

CSSS 554: Assignment 3

Department of Biostatistics @ University of Washington

Alejandro Hernandez

Winter Quarter 2025

Background

We aim to map lung cancer mortality data for men in the Valencia region of Spain from 1991-2000. For more details on the data, see Martinez-Beneito et al. (2019, Disease Mapping). Expected deaths are adjusted for reference rate only.

1. Let Y_i and E_i , $i = 1, \dots, n$, denote the observed and expected counts in region i , $i = 1, \dots, n$. Then consider the model

$$Y_i | \theta_i \sim \text{Poisson}(E_i \theta_i)$$

- (a) (b) Provide a map of the observed counts Y_i and observed counts E_i .

See Figure 1 for maps of observed and expected lung cancer mortality.

- (c) Provide a map of the standardized mortality rates (SMR), defined as $\text{SMR}_i = \hat{\theta}_i = \frac{Y_i}{E_i}$ for $i = 1, \dots, n$.

See Figure 2 for a map of standardized lung cancer mortality rates. Apart from a few near zero estimates the distribution of SMR is approximately normal, centered at 87%, with about half of the values sitting between 60% and 110%. 27 of the 540 regions have an SMR over 150%, and only 2 are greater than 200%.

- (d) Plot the SMRs versus the estimated standard errors, which are given by $\sqrt{\hat{\theta}_i / E_i}$.

2. In this question we will smooth the SMRs using the disease mapping Poisson-Lognormal model:

$$Y_i | \beta_i, \epsilon_i \sim_{ind} \text{Poisson}(E_i e^{\beta_0} e^{\epsilon_i}) \quad \epsilon_i | \sigma_i \sim N(0, \sigma_e^2)$$

for $i = 1, \dots, n$.

- (a) Using the `inla` function in R fit this model using the default priors for β_0 and σ_e . Report the posterior medians and 95% intervals for β_0 and σ_e .
- (b) Extract the posterior medians of the relative risk (RR) estimates and provide a map of these.
- (c) Plot these posterior RR estimates against the SMRs, and comment.

