

LICENCIATURA EN ESTADÍSTICA

“Una masacre estadística”

Trabajo Práctico 3

Autores: Tomás Anderson - Alejo Vascetti - Manuel Moresi

Docentes: Ignacio Evangelista - Tomás Capretto

05/06/2024

Tabla de contenidos

Prólogo: El comienzo de Guido, el fin de Sergio	1
Introducción	1
Capítulo 1: Elemental mi querido Guido	1
Medicina forense	1
Capítulo 2: Luego de la muerte, ¿Qué ocurre?	3
Distribuciones a <i>priori</i>	4
β_0 : Logaritmo natural de la diferencia entre la temperatura inicial del cuerpo y la ambiente	5
β_1 : Velocidad en la que la temperatura del cuerpo cae	5
h_m : Horario de la muerte	6
σ : Error de medición	7
Descenso de la temperatura a <i>priori</i>	8
Capítulo 3: Doctor formulemelo	9
<i>Posterior</i> de los parámetros con una observación	10
Capítulo 4: Crónica de una catástrofe bayesiana	14
<i>Posterior</i> de los parámetros con dos observaciones	14
<i>Posterior</i> de los parámetros con tres observaciones	18
Capítulo 5: “D.e.a.t.h.” Divagaciones entre aficionados tanteando hipótesis	21
Falencias del ajuste	22
Epílogo: ¡Y Guido pudo salir!	22
Conclusiones	22
Anexo	23

Prólogo: El comienzo de Guido, el fin de Sergio

Era una pacífica madrugada de miércoles en Córdoba, cuando suena un teléfono en la cercanía. Letárgico, luego de dejar sonando unos segundos el celular, Guido abre los ojos desganadamente y comprueba quién es la persona que está irrumpiendo en su sueño. “-¿El capitán?”, piensa Guido dudosamente. Al atender el móvil se escucha, “-Llegó la hora de tu primer trabajo Guido. Preparate que en 10 minutos te pasan a buscar para ir a Salsipuedes.” Como si todavía no hubiese entendido qué está pasando, lentamente Guido salió de su lecho para vestirse con su preciado nuevo uniforme de la policía científica de Córdoba. Recién cuando el vehículo llegó, pudo Guido entender lo que había pasado. “-¡Por fin! Mi primer trabajo.”, pensó Guido alegremente. Pero, ¿estaba preparado para afrontar lo que se iba a encontrar?

Introducción

Se suele decir que la muerte es certera pero... ¿es precisa? Acompañá a Guido, un estudiante de estadística, que como parte de su nuevo trabajo tiene que estimar a qué hora se produjo un asesinato en la ciudad de Salsipuedes.

Capítulo 1: Elemental mi querido Guido

Al llegar, Guido presenció el caos de la escena instantáneamente. Muebles volcados y destrozados. Utensilios de cocina desparramados y manchados. La brisa matutina se abría paso por los vidrios estallados y la débil luz de los móviles policiales dejaban ver que el cuerpo de Sergio Contreras yacía en un charco de sangre hacia un extremo del salón. Sorprendentemente, recordando lo que estudió, salió de su estupor y fue a averiguar los hechos del caso.

Medicina forense

Uno de los métodos que se utiliza en la medicina forense para poder estimar la hora del fallecimiento, es el estudio de la temperatura corporal. Cuando alguien fallece, la temperatura del cuerpo irá disminuyendo hasta que este llegue a un equilibrio térmico con la temperatura ambiente. Dicho proceso se lo conoce como *enfriamiento postmortem* o *algor mortis*. Pero el ritmo al que el cuerpo pierde calor no es constante. Dicho cambio se puede expresar como una ecuación diferencial que satisface la siguiente ley:

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

donde T_{amb} es la temperatura ambiente, r una constante positiva que indica la rapidez del *algor mortis* y $T(t)$ es la función que describe la temperatura del cuerpo en función del tiempo, donde $t = 0$ es el momento de la muerte de la persona.

Una función que satisface la ecuación diferencial es:

$$T(t) = T_{amb} + (T_i - T_{amb})e^{-rt}$$

Donde T_i es la temperatura inicial del cuerpo. Dicha relación se demuestra de la siguiente manera:

$$\begin{aligned}\frac{dT(t)}{dt} &= -r(T_i - T_{amb})e^{-rt} \\ \frac{dT(t)}{dt} &= -r(T_i - T_{amb})e^{-rt} + rT_{amb} - rT_{amb} \\ \frac{dT(t)}{dt} &= -r(T_{amb} + (T_i - T_{amb})e^{-rt}) + rT_{amb} \\ \frac{dT(t)}{dt} &= -rT(t) + rT_{amb} \\ \frac{dT(t)}{dt} &= r(T_{amb} - T(t))\end{aligned}$$

Si $T_{amb} = 23$ y $T_{ini} = 37$ la forma de la función es la siguiente:

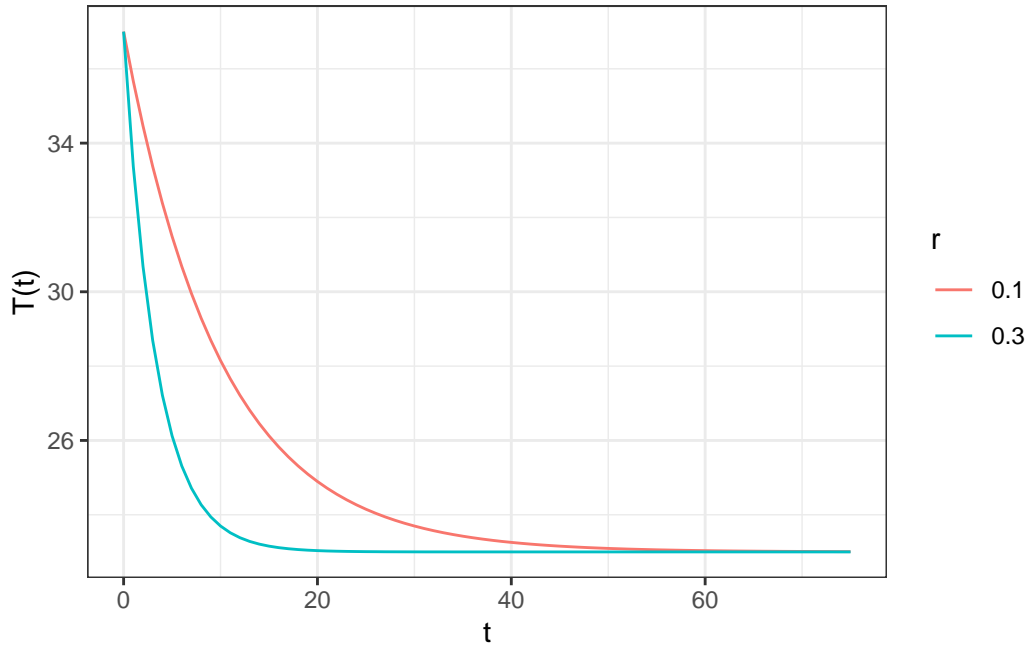


Figura 1: Temperatura de un cuerpo según el tiempo desde la muerte

Para simplificar el problema, no se utiliza la función $T(t)$ sino $\ln(T(t) - T_{amb})$, ya que esta tiene forma lineal.

$$T(t) - T_{amb} = T_{diff}e^{-rt} \quad \text{Donde : } T_{diff} = T_i - T_{amb}$$

$$Y(t) = \ln(T(t) - T_{amb}) = \ln(T_{diff}e^{-rt}) = \ln(T_{diff}) + \ln(e^{-rt}) = \ln(T_{diff}) - rt$$

Dicha función está representada por la ecuación de una recta $y = \beta_0 + \beta_1 * x$, donde β_0 es el logaritmo natural de la diferencia de temperatura del cuerpo inicial y de la ambiente. β_1 representa la velocidad de cambio de $Y(t)$.

