LICENCIATURA EN ESTADÍSTICA

UNA MASACRE ESTADÍSTICA

Regresión Lineal Bayesiana

Autores: Franco Santini - Nicolas Gamboa - Andrés Roncaglia

Docentes: Ignacio Evangelista - Tomás Capretto

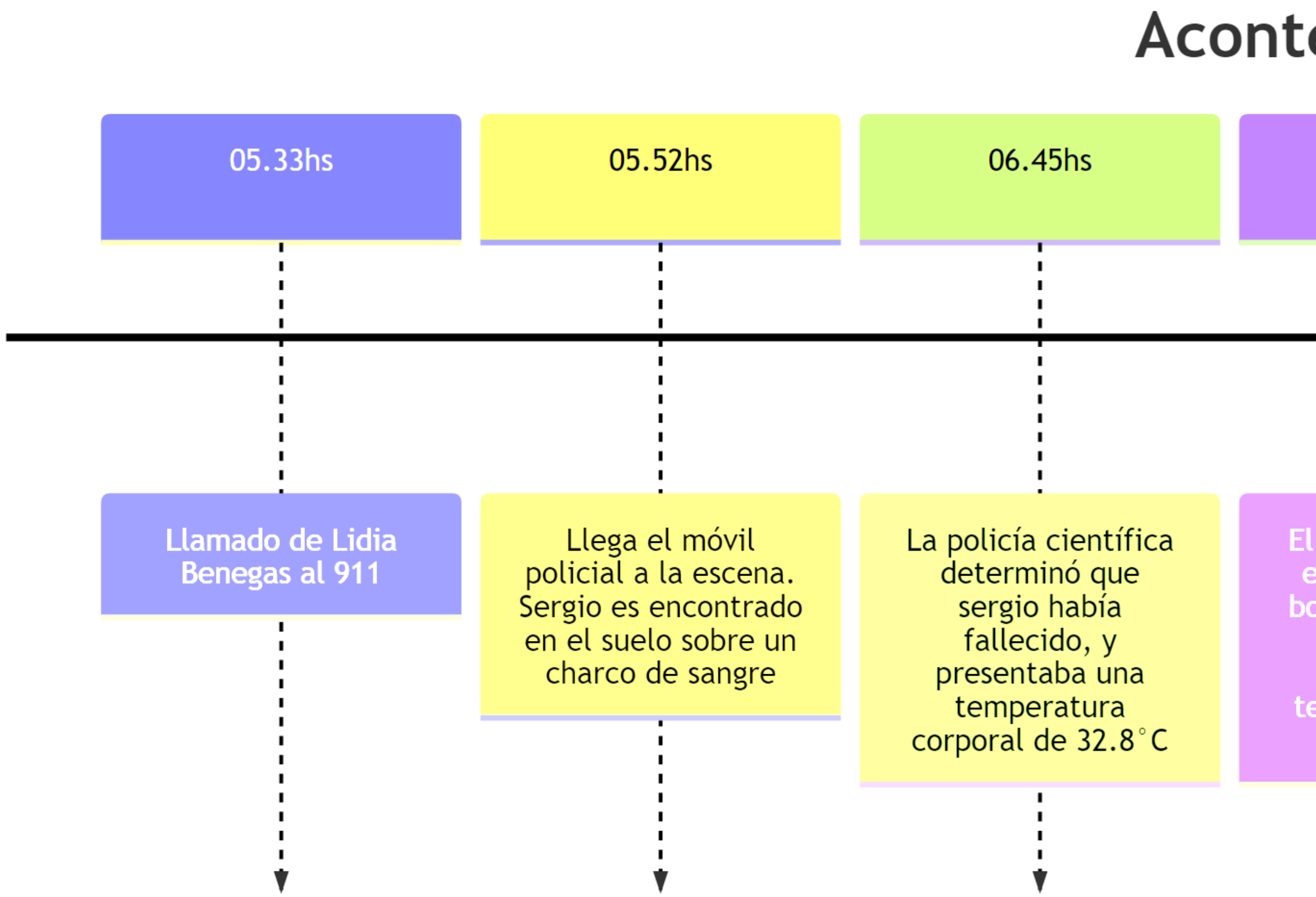
2024

Tabla de contenidos

Introducción

Siguiendo los acontecimientos de la muerte de sergio, una templada madrugada de miércoles en Salsipuedes, la policía forense busca determinar a qué hora murió. El encargado de esto es Guido, un estudiante de estadística fascinado por poder al fin aplicar sus conocimientos estadísticos en un caso de criminalística.

Los datos que la policía foresense propició a Guido son los siguientes:



Metodología

En la medicina forense existen muchas maneras de estimar la hora de la muerte. Entre ellas, una de las más utilizadas es el estudio de la temperatura corporal, la cual es posible estimar.

Se sabe que la temperatura del cuerpo de Sergio satisface la siguiente ley:

$$\frac{dT(t)}{dt} = r[T_{amb} - T(t)]$$

Donde T_amb es la temperatura ambiente en donde se encuentra el cuerpo, en este caso 22°C, r es una constante que determina el ritmo de enfriamiento y $T(t)$ es la función que describe la temperatura del cadaver en función del tiempo después de la muerte.

