

Waller SISMID 2022: NY Leukemia Spatial Regression with INLA

Lance A. Waller

7/18/2022

New York Leukemia Data Spatial Poisson Regression with INLA Example

NY Leukemia analysis with INLA from https://www.rdocumentation.org/packages/spdep/versions/0.6-15/topics/NY_data and <https://www.r-bloggers.com/spatial-data-analysis-with-inla/>

```
##Load libraries
```

```
library(spdep)
```

```
## Loading required package: sp
```

```
## Loading required package: spData
```

```
## To access larger datasets in this package, install the spDataLarge
```

```
## package with: `install.packages('spDataLarge',
```

```
## repos='https://nowosad.github.io/drat/', type='source')`
```

```
## Loading required package: sf
```

```
## Linking to GEOS 3.9.1, GDAL 3.4.0, PROJ 8.1.1; sf_use_s2() is TRUE
```

```
library(INLA)
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Loading required package: parallel
```

```
## This is INLA_21.11.22 built 2021-11-21 16:13:28 UTC.
```

```
## - See www.r-inla.org/contact-us for how to get help.
```

```
## - To enable PARDISO sparse library; see inla.pardiso()
```

```
# Set my working directory (Lance's here for example)
```

```
# setwd("~/OneDrive - Emory University/meetings/SISMID.2021/SISMID.2021.Waller.Rcode")
```

```
# Read in data (included in the 'foreign' library.)
```

```
library(foreign)
```

```
nydata <- read.dbf(system.file("etc/misc/nydata.dbf", package="spdep")[1])
```

Now format the data for INLA (this needs to be more clearly documented, but these steps put the right values from the dataset into the variable names needed for INLA)

```
# Format data for INLA
```

```
coordinates(nydata) <- c("X", "Y")
```

```
nyadjmat <- as.matrix(read.dbf(system.file("etc/misc/nyadjwts.dbf",  
                                           package="spdep")[1])[-1])
```

```
## Field name: 'Z600701190' changed to: 'Z600701190.1'
```

```

## Field name: 'Z600701190' changed to: 'Z600701190.2'
## Field name: 'Z600701200' changed to: 'Z600701200.1'
## Field name: 'Z600701210' changed to: 'Z600701210.1'
## Field name: 'Z600701210' changed to: 'Z600701210.2'
## Field name: 'Z600701220' changed to: 'Z600701220.1'
## Field name: 'Z600701270' changed to: 'Z600701270.1'
## Field name: 'Z600701320' changed to: 'Z600701320.1'
## Field name: 'Z600701330' changed to: 'Z600701330.1'
## Field name: 'Z605303010' changed to: 'Z605303010.1'
## Field name: 'Z605303010' changed to: 'Z605303010.2'
## Field name: 'Z605303040' changed to: 'Z605303040.1'
## Field name: 'Z605303040' changed to: 'Z605303040.2'
## Field name: 'Z605303050' changed to: 'Z605303050.1'
## Field name: 'Z606700170' changed to: 'Z606700170.1'
## Field name: 'Z606700360' changed to: 'Z606700360.1'
## Field name: 'Z606700560' changed to: 'Z606700560.1'
## Field name: 'Z606700610' changed to: 'Z606700610.1'
## Field name: 'Z606700610' changed to: 'Z606700610.2'
## Field name: 'Z606701100' changed to: 'Z606701100.1'
## Field name: 'Z606701110' changed to: 'Z606701110.1'
## Field name: 'Z606701120' changed to: 'Z606701120.1'
## Field name: 'Z606701120' changed to: 'Z606701120.2'
## Field name: 'Z606701120' changed to: 'Z606701120.3'
## Field name: 'Z606701140' changed to: 'Z606701140.1'
## Field name: 'Z606701520' changed to: 'Z606701520.1'
## Field name: 'Z606701650' changed to: 'Z606701650.1'
## Field name: 'Z606701680' changed to: 'Z606701680.1'
## Field name: 'Z606701690' changed to: 'Z606701690.1'

```

```

ID <- as.character(names(read.dbf(system.file("etc/misc/nyadjwts.dbf",
                                             package="spdep")[1]))[-1])

```

```

## Field name: 'Z600701190' changed to: 'Z600701190.1'
## Field name: 'Z600701190' changed to: 'Z600701190.2'
## Field name: 'Z600701200' changed to: 'Z600701200.1'
## Field name: 'Z600701210' changed to: 'Z600701210.1'
## Field name: 'Z600701210' changed to: 'Z600701210.2'
## Field name: 'Z600701220' changed to: 'Z600701220.1'