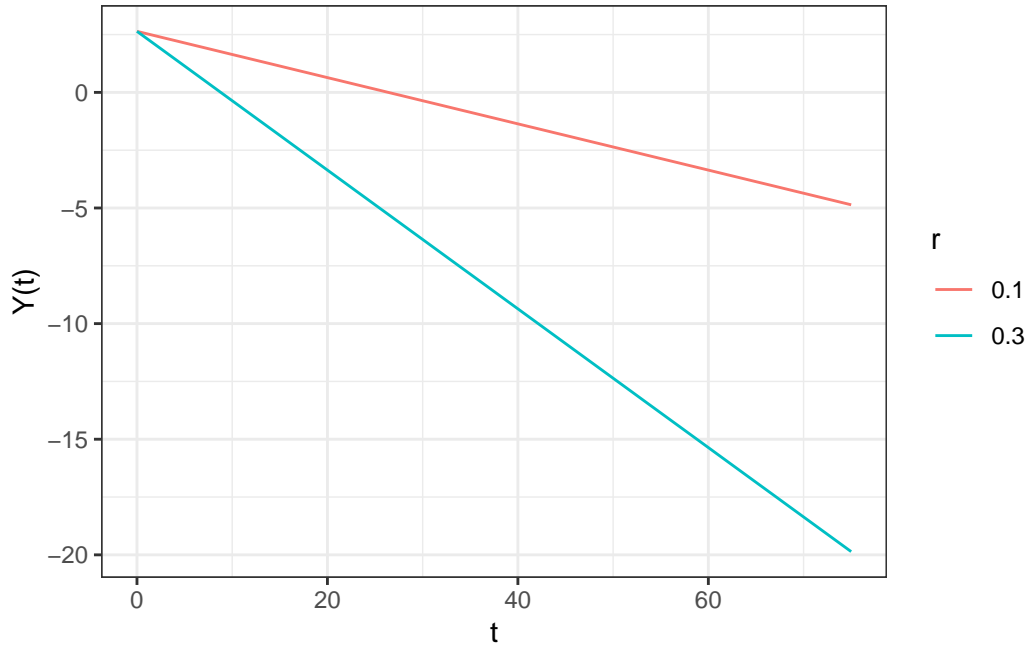


Figura 2: Transformación de la temperatura de un cuerpo según el tiempo desde la muerte

Capítulo 2: Luego de la muerte, ¿Qué ocurre?

Luego de que Martina, una policía que estaba en la escena del crimen, le informe a Guido que a las 5:33 am recibieron una llamada de una vecina sobre ruidos extraños, el joven estadístico le pregunta “-¿Y a qué hora llegaron?”. “-Llegamos a las 5:52 y en 8 minutos concluimos la inspección de la casa. En este mismo estado la encontramos.”. Sin poder continuar la conversación, el médico del equipo llama a Guido y le cuenta que la temperatura del cuerpo de Sergio es de 32.8 °C. “-¿Necesitás alguna información más

Guido?”. Sin mirarle a los ojos, Guido saca su celular, observa la hora, busca la temperatura ambiente actual y dice expectante “-No te preocupés doc, ya tengo todo lo que necesito”. Pero Guido, ¿tiene todo lo que necesita?

Distribuciones a *priori*

Fundamentalmente, la temperatura del cuerpo de Sergio a través del tiempo luego de fallecer depende de 4 aspectos:

- La temperatura del cuerpo al momento de morir ($^{\circ}\text{C}$).
- La temperatura ambiente ($^{\circ}\text{C}$).
- La velocidad en la que la temperatura del cuerpo cae.
- A qué hora falleció.

Además, hay que tener en cuenta el error de medición que pueda tener el equipo que se use para medir la temperatura.

Pero se quiere que el modelo pueda expresarse en función de una recta, por lo que se trabaja con el logaritmo natural de la diferencia de temperatura del cuerpo y la temperatura del ambiente.

Esto se explicita en el siguiente modelo:

$$Y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 t_i$$

Donde:

$$Y_i = \ln(T(t_i) - T_{amb})$$

$$\beta_0 = \ln(T_{ini} - T_{amb}) = \ln(T_{diff})$$

$$\beta_1 = -r$$

Y t_i es el i -ésimo tiempo en horas que pasó desde que Sergio murió.

Se conoce la temperatura ambiente, siendo $T_{amb} = 22^{\circ}\text{C}$. Los valores exactos de los demás parámetros se desconocen, pero se tiene cierta información sobre ellos.

β_0 : Logaritmo natural de la diferencia entre la temperatura inicial del cuerpo y la ambiente

Se sabe que la temperatura del cuerpo de una persona sana varía entre 36 y 37.5 °C, por lo que $\beta_0 = \ln(T_{diff})$ tiene que estar entre $\ln(36 - 22) = 2.639$ y $\ln(37.5 - 22) = 2.741$. Una distribución que representa lo requerido es:

$$\beta_0 \sim Normal(2.69, 0.017)$$

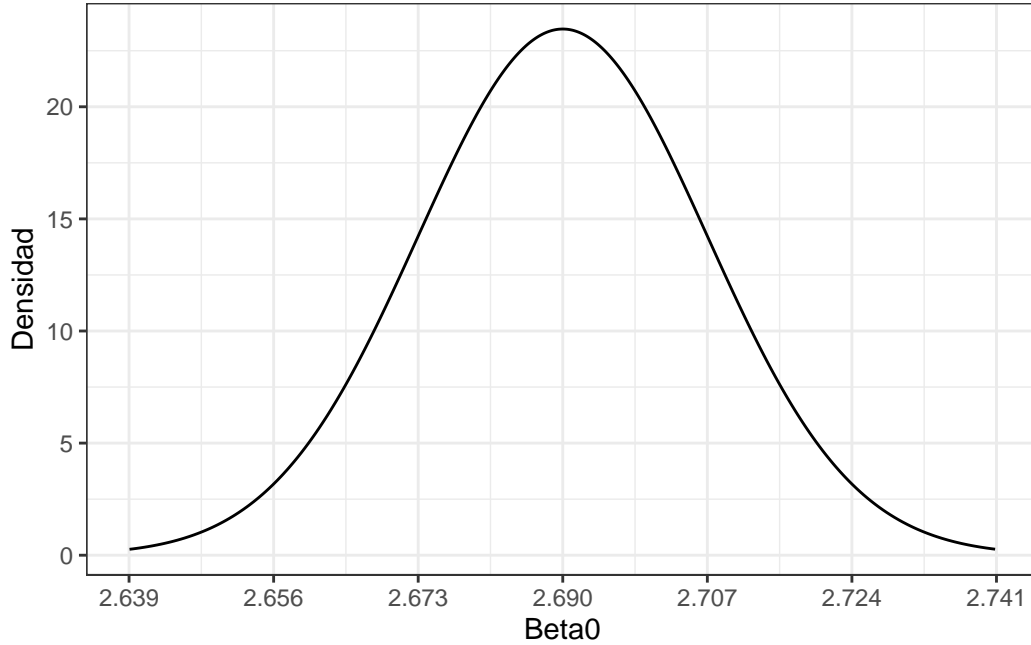


Figura 3: Distribución a priori de beta 0

β_1 : Velocidad en la que la temperatura del cuerpo cae

La velocidad a la que decrece la temperatura de los cuerpos humanos cuando mueren es un tema ya estudiado en la medicina forense.

“Greggio y Valtorta(1998) han demostrado que el descenso de la temperatura corporal post-mortem se realiza mediante un *plateau* inicial de una 3-4 horas de duración, con una caída promedio de 0.5 °C por hora, un decrecimiento lineal que sobrepasa las 12 hs de la muerte con una pérdida promedio de 1 °C cada hora y luego un *plateau* final, con descenso lento, de 0,5° o 0,25° por hora hasta igualar a la temperatura ambiente.” Fraraccio, J(2021). *Medicina forense contemporanea*. Dosyuna ediciones argentinas.

Basándose en este extracto, se tarda aproximadamente 22 horas en que un cuerpo llegue a una temperatura ambiente de 22 °C, por lo que se calcula el valor de $T(t)$ en $t = 22$ para varios valores de $\beta_1 = -r$ y temperaturas iniciales del cuerpo. Los β_1 que se acercan a 22 °C a las 22 horas están entre 0.194 y 0.302.