Una posible función $T(t)$ que satisface esta ecuación diferencial es la siguiente:

$$T(t) = T_{\text{amb}} + (T_i - T_{\text{amb}})e^{-rt} \Rightarrow \; \Rightarrow \frac{dT(t)}{dt} = -r(T_i - T_{\text{amb}}) \cdot e^{-rt} r(T_{\text{amb}} - T_{\text{amb}}) \cdot e^{-rt} - r(T_i - T_{\text{amb}}) \cdot e^{-rt} = \; = r\{T_{\text{amb}} - T_{\text{amb}}$$

Algunos valores posibles para la constante de enfriamiento r son:

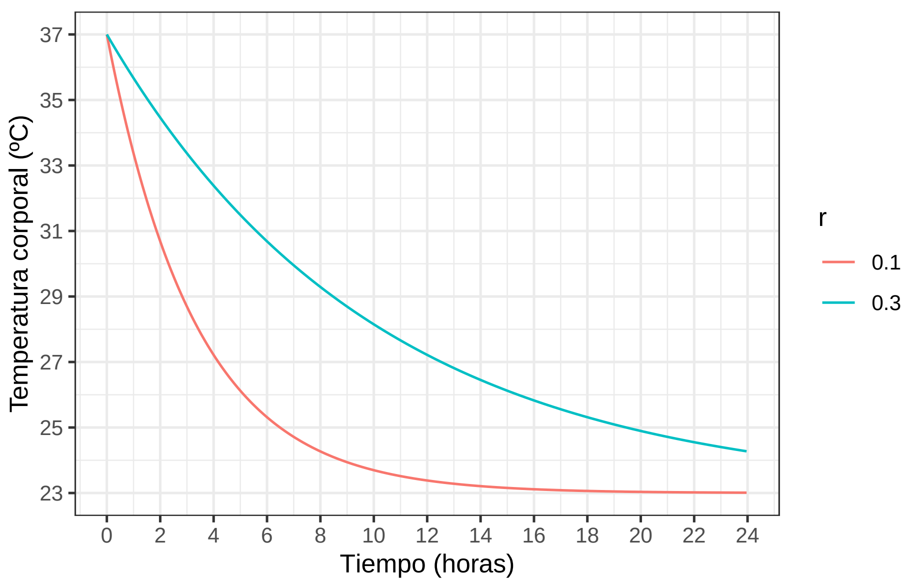


Figura 1: Temperatura del cuerpo según tiempo post-mortem Temperatura ambiente: 23°C

Al ser su primer caso oficial, Guido se sintió un poco desconcertado por no saber a qué hora podría haber ocurrido la tragedia.

Según lo publicado por la La Gaceta Internacional de Ciencias Forenses la tasa de enfriamiento del cuerpo varía en tres estapas, las primeras 3 o 4 horas donde la temperatura baja menos de medio grado por hora, las siguientes 6 a 10 horas donde la temperatura baja alrededor de 1 grado por hora, y a partir de las 10 horas el cuerpo desciende su temperatura entre 3/4 a 1/4 grados por hora, hasta alcanzar la temperatura ambiente. Teniendo en cuenta que r se supone constante, se puede asumir que un buen valor de r es mayor que 0,1 y menor que 0,45.

Con el objetivo de finalmente, después de tanto esfuerzo, estimar la hora de la muerte de Sergio, Guido decide utilizar un enfoque Bayesiano, planteando el siguiente modelo lineal.

$$T(t) - T_{\text{amb}} = T_{\text{diff}}e^{-rt} \Rightarrow ln(T(t) - T_{\text{amb}}) = ln(T_{\text{diff}}e^{-rt}) = \underbrace{ln(T_{\text{diff}})}_{\beta_0} - \underbrace{r}_{\beta_1} t = \beta_0 + \beta_1 \cdot t$$

Donde β_0 representa el logaritmo natural de la diferencia entre la temperatura inicial del cuerpo y la temperatura ambiente y β_1 es el cambio esperado en la diferencia de temperaturas del logaritmo por cada hora.

Una transformación posible para este modelo es la siguiente:

Para poder interpretar t como el tiempo después de haber encontrado el cuerpo, se realiza una traslación utilizando un nuevo parámetro δ.

$$T(t+\delta) - T_{\text{amb}} = T_{\text{diff}}e^{-r(t+\delta)} \Rightarrow ln(T(t+\delta) - T_{\text{amb}}) = ln(T_{\text{diff}}e^{-r(t+\delta)}) = \underbrace{ln(T_{\text{diff}})}_{\beta_0} - \underbrace{r}_{\beta_1} (t+\delta) = \beta_0 - \beta_1 \cdot (t+\delta)$$

El modelo que se usa para estimar la hora de muerte de Sergio resulta entonces:

$$ln(T(t+\delta) - T_{\text{amb}}) = \beta_0 - \beta_1 \cdot (t+\delta)$$

Donde:

- δ es el tiempo (en horas) que transcurrió entre la que persona murió y se encontró el cuerpo.
- Los demás parámetros conservan su significado.