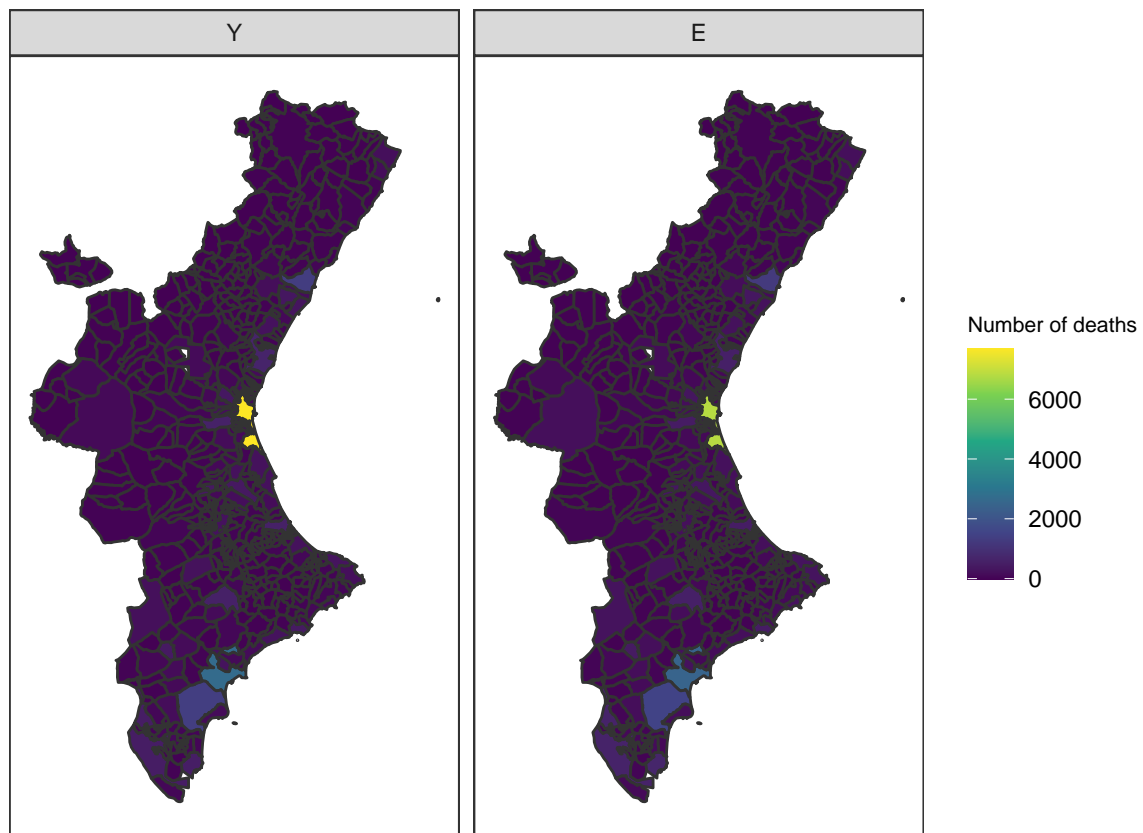


Figure 1: Lung cancer mortality among men in the Valencia region of Spain from 1991-2000.

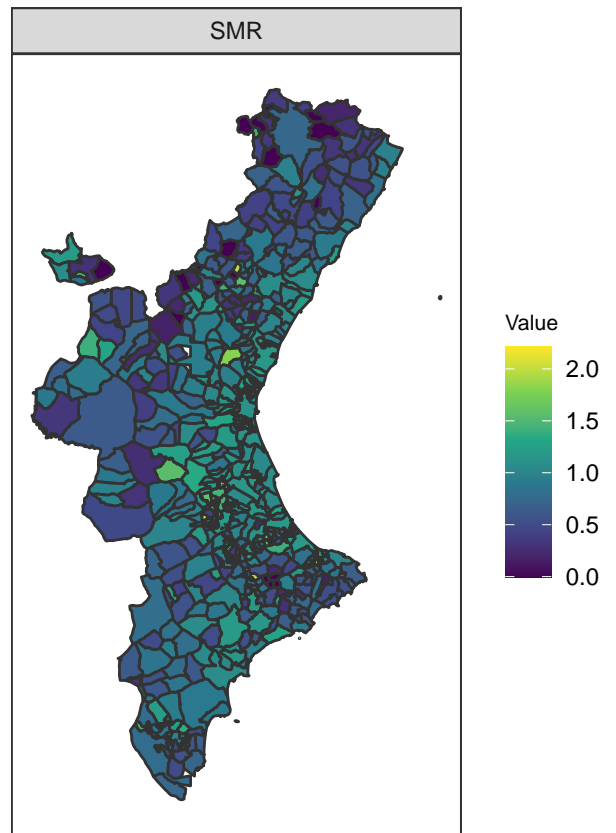


Figure 2: Standardized mortality rate of lung cancer mortality.