Entonces se llega a que:

$$\beta_1 \sim \text{Normal}(-0.248, 0.018)$$

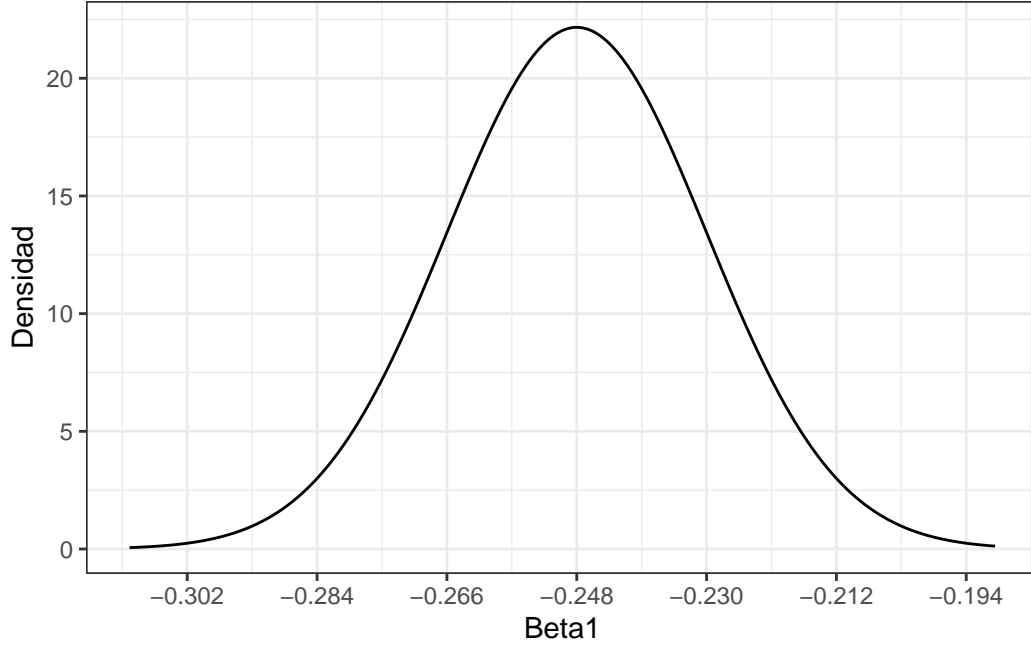


Figura 4: Distribución a priori de beta 1

h_m : Horario de la muerte

Como t_i depende de cuándo Sergio murió, y no se tiene dicha información, se le debe asignar una distribución a *priori*. Entonces se puede separar t_i en $t_i = h_i - h_m$.

Donde h_m es la hora en que se murió Sergio y h_i es la hora de la i -ésima observación.

Como hubo ruidos extraños a las 5:33 tiene sentido pensar que la muerte pudo haber ocurrido cerca de esa hora, y además no pudo haber ocurrido luego de las 5:52 ya que la policía lo hubiera notado. Entonces se plantea la siguiente distribución a *priori*:

$$h_m \sim (-1) * \text{Gamma}(5, 0.1) + 5.883$$

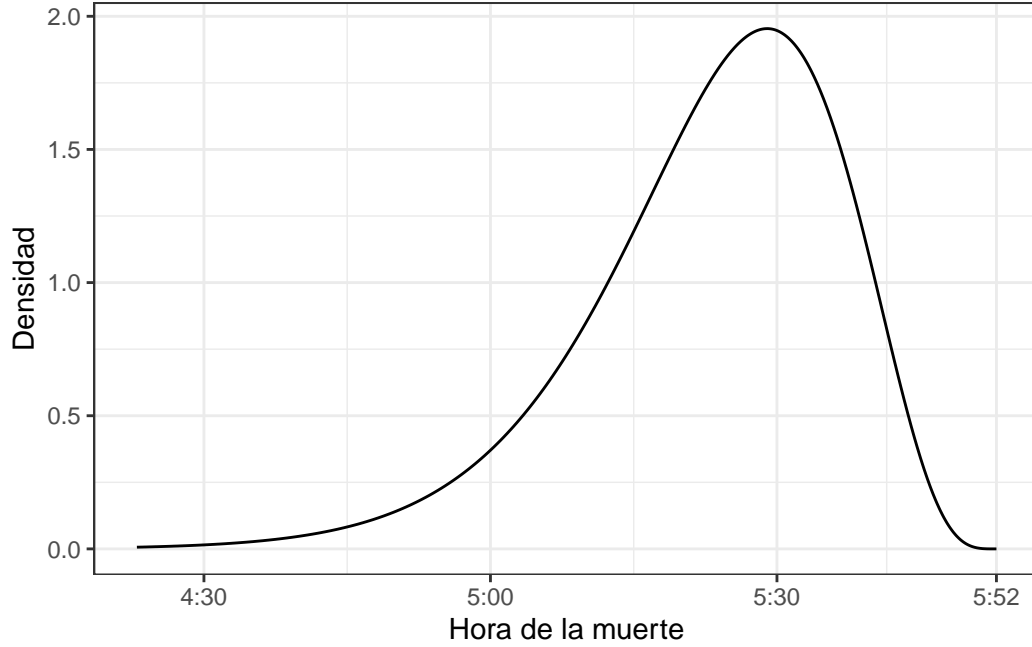


Figura 5: Distribución a priori de la hora de muerte

σ : Error de medición

Al investigar sobre el error de medición para la temperatura, se encuentra que el error máximo de un termómetro es de 0.7°C , entonces se calcula la diferencia entre el logaritmo natural de una temperatura T y la temperatura con su error máximo:

$$\ln(T + 0.7) - \ln(T) = \ln\left(\frac{T + 0.7}{T}\right) = \ln\left(1 + \frac{0.7}{T}\right)$$

Y esta expresión será máxima cuando T sea mínimo, es decir, $T = 22$. Entonces,

$$\ln\left(1 + \frac{0.7}{22}\right) \approx 0.03$$

Por lo tanto, el *prior* resultante es:

$$\sigma \sim \text{Normal}_{+}(0, 0.01)$$

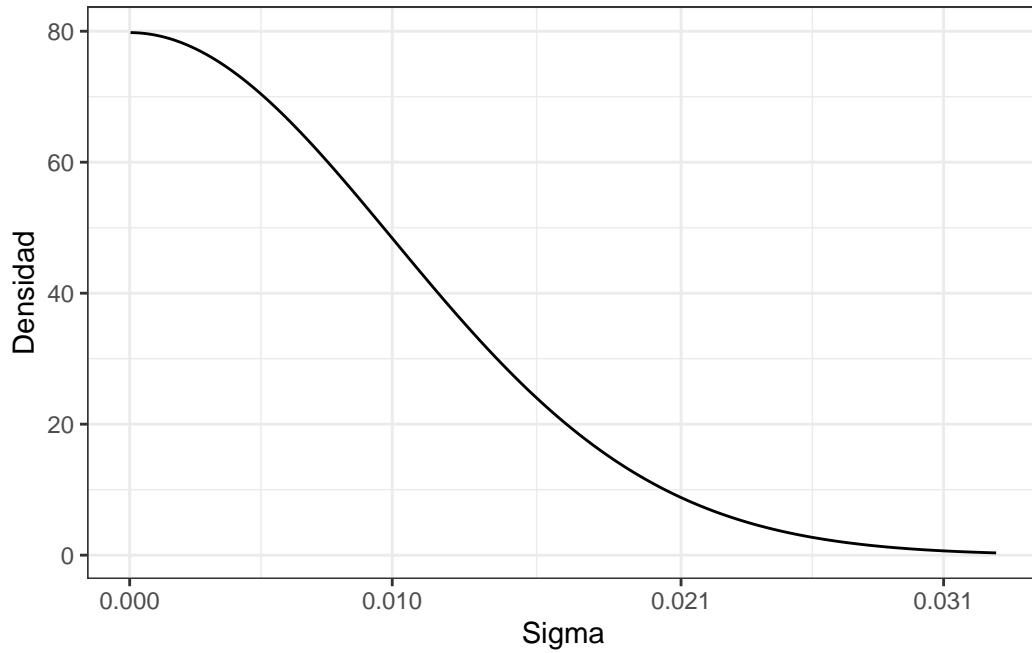


Figura 6: Distribución a priori del error de medición

Donde dichos valores representan el error de medición en términos del porcentaje de la temperatura medida.

Descenso de la temperatura a *priori*

Una vez definidas las distribuciones a *priori* de todos los parámetros de los que depende el cambio de temperatura del cuerpo a través del tiempo, esta se puede graficar para determinar si lo que uno supone es razonable.