Guido recuerda haber mirado el pronóstico del tiempo el día de la muerte de Sergio y recuerda que anunciaban una temperatura ambiente de 22ºC. Por otro lado, dado que la temperatura corporal de una persona sana ronda los 37°C, consideró esta la temperatura inicial del cuerpo de Sergio.

$$ln(T(0) - T_{amb}) = \beta_0 - \beta_1 \cdot (0)\beta_0 = ln(37 - 22) = ln(15) = 2.7081$$

Al ser un enfoque bayesiano, Guido debe pensar una distribución a priori para los parámetros δ y β_1.

En principio, dado que el cuerpo fue encontrado con una temperatura corporal de 32.8°C, y sabiendo que Sergio murió desangrado, lo cual acelera considerablemente el descenso de la temperatura, cree que debe haber muerto hace aproximadamente 2 horas, aunque no está para nada seguro, por lo que utilizará una distribución uniforme, que irá entre 1 hora y media y 2 horas y media.

Con respecto al parámetro β_1, y considerando la hora a la que estima que murió Sergio y la temperatura a la que fue encontrado, observa lo siguiente.

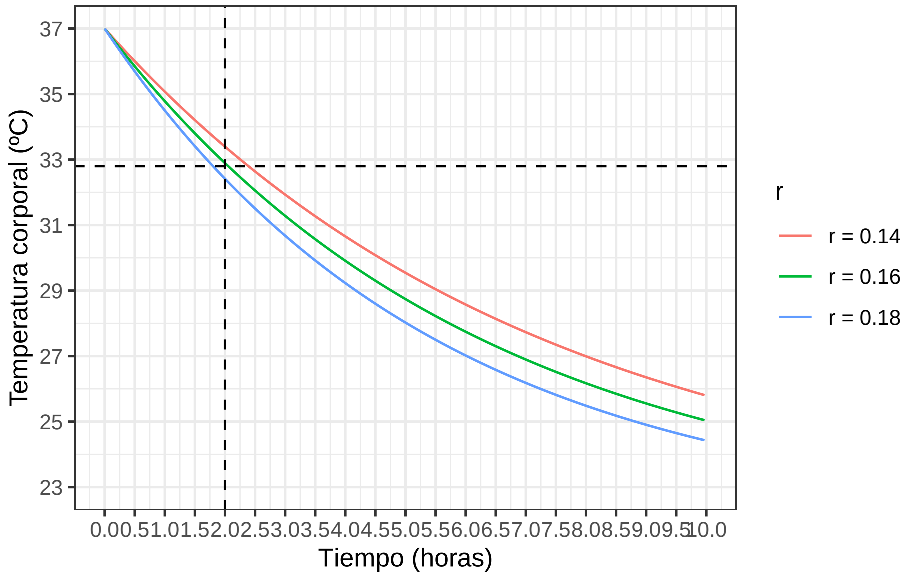


Figura 2: Temperatura del cuerpo según tiempo post-mortem Temperatura ambiente: 22°C

En base a esto, tomará valores similares a los graficados para la distribución de β_1, considerando una distribución normal.

De manera un poco más formal, el modelo que planteado por Guido es el siguiente:

$$ln(Y)/\mu_i,\ \sigma \sim N(\mu_i,\ \sigma)\mu_i = \beta_0+\beta_1\cdot(t+\delta)\beta_0 = 2.7081\beta_1 \sim N^+(0.16,\ 0.075)\delta \sim U(1.5,\ 2.5)\sigma \sim N^+(0,\ 0.2)$$

Los parámetros, a priori, son independientes y siguen las distribuciones mencionadas.

El interés se centra principalmente en el parámetro δ

Resultados

El objetivo de Guido es estimar la hora de la muerte y para hacerlo usará toda la información que tenga a su disposición. Pero antes de esto quiere confirmar que sus creencias iniciales tienen sentido, por lo que decide realizar pruebas predictivas a priori.