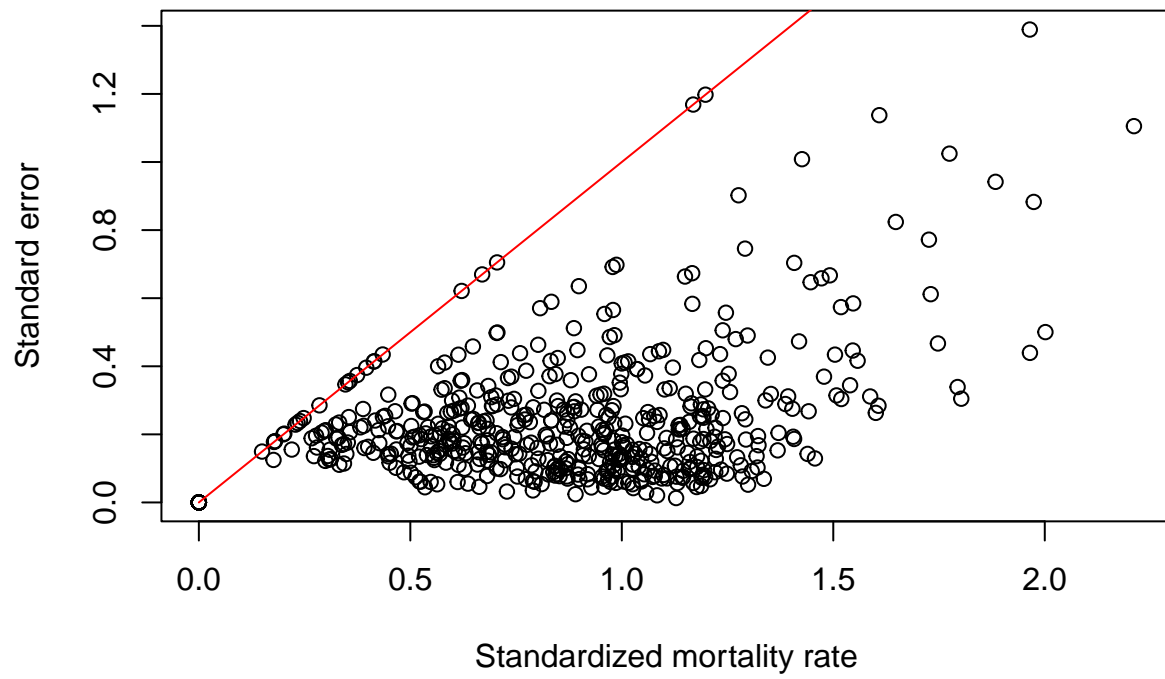


Figure 3: Standardized mortality rate versus estimate standard error

- (d) Plot the posterior standard deviations of the RRs against the standard errors of the SMRs and comment.

3. In this question we will smooth the SMRs using the disease mapping Poisson-LognormalSpatial model:

...

- (a) Using the `inla` function in R fit this model using the `bym2` model, with the default prior for β_0 and the following prior specification for the spatial and non-spatial random effects (note that you must be in the directory that contains the `VR.graph` file):

```
f(Region, model="bym2", graph="VR.graph", scale.model=T, constr=T, hyper=list(phi=list(prior="pc",
param=c(0.5, 0.5), initial=1), prec=list(prior="pc.prec", param=c(0.3,0.01), initial=5)))
```

These choices correspond to the prior belief that there is a 1% chance that the total residual standard deviation is greater than 0.3, and a 50% chance that the proportion of the variance that is spatial is bigger than 0.5.

Report both the posterior medians and 95% intervals for β_0 , the total variance of the random effects, and the proportion of the total variance attributed to the spatial random effect.

- (b) Extract the relative risk estimates and provide a map of these. Compare these estimates with the SMRs and with those obtained from the Poisson-Lognormal model (i.e., the model with IID random effects only) that you fit in Question 2.

4. Bonus Question: Suppose that instead of having available the counts and expected numbers we have access to the relative risks and their standard errors. Take the data as

and fit the model

Fit this model using `inla` and plot the estimated relative risks from this model against the estimates from the Poisson-Lognormal model, and comment.

End of report. Code appendix begins on the next page.

Code Appendix

```
# Clear environment
rm(list=ls())

# Setup options
knitr::opts_chunk$set(echo=FALSE, warning=FALSE, message=FALSE, results='hide')
options(knitr.kable.NA = '-', digits = 2)
labs = knitr::all_labels()
labs = labs[!labs %in% c("setup", "allcode")]
# Load relevant packages
# library(surveyPrev)
# library(sf)
library(SUMMER)

## Load data
load("../data/HW3data.Rdata")
# Shapefile for plotting (VR.cart$CODMUNI corresponds to rownames(Exp.mv3))
geo <- VR.cart

# Make a table of observed (Y) and expected (E) counts
ad1 <- data.frame(ID = rownames(Exp.mv3),
                  Y = data.frame(Obs.mv3)$Lung,
                  E = data.frame(Exp.mv3)$Lung)
# Compute SMR (standardized mortality ratio)
ad1$SMR = with(ad1, Y/E)
# Compute S.E. of SMR
ad1$SMR.se = with(ad1, sqrt(SMR/E))
head(ad1)

#####
#### QUESTION 1 ####
#####

# Map of observed and expected counts
SUMMER::mapPlot(data = ad1, variables = c("Y", "E"),
                values = c("Y", "E"), geo = geo,
                by.data = "ID", by.geo = "CODMUNI",
                legend.label = "Number of deaths")

# Map of SMR
mapPlot(data = ad1, variables = "SMR", values = "SMR",
        geo = geo, by.data = "ID", by.geo = "CODMUNI")

## Distribution of SMR
# Histogram
hist(ad1$SMR, main = "", xlab = "Standardized mortality rate",
     labels = TRUE, breaks = 20)
abline(v=quantile(ad1$SMR, probs = c(.25, .5, .75)), col = "red")
# Summary statistics
summary(ad1$SMR)

plot(x = ad1$SMR, y = ad1$SMR.se,
```

```
xlab = "Standardized mortality rate",  
ylab = "Standard error")  
lines(0:10, 0:10, col = "red")
```

```
#####  
#### QUESTION 2 ####  
#####
```

```
#####  
#### QUESTION 3 ####  
#####
```

```
#####  
#### QUESTION 4 ####  
#####
```

End of document.