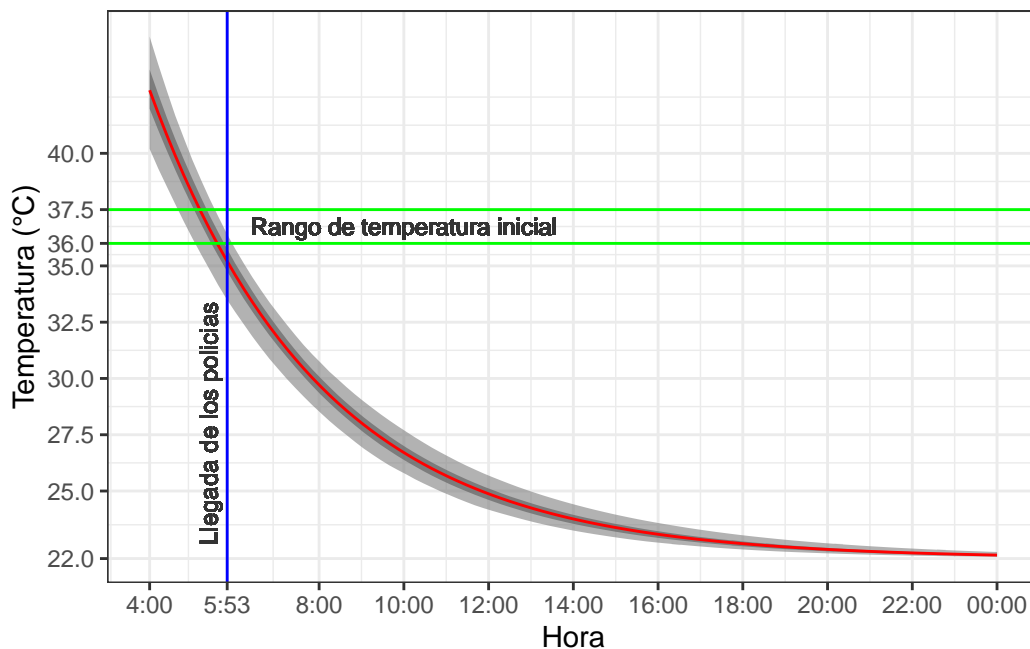


Figura 7: Distribución a priori de la temperatura del cuerpo de Sergio

En la figura 7 se observa que entre las 4:30 y las 6:00 es donde el cuerpo tiene una temperatura normal para una persona viva, es decir, se está prediciendo a la hora de muerte en ese rango horario.

Capítulo 3: Doctor formulemelo

Después de volver de la estación de servicio, mientras los demás miembros del equipo sacan fotos y recogen pruebas de la escena, nuestro protagonista observa detenidamente el proceso reflexionando sobre el ambiente. -“¿Hoy es un día lúgubre porque ocurrió un asesinato, u ocurrió un asesinato porque es un día lúgubre?” Pero antes de poder continuar con sus meditaciones, observa atentamente cómo el doctor le toma la temperatura al cadáver antes de envolverlo en una bolsa de óbito. -“Esta vez tiene una temperatura de 30.5 °C.” dice el doctor con súbito sosiego. -“Ahora lo transportan a la morgue en Capital para que le hagan la autopsia. ¿Querés los resultados?”. “Bueno doc, cuando los tengas mandámelos.”. Al tener todos los datos que creía necesario, Guido suspira con anticipo y se dirige al auto para que lo devuelvan a su casa.

Posterior de los parámetros con una observación

Si se utiliza toda la información postulada en el capítulo 2, se puede plantear el siguiente modelo con los supuestos detallados en el mismo capítulo.

$$Y_i | \beta_0, \beta_1, h_m, \sigma \sim \text{Normal}(\mu_i, \sigma)$$

Con: $\mu_i = \beta_0 + \beta_1(h_i - h_m)$

La temperatura del cuerpo a las 6:45 es de 32.8 °C, es decir, $h_1 = 6.75$ y $Y_1 = \ln(32.8 - 22) = 2.55$. Entonces, se obtiene la distribución a *posteriori* del modelo al incorporar esta información. Esto se logra mediante la aplicación del algoritmo Hamiltonian Montecarlo a través del lenguaje Stan.

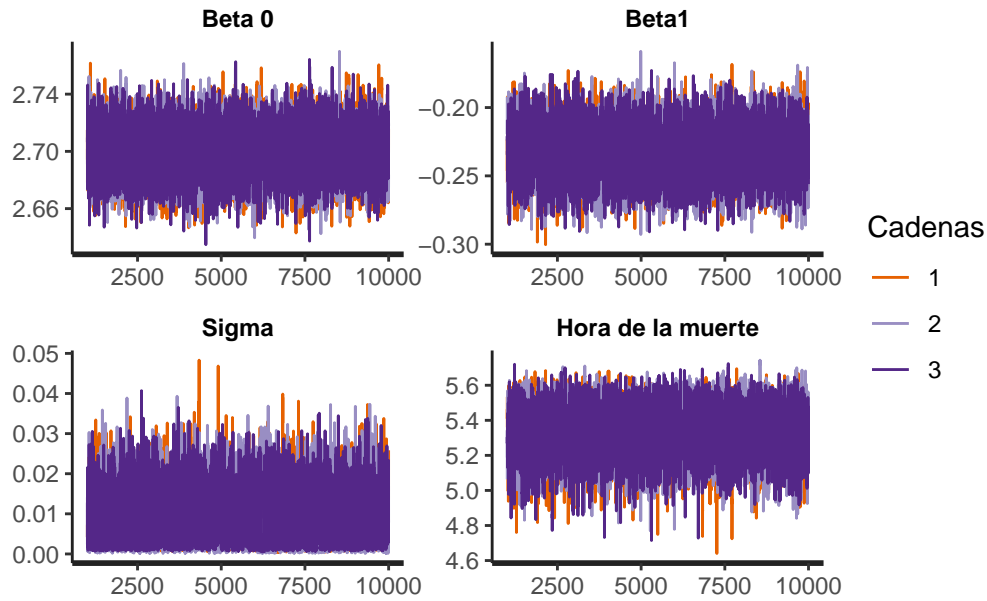


Figura 8: Trazas de las cadenas de cada parámetro para el modelo con una observación

Todos los parámetros tienen un \hat{R} igual a 1, por lo que las distintas cadenas convergen a la misma distribución. Además, las trazas para cada cadena parece seguir un orden aleatorio por lo que se considera que la muestra extraída es representativa de la distribución del *posterior*. Algunas medidas de las muestras extraídas son:

Parámetro	2.5%	Media	97.5%	N_{eff}
β_0	2.67	2.7	2.73	8002
β_1	-0.27	-0.23	-0.22	1694
σ	0	0.01	0.02	1981
h_m	5.06	5.35	5.58	2636

Tabla 1: Estadísticas del *posterior* de cada parámetro con una observación

Una vez ajustado el modelo, consecuentemente se obtiene la distribución a *posteriori* de la hora de muerte de Sergio.

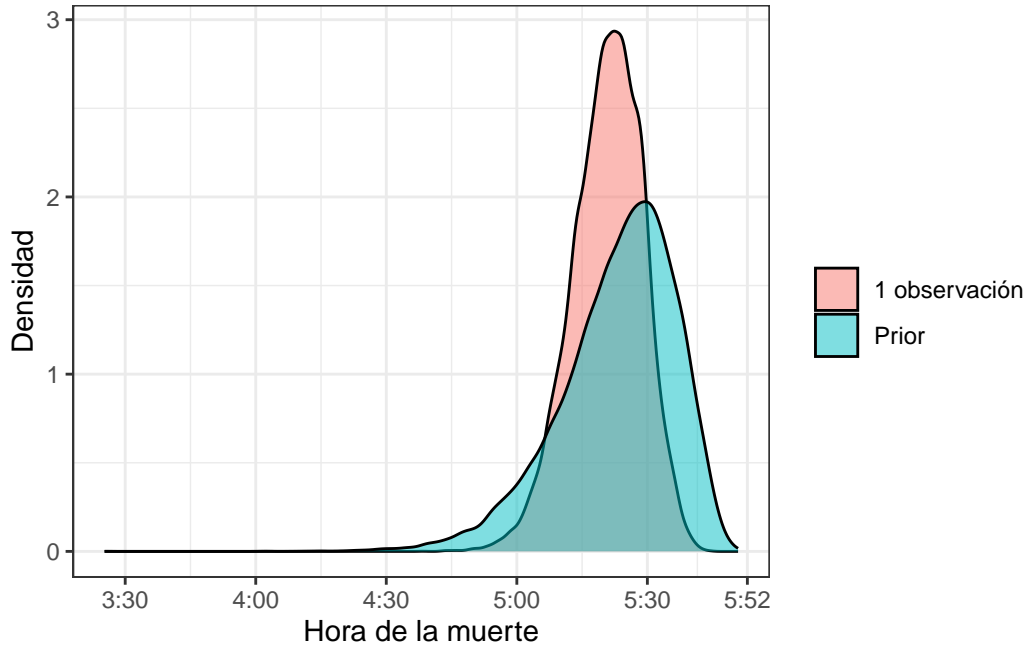


Figura 9: Distribución a posteriori de la hora de muerte

Utilizando solo la primera medición de la temperatura del cuerpo, con una probabilidad del 95%, la hora de muerte se encuentra entre las 5:04 y las 5:35, y su mediana se da en las 5:21.