Pruebas predictivas a priori

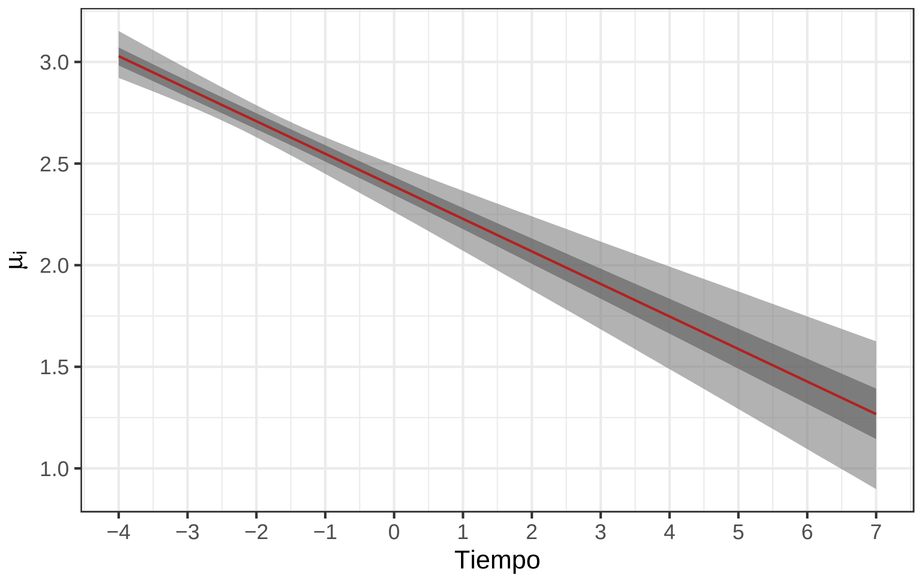


Figura 3: Distribución a Priori del predictor lineal.

En un principio, el prior de Guido parece tener sentido.

Busca un intervalo de tiempo en el que pueda decir con 90% de certeza que Sergio murió dentro de ese rango horario.

A priori Guido concluye con una credibilidad del 90% que sergio murió entre las 4.19hs y las 5.12hs.

Pruebas predictivas a Posteriori

Ahora sí, finalmente Guido utilizará los datos provistos para estimar la hora de la muerte, pero con el objetivo de desafiarse a sí mismo decide hacerlo de manera secuencial, para ver como los datos van afectando sus estimaciones.

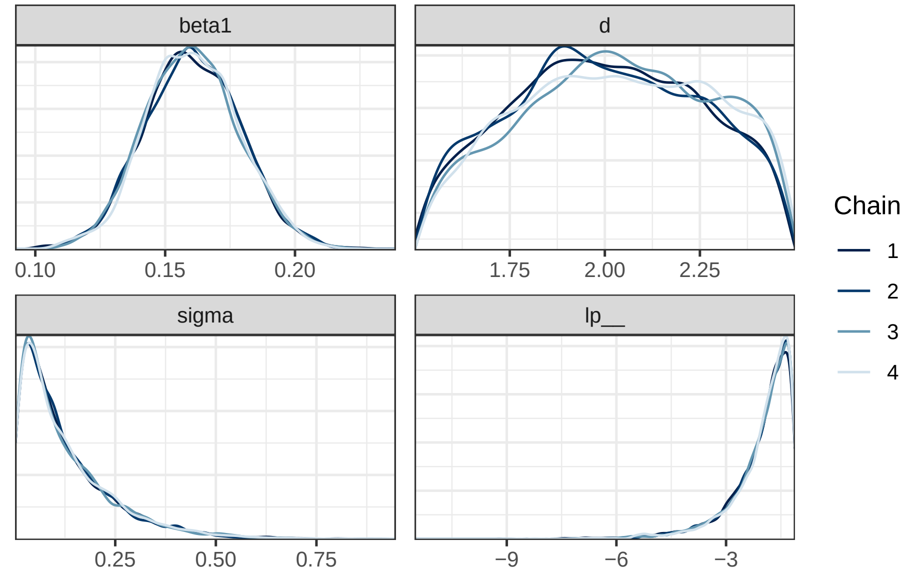
Después de que la computadora de bajos recursos del departamento de la Policía Científica estuviera encendida durante días, Guido consiguió los siguientes resultados.

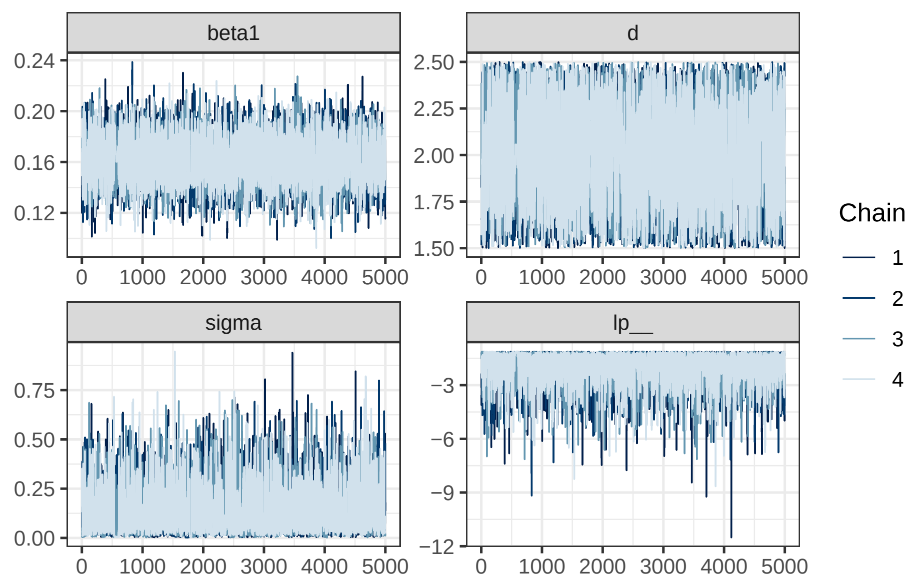
Los datos disponibles son los siguientes:

$Tiempo$	$Temp_{diff}$	$ln(Temp_{diff})$
0.00	10.8	2.380
1.50	8.5	2.140
6.75	1.7	0.531

A Guido le remarcaban mucho en clases de Estadística Bayesiana la importancia de verificar la convergencia de las cadenas de Markov, en este caso obtenidas por el método NUTS (*No-U-Turn*

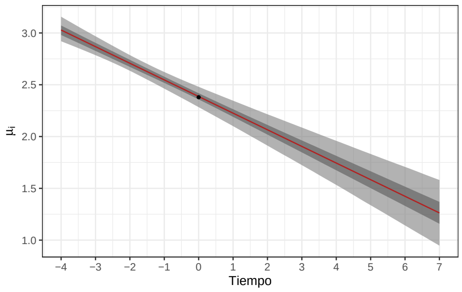
Sampler), para los distintos párametros del modelo. Es por ello que para cada una de las estimaciones de los parámetros, presenta gráficos de densidad y de traza.



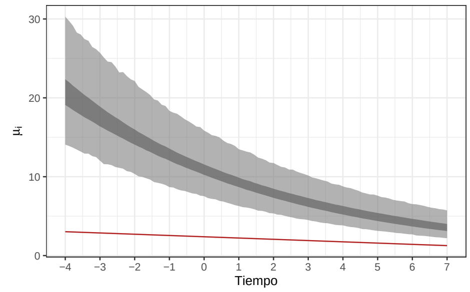


Con la información provista hasta las 7 de la mañana las cadenas de markov para cada uno de los parametros parecen converger a la misma distribucion a posteriori. Esto, además de haber obtenido un $\hat{R} = 1$ para todos los párametros le confirman a Guido que las muestras obtenidas son representativas de los parámetros.

La curiosidad de Guido lo llevaba a querer visualizar como los datos afectan al prior, por lo que graficó bandas de credibilidad para el predictor lineal.

