**Explicación del código del paquete SEROFOI**

**Las librerías**

Las siguientes son librerías necesarias para el paquete de serofoi, la descripción de su uso se encuentra a continuación.

**Rstan:** El paquete R rstan proporciona RStan, la interfaz R para Stan. El paquete rstan permite ajustar convenientemente los modelos Stan de R (R Core Team 2014) y acceder a la salida, incluidas las inferencias posteriores y las cantidades intermedias, como las evaluaciones de la densidad logarítmica posterior y sus gradientes.

**Tidyverse:** El tidyverse es una colección obstinada de paquetes R diseñados para la ciencia de datos. Todos los paquetes comparten una filosofía de diseño, gramática y estructuras de datos subyacentes. El tidyverse central incluye los paquetes que probablemente usará en los análisis de datos cotidianos.

**Reshape2:** Reshape2 es un reinicio del paquete de remodelación. Reshape2 usa ese conocimiento para crear un nuevo paquete para remodelar datos que es mucho más centrado y mucho más rápido. Existe una librería de tidyverse que puede reemplazarla.

**Bayesplot:** bayesplot es un paquete de R que proporciona una amplia biblioteca de funciones de trazado para su uso después de ajustar modelos bayesianos (normalmente con MCMC). Los gráficos creados por bayesplot son objetos ggplot, lo que significa que después de crear un gráfico, se puede personalizar aún más utilizando varias funciones del paquete ggplot2.

**Loo:** Validación cruzada eficiente aproximada de dejar uno fuera (LOO) para el ajuste de modelos bayesianos utilizando la cadena de Markov Monte Carlo. La aproximación utiliza el muestreo de importancia suavizado de Pareto (PSIS), un nuevo procedimiento para regularizar los pesos de importancia. Como subproducto de los cálculos, también obtenemos errores estándar aproximados para errores predictivos estimados y para la comparación de errores predictivos entre modelos. El paquete también proporciona métodos para utilizar el apilamiento y otras técnicas de ponderación de modelos para promediar las distribuciones predictivas bayesianas.

**Pracma:** Proporciona una gran cantidad de funciones de análisis numérico y álgebra lineal, optimización numérica, ecuaciones diferenciales, series de tiempo, además de algunas funciones matemáticas especiales conocidas. Utiliza nombres de funciones de matlab cuando corresponda para simplificar la transferencia.

**Cowplot:** El paquete cowplot proporciona varias funciones que ayudan a crear figuras con calidad de publicación, como un conjunto de temas, funciones para alinear gráficos y organizarlos en figuras compuestas complejas, y funciones que facilitan la anotación de gráficos o la combinación de gráficos con imágenes.

**Grid:** El paquete grid en R implementa las funciones gráficas primitivas que subyacen en el sistema de trazado ggplot2. El paquete grid proporciona un conjunto de funciones y clases que representan objetos gráficos o grobs, que se pueden manipular como cualquier otro objeto R.

**GridExtra:** Proporciona una serie de funciones a nivel de usuario para trabajar con gráficos de "cuadrícula", en particular para organizar múltiples gráficos basados en cuadrículas en una página y dibujar tablas.

**Hmisc:** La librería Hmisc contiene muchas funciones útiles para el análisis de datos, gráficos de alto nivel, operaciones de servicios públicos, funciones para calcular el tamaño y la potencia de la muestra, traducir conjuntos de datos SAS a R, imputar valores faltantes, creación avanzada de tablas, agrupamiento de variables, manipulación de cadenas de caracteres, conversión de objetos R a código LaTeX, las variables de codificación y el análisis de medidas repetidas de arranque.

**Dplyr:** Uno de los paquetes principales de tidyverse en el lenguaje de programación R, dplyr es principalmente un conjunto de funciones diseñadas para permitir la manipulación de marcos de datos de una manera intuitiva y fácil de usar. Se suele utilizar dplyr para transformar los conjuntos de datos existentes en un formato más adecuado para algún tipo particular de análisis o visualización de datos.

**Epitrix:** Una colección de pequeñas funciones útiles para el análisis de epidemias y el modelado de enfermedades infecciosas. Esto incluye el cálculo de los números básicos de reproducción a partir de las tasas de crecimiento, la generación de etiquetas hash para anonimizar los datos y el ajuste de distribuciones Gamma discretas.

**Gsubfn:** La función gsubfn es como gsub pero puede tomar una función de reemplazo u otros objetos en lugar de la cadena de reemplazo. Las coincidencias y las referencias anteriores se ingresan en la función de reemplazo y se reemplazan por la salida de la función. gsubfn se puede usar para dividir cadenas según el contenido en lugar de los delimitadores y para la interpolación de cadenas al estilo cuasi-perl. El paquete también tiene facilidades para traducir fórmulas a funciones y permitir tales fórmulas en llamadas a funciones en lugar de funciones. Esto se puede usar con funciones de R como apply, sapply, lapply, optim, integra, xyplot, Filter y cualquier otra función que espere otra función como argumento de entrada o funciones como llamadas cat o sql que pueden involucrar cadenas donde la sustitución es deseable. También hay una función para devolver múltiples objetos de funciones y una versión de transform que permite que RHS se refiera a LHS utilizado en la misma transform.

