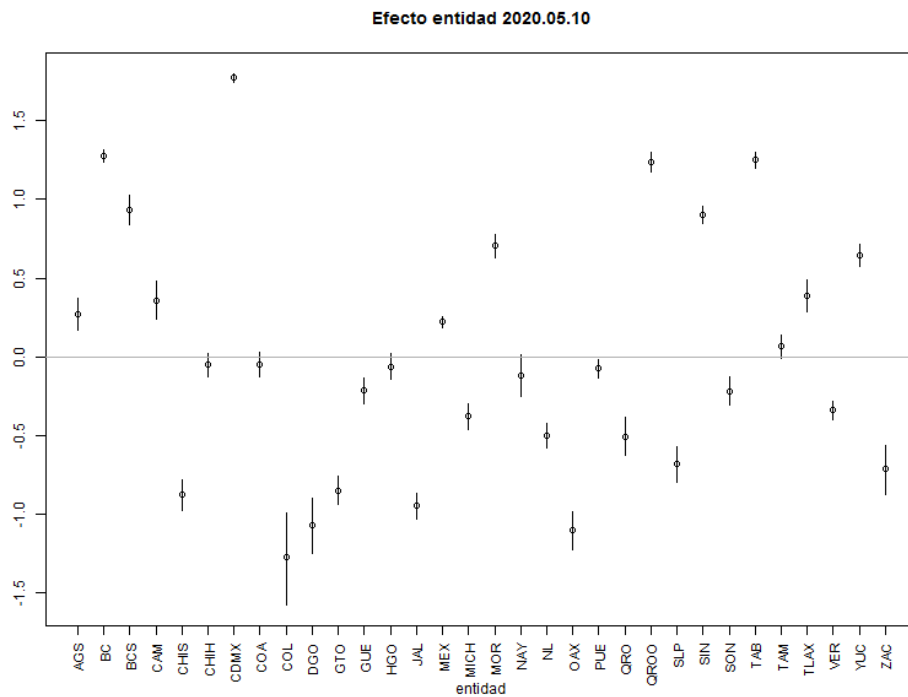


Figura 6. Estimadores de los efectos por entidad  $\gamma_j, j = 1, \dots, 32$  con datos al 10/05/2020.

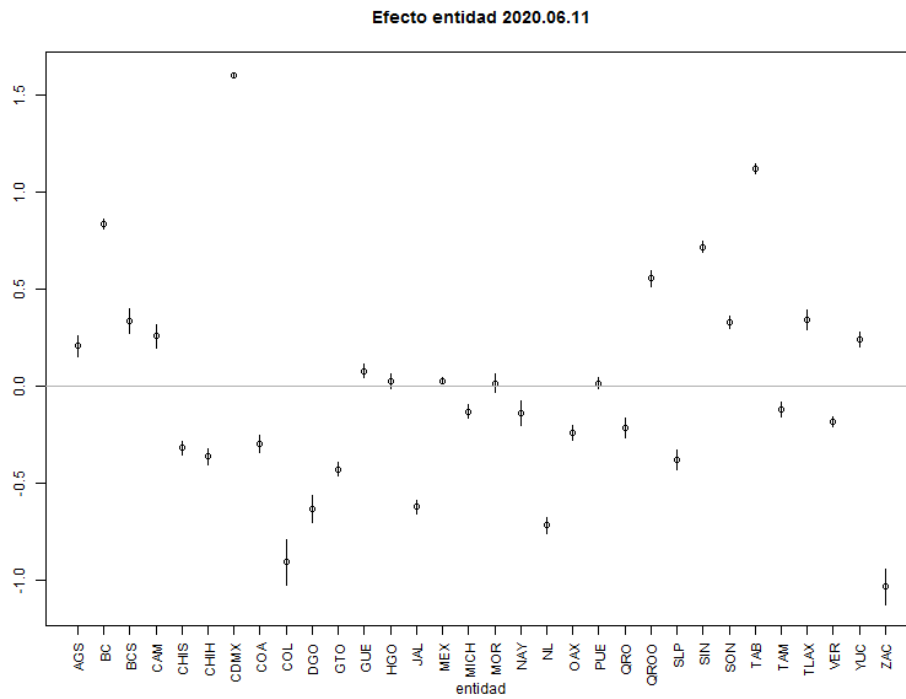


Figura 7. Estimadores de los efectos por entidad  $\gamma_j, j = 1, \dots, 32$  con datos al 11/06/2020.