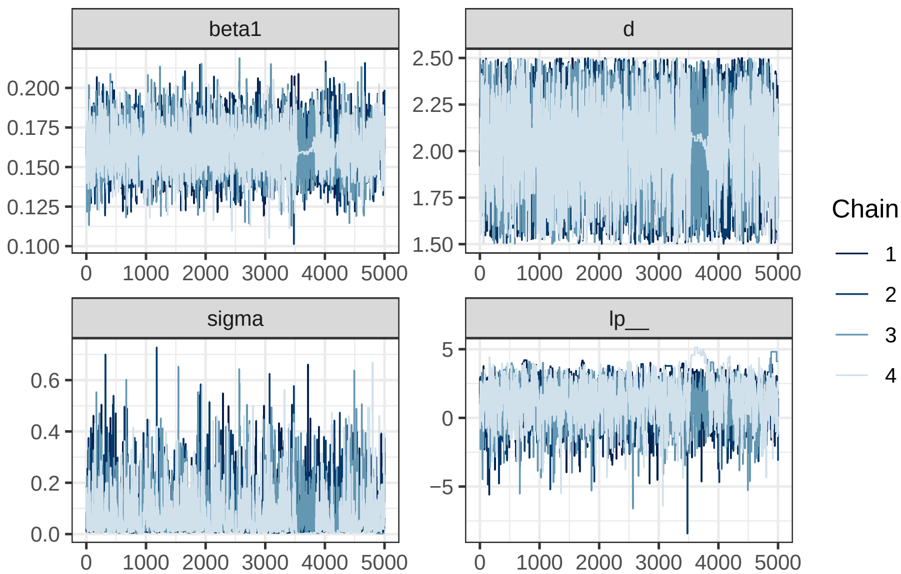


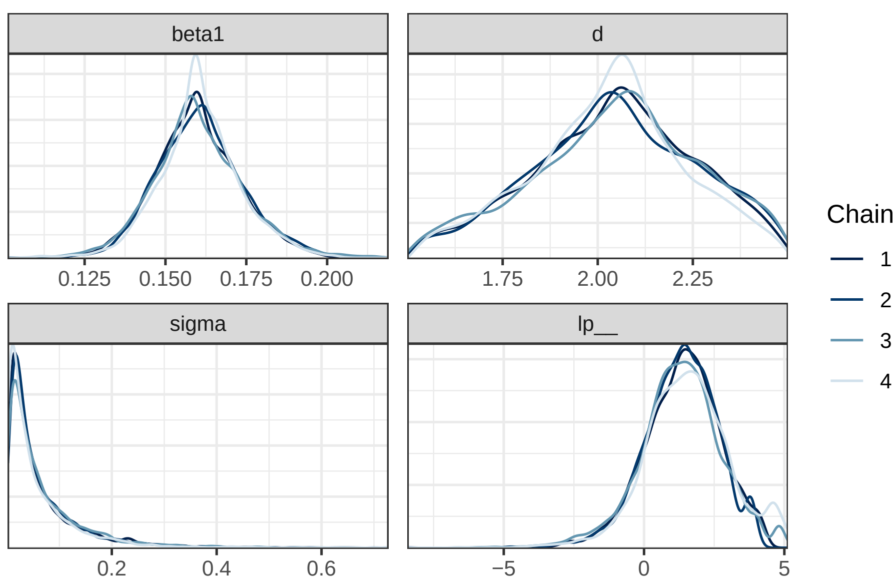
(a) Del predictor lineal



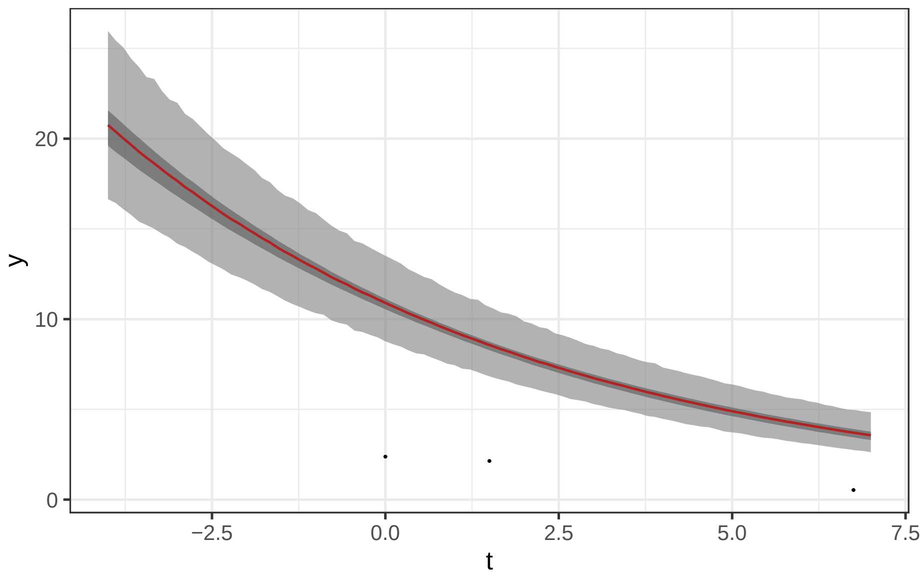
(a) De $T(t) - T_{amb}$

Distribuciónes a Priori con la información hasta las 7hs.