**vscDebbuger:** El depurador funciona de la siguiente manera: se inicia un proceso R dentro de un proceso secundario. Se carga el paquete R vscDebugger. El depurador inicia y controla los programas R enviando información a la entrada estándar del proceso secundario

**Las funciones**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Nombre** | **Agr. código** | **Agr. esquema** | **Descripción** | **Comentarios** |
| **1** | dat0 <- readRDS | Main | read\_data | Lee los datos de la carpeta data que están en formato RDS |  |
| **2** | Model i <- readRDS (i=0,1,2) | Main | declare\_model | Lee los modelos contenidos en la carpeta stanmodels |  |
| **3** | my\_dir | Main | define\_model | La carpeta donde los resultados serán guardados |  |
| **4** | RunSaveModels | fRun |  | Corre los modelos con un número de iteraciones dado |  |
| **5** | fRun | fRun | declare\_model  fit\_model  summary  define\_comparison | Las funciones presentes en fRun toman los datos para los tres modelos a probar y los corren. Luego, hacen un ajuste (fit), plotean y hacen una revisión estadística. Finalmente compara y guarda el mejor de los tres modelos, genera un plot del resultado, toma el tiempo que llevó la compilación y guarda en RDS los resultados de la encuesta y el fitting. | Puede tomar un tiempo significativo |
| **6** | fUtility | fUtility | define\_model  test\_hypothesis  summary | Las funciones presentes en fUtility tienen como fin entre otros obtener la matriz de exposición, la prevalencia extendida y obtener de los datos la prevalencia predicha y observada (data.frame y dat) para finalmente obtener la prevalencia final. Por último, hay una función para obtener los residuos. |  |
| **7** | fCheck | fCheck | define\_comparison  calculate\_force\_infection  summary | fCheck contiene funciones para comparar y salvar el mejor de los tres modelos y permite medir la precisión mediante el logaritmo esperado de densidad predictiva puntual (elpd). Se hace una estimación de FOI a partir de los modelos de rstan. |  |
| **8** | fFitting | fFitting | fit\_model | Las funciones de fFitting tienen como objetivo realizar el ajuste de los tres modelos básicos planteados. | Corre bastante rápido. |
| **9** | fPlot | fPlot | plot\_data  plot\_force\_infection  plot\_comparison | fPlot contiene las funciones que se encargan de la gráfica de datos, fuerza de infección (FOI) y la comparación de los modelos. | Corre sin ningún problema y bastante rápido. |
| **10** | Constant\_foi\_Bi | Stanmodels | declare\_design  declare\_model | Se define el modelo para una fuerza de infección constante con un prior dado por una distribución uniforme con parámetros 0 y 2. | Sería útil poner a variar los parámetros de la distribución del prior entre un intervalo de valores con el fin de no dejarlos fijos. |
| **11** | Continuous\_foi\_normal\_Bi | Stanmodels | declare\_design  declare\_model | Se define el modelo para una fuerza de infección variable de manera anual con prior y upper prior dados por distribuciones normales y de Pareto y Cauchy respectivamente. | Variar los parámetros entre valores en un intervalo aumentará irremediablemente el tiempo de cálculo, pero podría hacer más versátil el modelo. |
| **12** | Continuous\_foi\_normal\_log | Stanmodels | declare\_design  declare\_model | Define el modelo para una fuerza de infección variable anual con un prior de baja transmisión dada por el logaritmo de la fuerza de infección. Los prior están dados por distribuciones normales, mientras el upper prior viene dado por una distribución de Cauchy. |  |

**Funciones finales**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Nombre** | **Agr. código** | **Agr. Esquema** | **Descripción** | **Comentarios** |
| **1** | check\_all\_diagnostics | fUtility |  |  |  |
| **2** | cleanObject | fUtility |  |  |  |
| **3** | compare\_and\_save\_best\_model | fCheck |  | Compara y guarda el mejor modelo. |  |
| **4** | dir\_results | fUtility |  | Guarda los resultados en un lugar específico. |  |
| **5** | extract\_and\_save | fCheck |  |  |  |
| **6** | extract\_summary\_mod | fCheck |  |  |  |
| **7** | fCombinedPlots | fPlot |  | Combina los plots en uno solo. |  |
| **8** | fFitModel | fFitting |  | Ajusta los modelos. |  |
| **9** | fFitModel\_log | fFitting |  | Ajusta el modelo log. |  |
| **10** | get\_exposure\_matrix | fUtility |  | Obtiene la matriz de exposición. |  |
| **11** | get\_model\_comparison\_plot | fPlot |  | Obtiene la gráfica para comparar los modelos. |  |
| **12** | get\_posterior\_summary | fUtility |  |  |  |
| **13** | get\_prev\_expanded | fUtility |  |  |  |
| **14** | get\_residuals | fUtility |  | Obtiene los residuos. |  |
| **15** | get\_table\_rhats | fCheck |  | Obtiene la tabla de convergencia de los modelos. |  |
| **16** | make\_yexpo | fUtility |  |  |  |
| **17** | obtain\_prevalence\_extended | fUtility |  |  |  |
| **18** | partition\_div | fUtility |  | Devuelve arreglos de parámetros separados en transiciones divergentes y no divergentes. |  |
| **19** | plot\_info\_table | fPlot |  | Plotea la tabla con información. |  |
| **20** | RunSaveModels | fRun |  | Corre los modelos considerados. |  |
| **21** | sub\_sample | fUtility |  |  |  |
| **22** | thin\_chain | fUtility |  |  |  |
| **23** | vertical\_plot\_arrange\_per\_model | fPlot |  |  |  |
| **24** | get\_data\_binned | fUtility |  | Obtiene datos no agrupados. | No necesaria y su eliminación reduce el tiempo de ejecución. |

**Referencias**

1. <https://www.rdocumentation.org/packages/>
2. <https://cran.r-project.org/web/packages/>