```

```
## Field name: 'Z600701270' changed to: 'Z600701270.1'
## Field name: 'Z600701320' changed to: 'Z600701320.1'
## Field name: 'Z600701330' changed to: 'Z600701330.1'
## Field name: 'Z605303010' changed to: 'Z605303010.1'
## Field name: 'Z605303010' changed to: 'Z605303010.2'
## Field name: 'Z605303040' changed to: 'Z605303040.1'
## Field name: 'Z605303040' changed to: 'Z605303040.2'
## Field name: 'Z605303050' changed to: 'Z605303050.1'
## Field name: 'Z606700170' changed to: 'Z606700170.1'
## Field name: 'Z606700360' changed to: 'Z606700360.1'
## Field name: 'Z606700560' changed to: 'Z606700560.1'
## Field name: 'Z606700610' changed to: 'Z606700610.1'
## Field name: 'Z606700610' changed to: 'Z606700610.2'
## Field name: 'Z606701100' changed to: 'Z606701100.1'
## Field name: 'Z606701110' changed to: 'Z606701110.1'
## Field name: 'Z606701120' changed to: 'Z606701120.1'
## Field name: 'Z606701120' changed to: 'Z606701120.2'
## Field name: 'Z606701120' changed to: 'Z606701120.3'
## Field name: 'Z606701140' changed to: 'Z606701140.1'
## Field name: 'Z606701520' changed to: 'Z606701520.1'
## Field name: 'Z606701650' changed to: 'Z606701650.1'
## Field name: 'Z606701680' changed to: 'Z606701680.1'
## Field name: 'Z606701690' changed to: 'Z606701690.1'
identical(substring(ID, 2, 10), substring(as.character(nydata$AREakey), 2, 10))

## [1] TRUE

nyadjlw <- mat2listw(nyadjmat, as.character(nydata$AREakey))
listw_NY <- nb2listw(nyadjlw$neighbours, style="B")
```

Now let's calculate the overall rate (total number of cases over total population size).

Next, find the number of cases in each tract if this overall rate were applied to the tract-level population size.

Finally, calculate the standardized mortality ratio, the number observed divided by the number expected (if the overall rate applied equally everywhere).

```
# Calculate overall rate
rate <- sum(nydata$TRACTCAS) / sum(nydata$POP8)

# Set expected number of cases (population * rate)
Expected <- nydata$POP8 * rate

# Calculate standardized mortality ratio (SMR)
SMR <- nydata$TRACTCAS / nydata$Expected
```

Now fit the ICAR model and output the estimates.

```
# Fit ICAR model
# Assuming here that the order of nydata and nyadjmat are the same
ID_new <- seq(nrow(nydata)) # I think INLA needs the IDs to be 1:n
nydata$ID_new <- ID_new
nyadjlw <- mat2listw(nyadjmat, row.names = ID_new)
listw_NY <- nb2listw(nyadjlw$neighbours, style="B")
# FEED THE FOLLOWING INTO INLA'S graph ARGUMENT instead of listw_NY
nyadj_matrix <- as(nb2mat(nyadjlw$neighbours, style = "B"), "Matrix")

m.icar <- inla(trunc(TRACTCAS) ~ 1 + PCTAGE65P + AVGIDIST + PCTOWNHOME +
              f(ID_new, model = "besag", graph = nyadj_matrix),
              data = as.data.frame(nydata), E = Expected, family = "poisson",
              control.predictor = list(compute = TRUE),
              control.compute = list(dic = TRUE, waic = TRUE))

summary(m.icar)

##
## Call:
##   c("inla(formula = trunc(TRACTCAS) ~ 1 + PCTAGE65P + AVGIDIST +
##     PCTOWNHOME + ", " f(ID_new, model = \"besag\", graph = nyadj_matrix),
##     family = \"poisson\", ", \" data = as.data.frame(nydata), E = Expected,
##     control.compute = list(dic = TRUE, \", \" waic = TRUE), control.predictor
##     = list(compute = TRUE))\" )
## Time used:
##   Pre = 4.07, Running = 0.409, Post = 0.0247, Total = 4.5
## Fixed effects:
##           mean      sd 0.025quant 0.5quant 0.975quant   mode kld
## (Intercept) -0.474 0.174    -0.821   -0.472    -0.139 -0.468   0
## PCTAGE65P    4.174 0.615     2.944    4.181     5.361  4.195   0
## AVGIDIST      0.289 0.085     0.114    0.292     0.447  0.298   0
## PCTOWNHOME   -0.396 0.200    -0.786   -0.398     0.001 -0.400   0
##
## Random effects:
##   Name      Model
##   ID_new Besags ICAR model
##
## Model hyperparameters:
##           mean      sd 0.025quant 0.5quant 0.975quant   mode
## Precision for ID_new 16094.02 19253.05      6.21  9514.72  68581.82  5.56
##
## Deviance Information Criterion (DIC) .....: 968.59
## Deviance Information Criterion (DIC, saturated) ....: -823.21
## Effective number of parameters .....: 7.26
##
## Watanabe-Akaike information criterion (WAIC) ...: 968.91
## Effective number of parameters .....: 8.26
##
## Marginal log-Likelihood: -698.51
## Posterior summaries for the linear predictor and the fitted values are computed
## (Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```