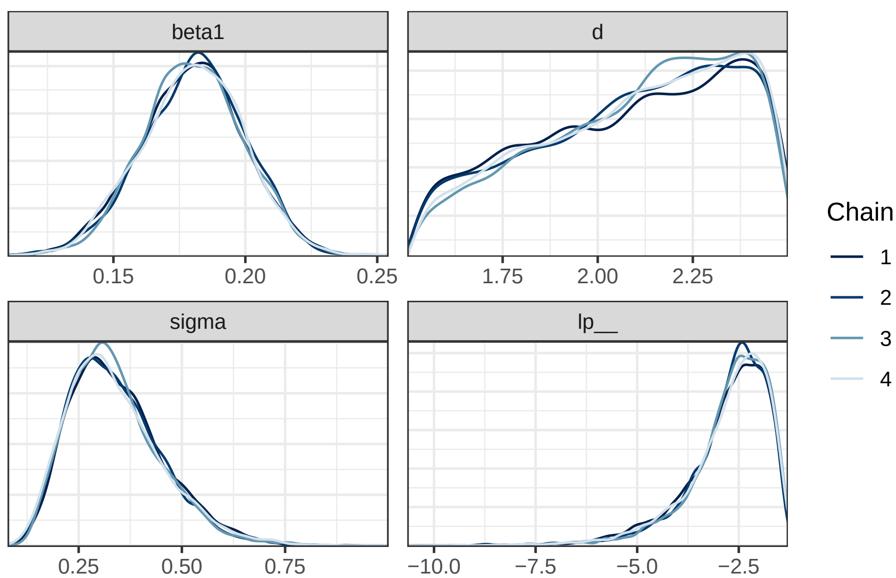




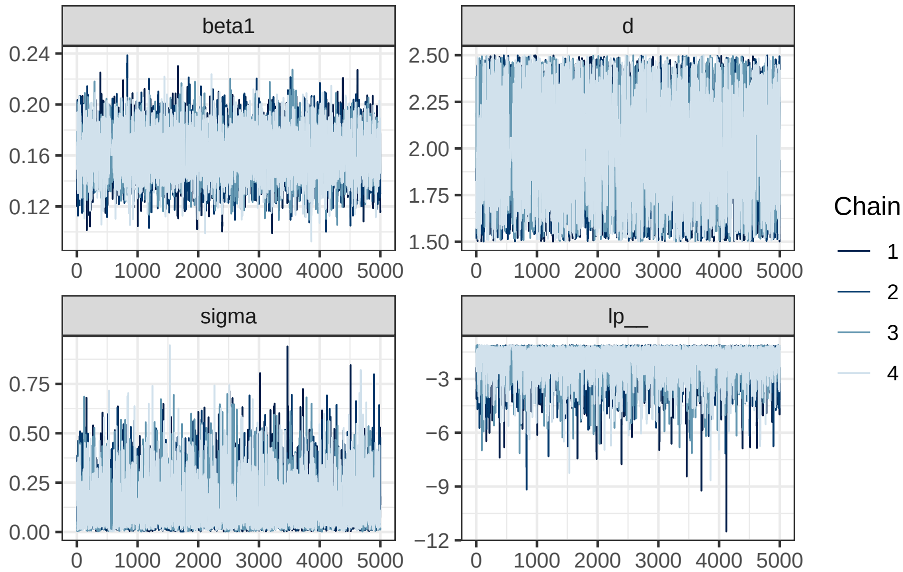
```
[1] 0.8212
```

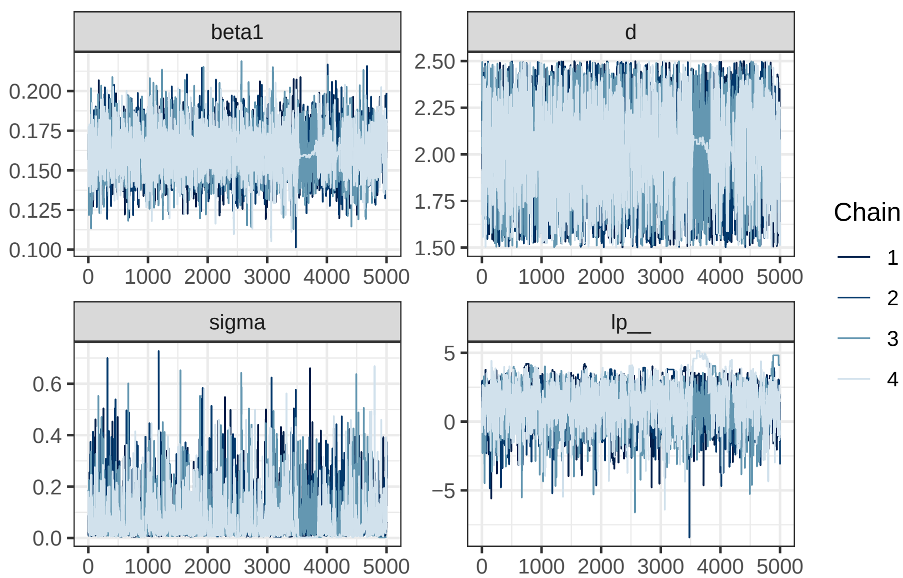


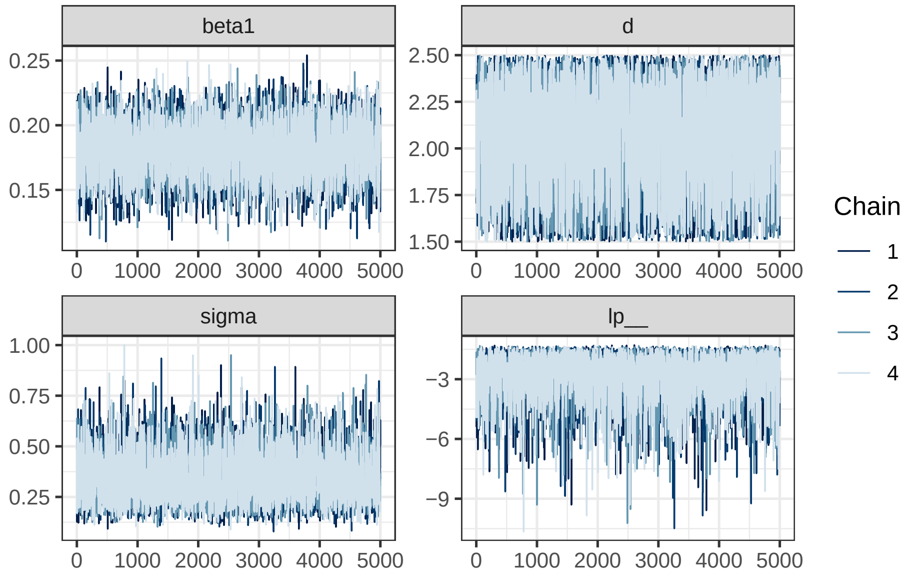
20
10
0
y
−2.5
0.0
2.5
5.0
7.5
t



beta1
d
sigma
lp__
0.15
0.20
0.25
1.75
2.00
2.25
0.25
0.50
0.75
−10.0
−7.5
−5.0
−2.5
Chain
1
2
3
4


beta1
d
sigma
lp__
0.24
0.20
0.16
0.12
2.50
2.25
2.00
1.75
1.50
0.75
0.50
0.25
0.00
−3
−6
−9
−12
0 1000 2000 3000 4000 5000
Chain
1
2
3
4


beta1
d
sigma
lp__
0.200
0.175
0.150
0.125
0.100
2.50
2.25
2.00
1.75
1.50
0.6
0.4
0.2
0.0
5
0
−5
0 1000 2000 3000 4000 5000
Chain
1
2
3
4