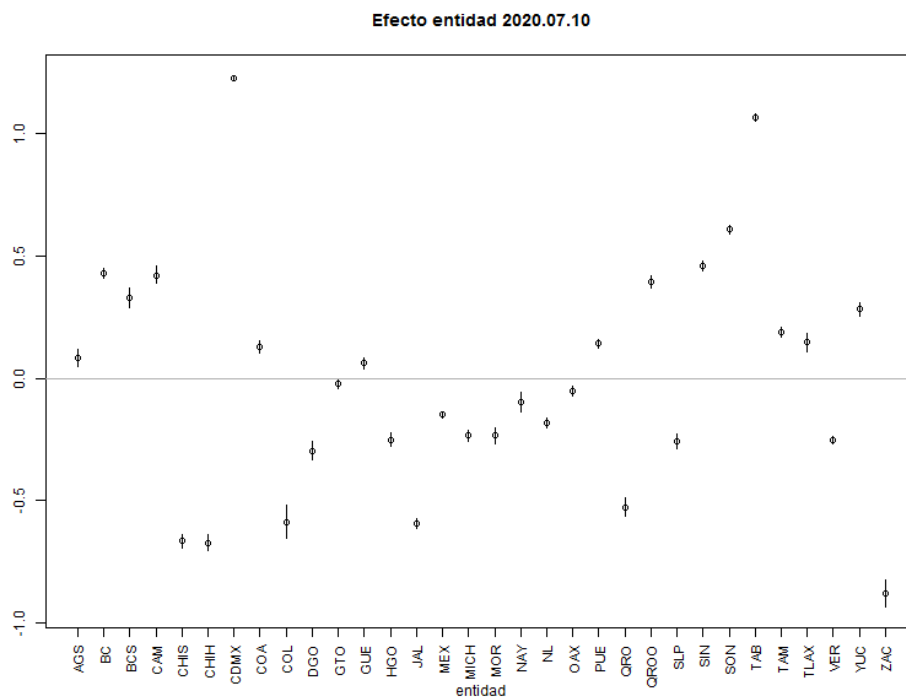


Figura 8. Estimadores de los efectos por entidad  $\gamma_j, j = 1, \dots, 32$  con datos al 10/07/2020.