Al incorporar nueva información, se redujo la incertidumbre sobre la hora de muerte de Sergio, pero al usar tan solo una observación, realmente no se puede obtener una mejoría sustancial de la creencia de los parámetros β_0 , β_1 y σ , ya que estos necesitan más de una observación para poder ser estimados correctamente, y por lo tanto, reducir la incerteza sobre ellos.

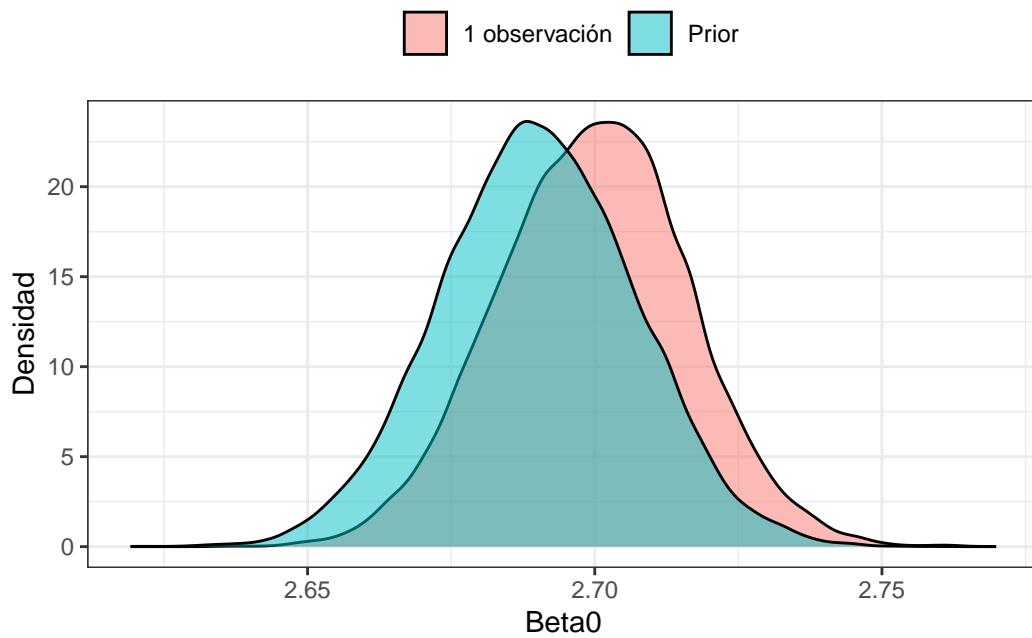


Figura 10: Distribución a posteriori de β_0

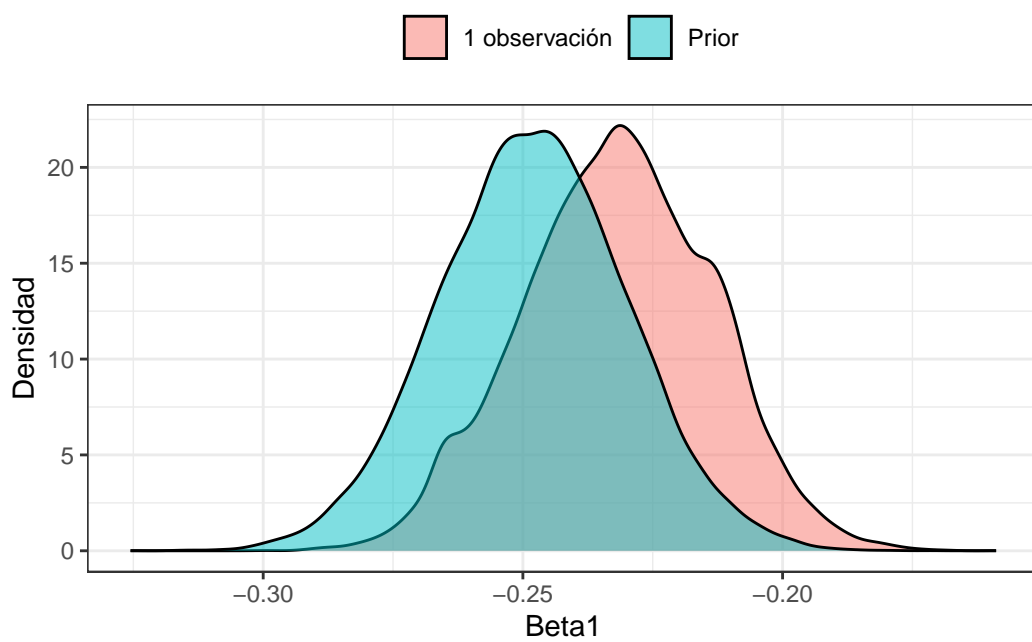


Figura 11: Distribución a posteriori de β_1

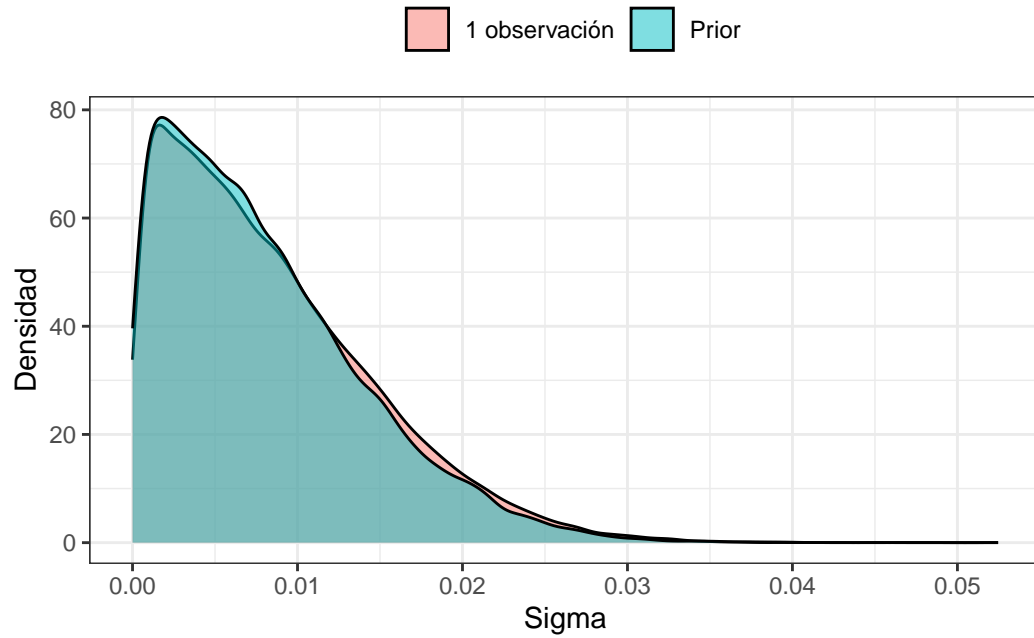


Figura 12: Distribución a posteriori del error de medición

Al actualizar el modelo con una observación y realizar la transformación necesaria para volver a la escala original, se obtiene la siguiente predicción de la temperatura de Sergio a través del tiempo.