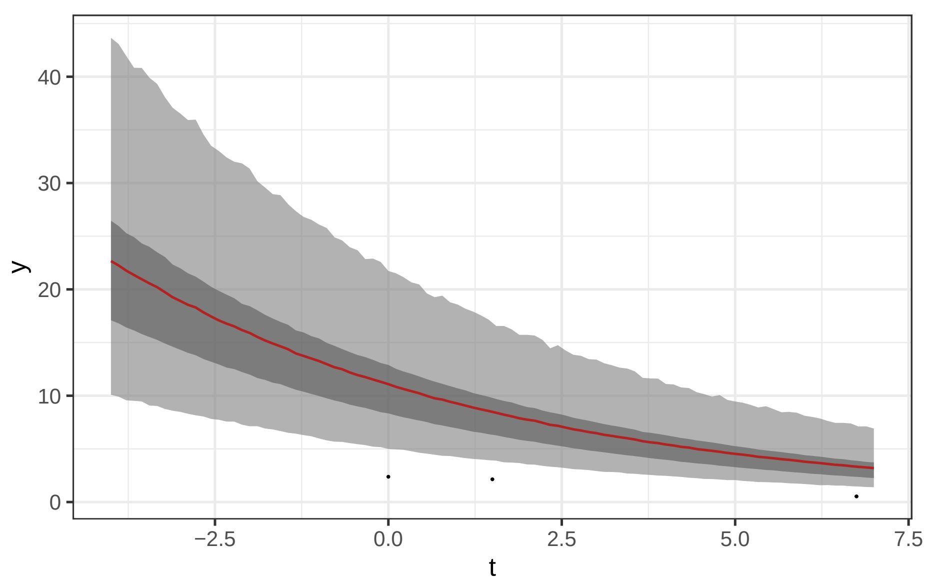


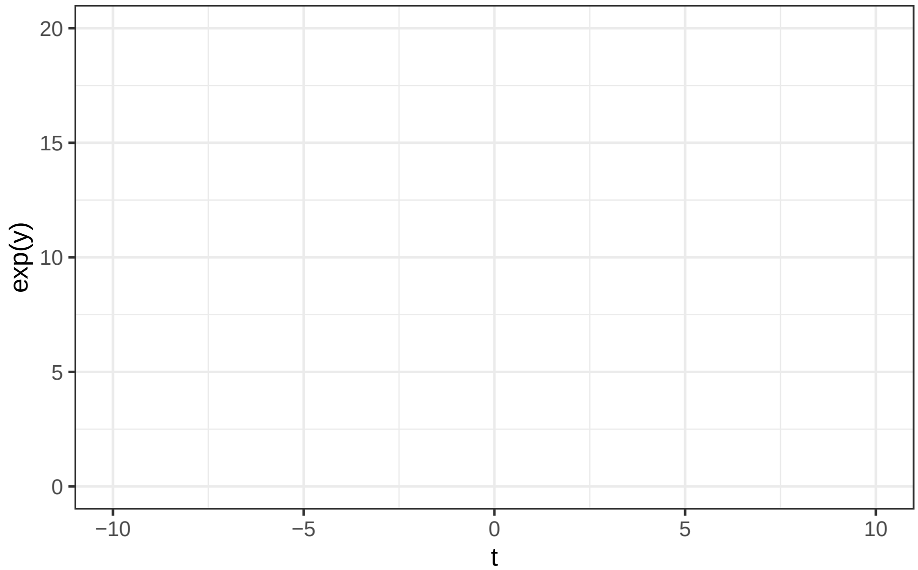

```
      10%      90%
1.601943 2.398954

      10%      90%
1.638244 2.380055

      10%      90%
1.717267 2.343933

      10%      90%
1.655706 2.422513
```





Punto 9

Modificaría algún aspecto del modelo si supiera que Sergio no gozaba de buena salud al momento de fallecer?.

Si se tuviese información que indique que Sergio no gozaba de buena salud al momento de fallecer esto podría modificar nuestra creencia a Priori ya que la velocidad de enfriamiento se retarda en casos de enfermedades febriles, intoxicaciones por nicotina, rodenticidas, entre otros y obesidad. También la perdida de calor del cuerpo puede acelerarse en personas con bajo peso, enfermedades crónicas e intoxicadas por fósforo, arsénico y alcohol.

Una posible solución podría ser una variable que indique la enfermedad.

Se podría entonces agregar una nueva variable que indique como se modifica la tasa de enfriamiento del cuerpo según su estado de salud previo a la muerte. Resultando el modelo:

$$\underbrace{ln(T_{\text{diff}})}_{\beta_0} + \underbrace{(-r)}_{\beta_1} t + (\alpha)t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t$$

No modificaría en ningún aspecto el modelo el saber que Sergio no gozaba de buena salud, dado que nuestro prior abarca hasta el caso más extremo de la temperatura corporal de una persona, considerando que la temperatura del ambiente se encuentra a 22°.