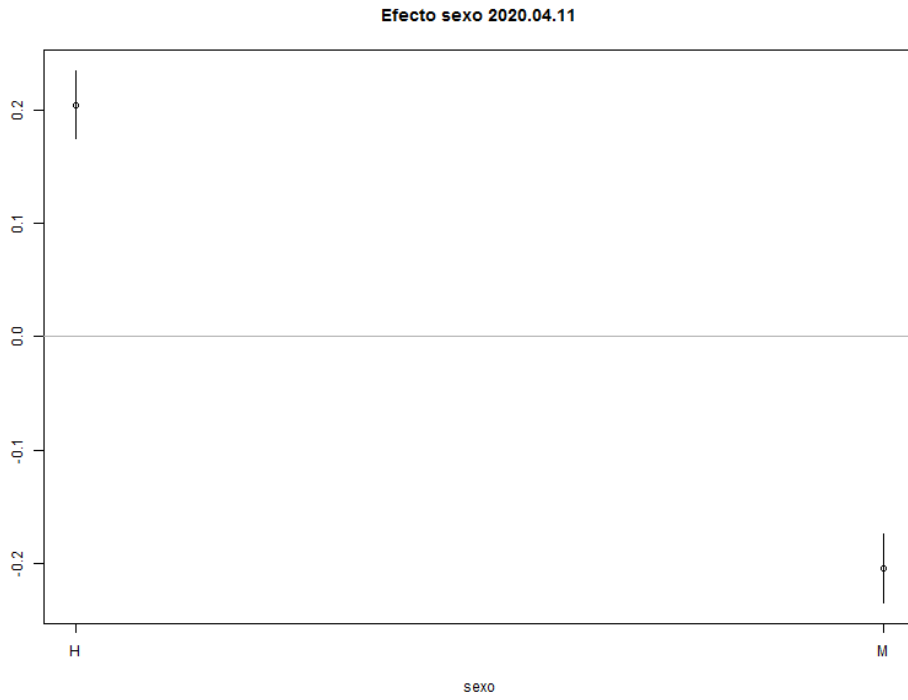


Figura 9. Estimadores de los efectos por sexo  $\delta_k, k = 1,2$  con datos al 11/04/2020.



Figura 10. Estimadores de los efectos por sexo  $\delta_k, k = 1,2$  con datos al 10/05/2020.

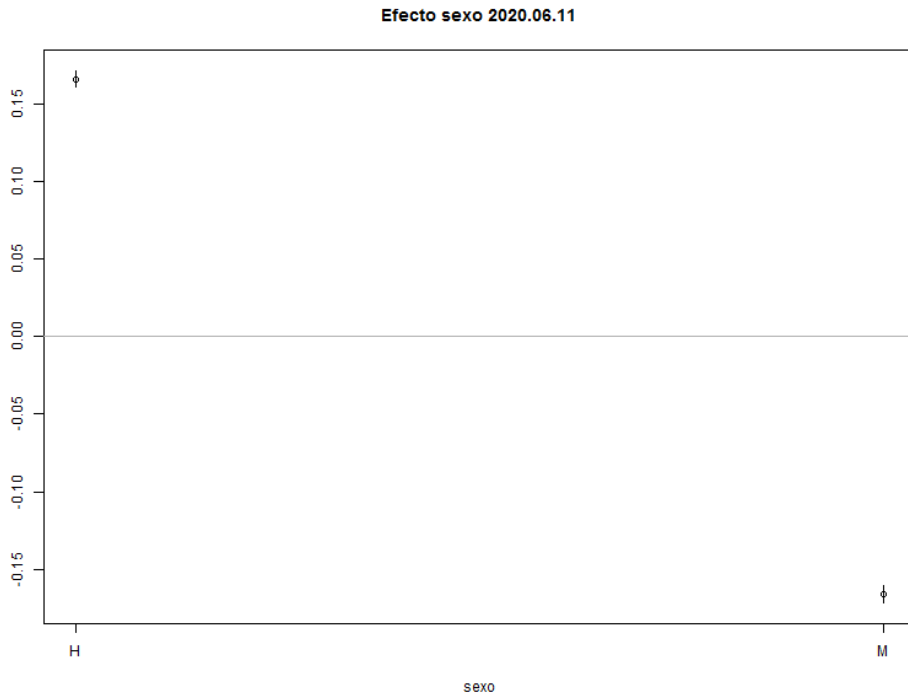


Figura 11. Estimadores de los efectos por sexo  $\delta_k, k = 1,2$  con datos al 11/06/2020.



Figura 12. Estimadores de los efectos por sexo  $\delta_k, k = 1,2$  con datos al 10/07/2020.



La evolución de las tasas de incidencia por entidad  $e^{\alpha+\gamma_j}, j = 1, \dots, 32$  se muestran en mapas dinámicos con corte el 10 de julio:

[http://allman.rhon.itam.mx/~lnieto/index\\_archivos/mapa\\_animate\\_2020-07-10\\_oficial.gif](http://allman.rhon.itam.mx/~lnieto/index_archivos/mapa_animate_2020-07-10_oficial.gif)

Adicionalmente, las tasas de incidencia media nacional, por entidad, por entidad y sexo, por entidad y grupo de edad se reportan junto con sus intervalos de credibilidad con corte el 10 de julio:

[http://allman.rhon.itam.mx/~lnieto/index\\_archivos/tasas2020.07.10.csv](http://allman.rhon.itam.mx/~lnieto/index_archivos/tasas2020.07.10.csv)