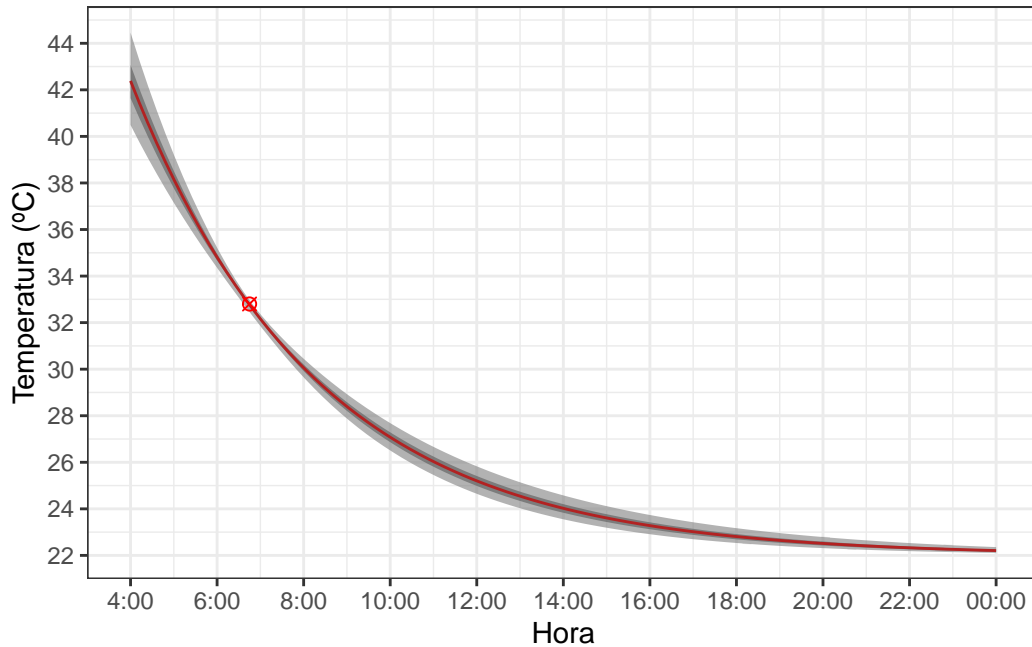


Figura 13: Temperatura del cuerpo de Sergio incorporando una observación al modelo

Al tener una sola observación, no se puede obtener una representación fidedigna de la hora de muerte.

Capítulo 4: Crónica de una catástrofe bayesiana

Luego de volver a su casa Guido se prepara un café y comienza a armar el informe sobre el caso. Utilizando toda la información que recolectó, estima la hora de muerte de Sergio. Luego, por la tarde, le llega a su casilla de mail los resultados de la autopsia con una temperatura del cuerpo de 23.7 °C al momento de hacerla. Con un semblante de confianza, Guido decide agregar esta observación a su análisis. En menos de lo que canta un gallo, toda su seguridad en sí mismo desaparece inadvertidamente.

Posterior de los parámetros con dos observaciones

En el capítulo anterior se trabajó con toda la información obtenida hasta el capítulo 2, pero es un sinsentido no utilizar la totalidad de lo que se sabe. Incorporando al modelo la segunda observación de 30.5 °C de temperatura del cuerpo a las 8:15, se obtienen los siguientes resultados.

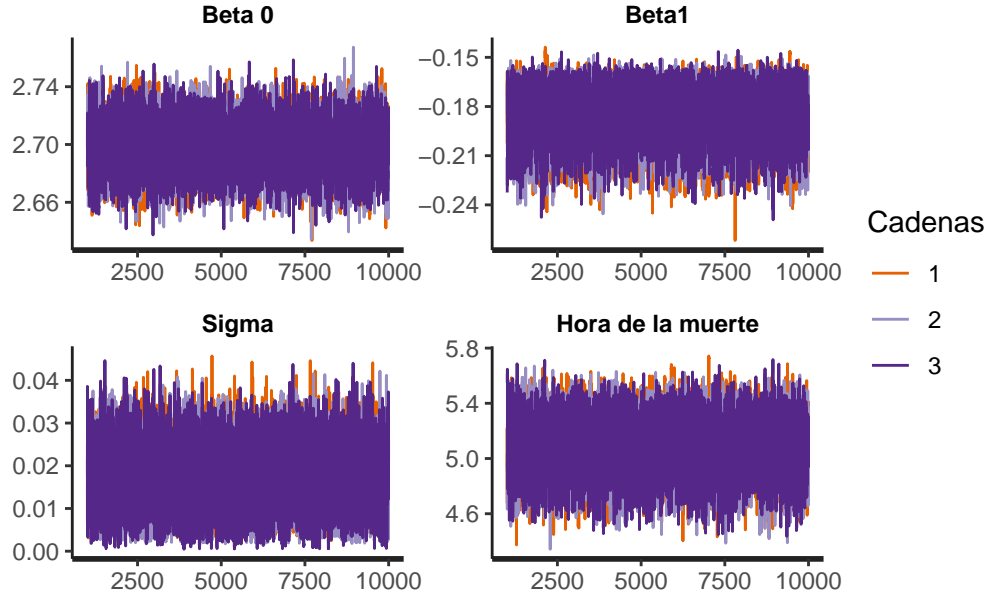


Figura 14: Trazas de las cadenas de cada parámetro para el modelo con dos observaciones

Todos los parámetros tienen un \hat{R} igual a 1, por lo que las distintas cadenas convergen a la misma distribución. Además, los saltos entre iteraciones parecen hacerse de manera aleatoria por lo que se considera que las muestras extraídas de los *posteriors* son buenas.

Parámetro	2.5%	Media	97.5%	N_{eff}
β_0	2.67	2.70	2.73	9043
β_1	-0.22	-0.19	-0.16	4989
σ	0.00	0.02	0.03	5530
h_m	4.67	5.09	5.49	5222

Tabla 2: Estadísticas del *posterior* de cada parámetro con dos observaciones

Para observar mejor los cambios de los *posteriors* entre el modelo ajustado con una observación y el ajustado con dos, se presentan los siguientes gráficos:

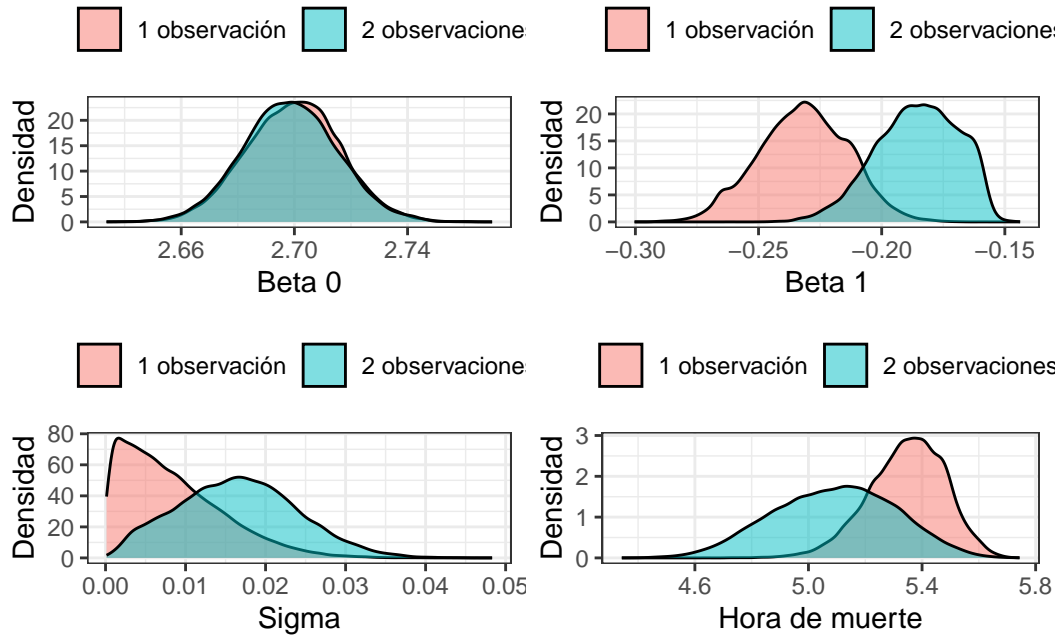


Figura 15: Comparación de los posteriors

En la figura 15 se observa que los *posteriors* de β_1 , σ y la hora de muerte cambian sustancialmente. La distribución de β_1 se traslada hacia la derecha, por lo que la velocidad a la que el cuerpo pierde calor es menor. A la distribución de σ le ocurre lo mismo, por lo que el error de medición será más grande. Y por último, la creencia de la hora de muerte se ve desplazada a la izquierda y se torna más variable, por lo tanto, el nuevo intervalo de confianza para ella será de las 4:40 y las 5:29.

Al actualizar el modelo con la segunda observación y realizar la transformación necesaria para volver a la escala original, se obtiene la siguiente predicción de la temperatura de Sergio a través del tiempo.

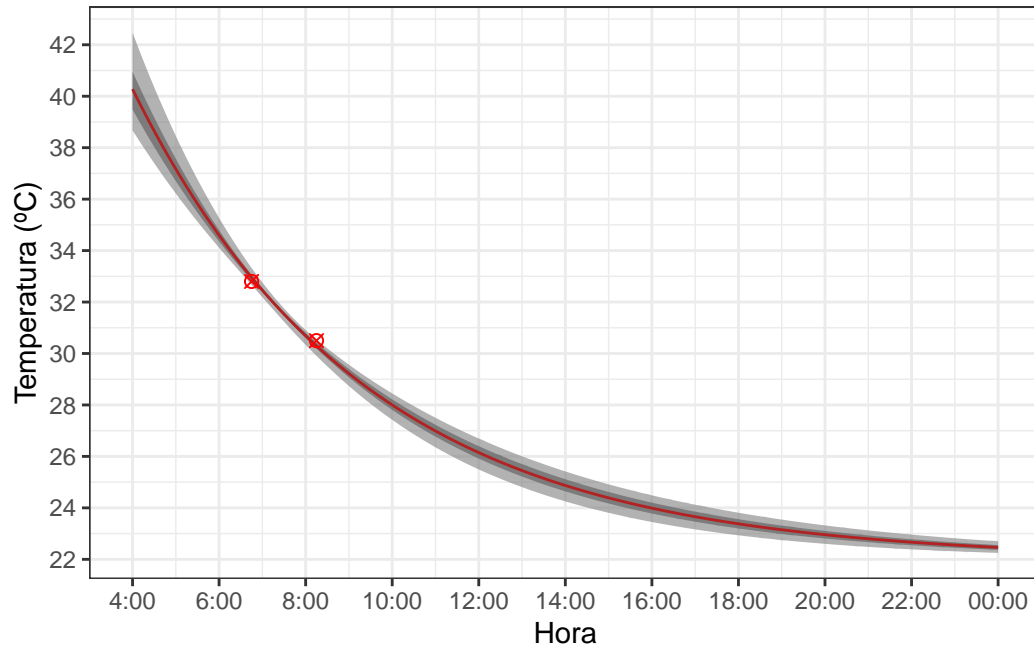


Figura 16: Temperatura del cuerpo de Sergio incorporando dos observaciones al modelo

Para garantizar de que el ajuste es correcto, se realiza una validación interna donde se obtienen réplicas de la medición de la temperatura para distintas temperaturas posibles del cuerpo de Sergio en el momento que se obtuvieron los datos reales y comparándolos.

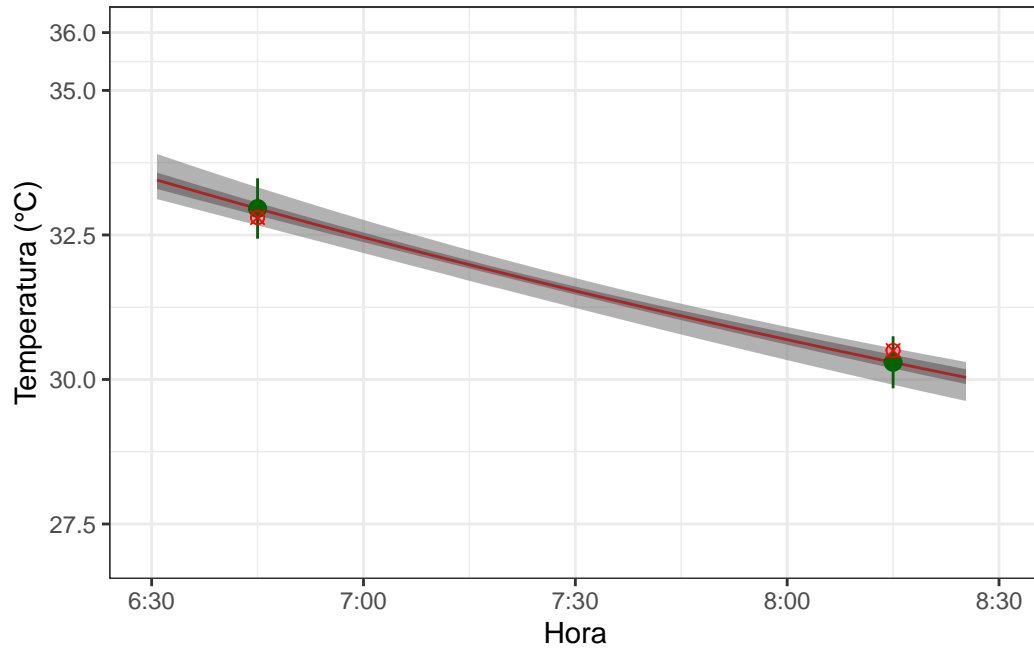


Figura 17: Validación interna de los datos observados

Al estar los puntos reales contenidos adentro de la distribución de las réplicas del conjunto de datos, se considera válido el ajuste.

Posterior de los parámetros con tres observaciones

Teniendo en cuenta la idea de usar toda la información, se utiliza la última temperatura brindada de 23.7 °C para actualizar el modelo.

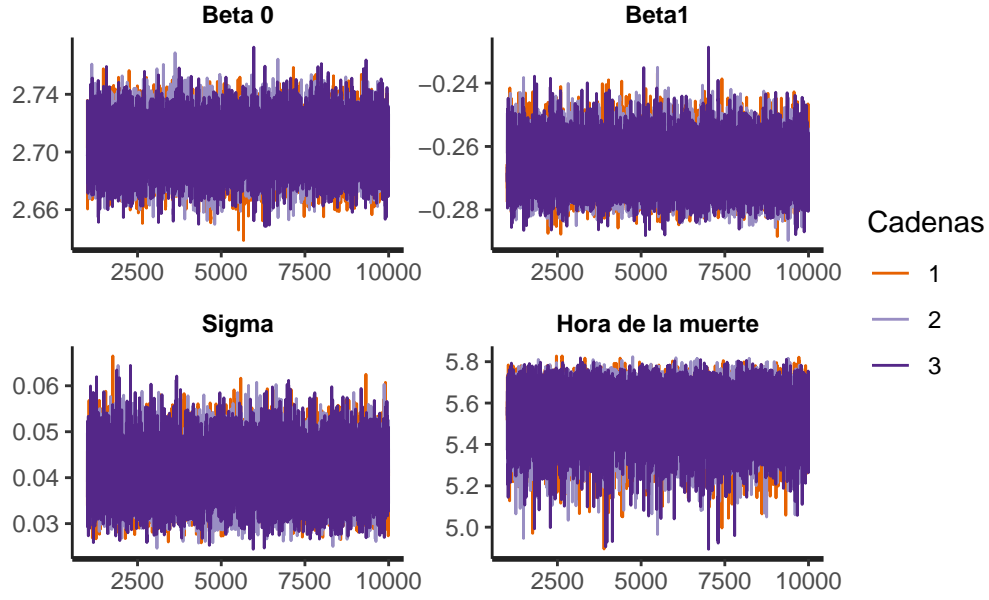


Figura 18: Trazas de las cadenas de cada parámetro para el modelo con tres observaciones

Todos los parámetros tienen un \hat{R} igual a 1, por lo que las distintas cadenas convergen a la misma distribución. Las trazas de cada cadena parecieran distribuirse aleatoriamente por el espacio de probabilidades logrando así una buena representatividad del *posterior*. Algunas de las características de dichas distribuciones son las siguientes:

Parámetro	2.5%	Media	97.5%	N_{eff}
β_0	2.67	2.70	2.72	14439
β_1	-0.28	-0.27	-0.25	11956
σ	0.03	0.04	0.05	13525
h_m	5.29	5.55	5.75	11634

Tabla 3: Estadísticas del *posterior* de cada parámetro con dos observaciones

En un principio, no se ve nada raro con las estimaciones del *posterior* en este nuevo modelo. Por lo que se procede a compararlo con el anterior modelo:

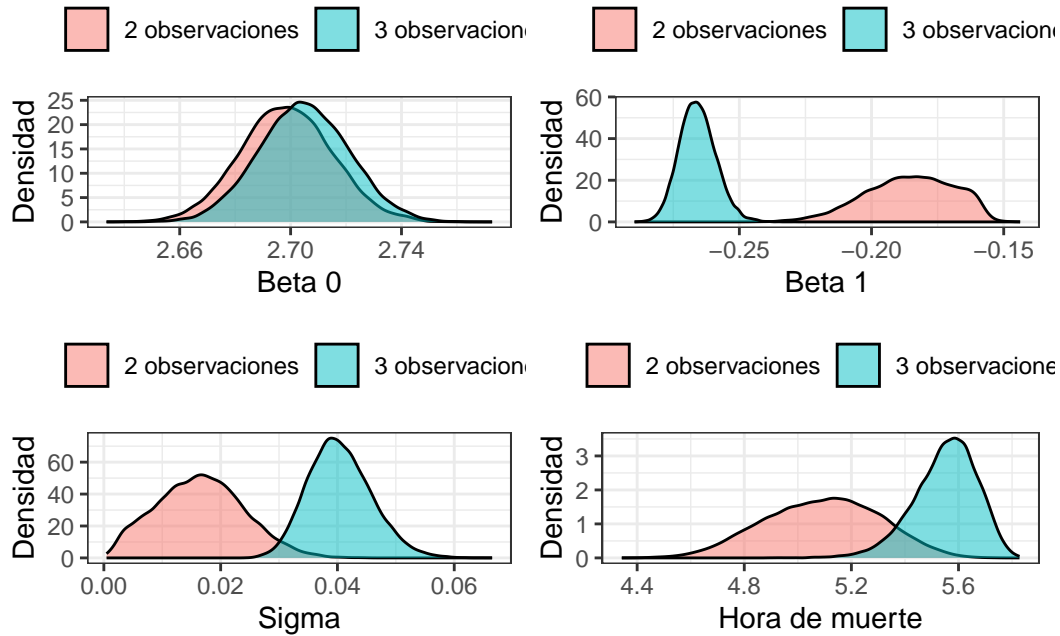


Figura 19: Comparación de los posteriors

Los *posteriors* de ambos modelos se encuentran muy desplazados entre sí, exceptuando el del β_0 , sucediendo esto al agregar un solo dato más. La distribución del β_1 se encuentra muy desplazada a valores menores en este nuevo modelo, es decir, el intercambio de temperatura del cuerpo con el ambiente es más rápido. La distribución de σ se encuentra desplazada a la derecha, lo que implica que el error de medición podría llegar a ser mayor al que se pensaba. Por último, la hora de muerte se ve desplazada a horarios más cercanos al punto de encuentro con el cadáver, influenciado por la nueva estimación de un descenso rápido en su temperatura. Esta nueva estimación de la hora de muerte garantiza que el hecho se concretó entre las 5:17 y las 5:44.

Estos grandes cambios al actualizar el *posterior* con un único dato no se deben pasar por alto, ya que no resultan del todo lógico, por lo que se procede a hacer una validación interna del modelo.

Warning: Removed 120 rows containing missing values or values outside the scale range (``geom_line()``).

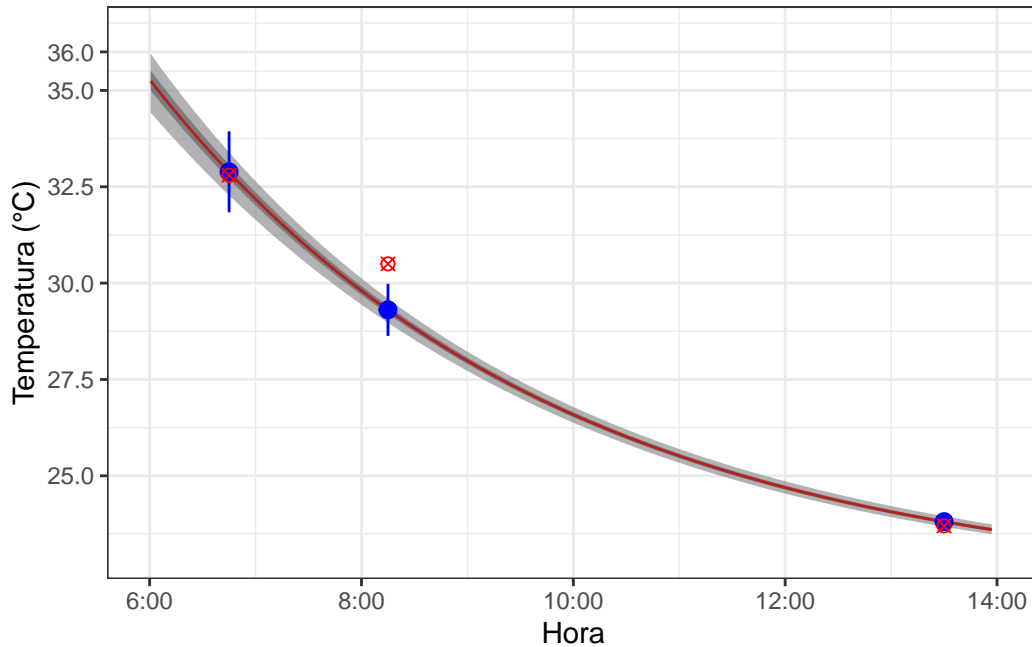


Figura 20: Validación interna del modelo con 3 observaciones

Según el modelo con 3 observaciones, la distribución de la temperatura del cuerpo de Sergio estimada no coincide con los datos observados. Esto es un indicio de que el modelo planteado no se ajusta bien al comportamiento observado. Puede haber dos razones para explicar este malfuncionamiento del modelo: la primera es que alguno/s de los datos observados son erróneos, por ejemplo, el cuerpo fue expuesto a diferentes condiciones que modifican el intercambio de temperatura con el ambiente, y la segunda razón, que el modelo planteado no sea correcto para representar dicho intercambio.

Capítulo 5: “D.e.a.t.h.” Divagaciones entre aficionados tanteando hipótesis

“-Quien hubiera pensado que la muerte fuera tan compleja.” dice Guido sarcásticamente, castigándose a si mismo por no haber podido contener a su propia desmedida arrogancia de ser motivo de semejante desidia. Sintiéndose tonto por pensar que en su primer trabajo todo iba a salir bien, Guido continúa trabajando hasta tarde para poder entregar su informe a tiempo.

Falencias del ajuste

Indudablemente, el modelo con 3 observaciones no es uno que se pueda usar para realizar una estimación certera de la hora muerte. Una de las posibles causas es que la última medición de la temperatura de Sergio no es verídica. Al cambiar las condiciones donde el cuerpo reside y además perder el seguimiento hasta su autopsia, las posibilidades de que la última medición sea incorrecta son altas.

O bien, en el caso de que las tres observaciones sean verídicas, el planteamiento teórico del modelo y del intercambio de temperatura pueden llegar a ser erróneos, por lo que los datos observados no pueden tomar la forma que se espera.

Por lo postulado anteriormente, de los tres modelos planteados, el mejor podría llegar a ser aquel que utiliza 2 observaciones, aunque no existe seguridad alguna de ello.

Adicionalmente, todos los modelos propuestos suponen que Sergio se encontraba en un estado saludable al momento de morir. El incumplimiento de dicha suposición puede perjudicar las estimaciones al usar distribuciones *a priori* incorrectas. En este caso, sería correcto aumentar la incertidumbre sobre el β_0 , además de que la construcción de los posibles valores para β_1 , por cómo se estima en este trabajo, cambiaría porque dependen del β_0 .

Entonces, ¿cuándo murió Sergio?

Epílogo: ¡Y Guido pudo salir!

Luego de entregar su informe, Guido sintió una sensación ambivalente de satisfacción y disconformidad “-Al menos ya todo terminó.” dice el fatigado estudiante y sin ningún tipo de reparo en su vestimenta, se desploma en su lecho para poder tener unas ganadas horas de sueño.

Conclusiones

Si se tuviese que elegir algún modelo de los planteados para predecir la hora de muerte de Sergio, se utilizaría el de 2 observaciones. Entonces, la hora estimada se encuentra entre las 4:40 y las 5:29. Pero no se tiene ninguna certeza de que dicho modelo sea correcto. Para poder plantear un mejor modelo que se asemeje a la realidad del descenso de la temperatura de un cuerpo humano a través del tiempo se requiere una profunda indagación en la literatura de dicho efecto fisiológico y físico.

Anexo

Lo còdigos para generar este trabajo se encuentran en el siguiente repositorio:

https://github.com/TomasAnderson0/TP3_Bayes