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.

Analysis

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

$Y_i|\theta_i \sim Poisson(E_i\theta_i)$

(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 $\text{SMR}_\text{i} = \hat{\theta}_i = \frac{Y_i}{E_i}$ for $i = 1, ..., n$.

See Figure 2 for a map of standardized lung cancer mortality rates. Apart from a few near zero estimates, its distribution is approximately normal. SMR is 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 $\sqrt{\hat{\theta}_i/E_i}$.

See Figure 3 for the standardized mortality rates versus their estimated standard errors. The difference between an SMR estimate and its standard error increases as SMR grows.

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

$Y_i|\beta_i, \epsilon_i \sim_{ind} Poisson(E_i e^{\beta_0} e^{e_i})$
$e_i|\sigma_e^2 \sim_{iid} N(0, \sigma_e^2)$
for i, $i = 1, ..., n$.

(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.

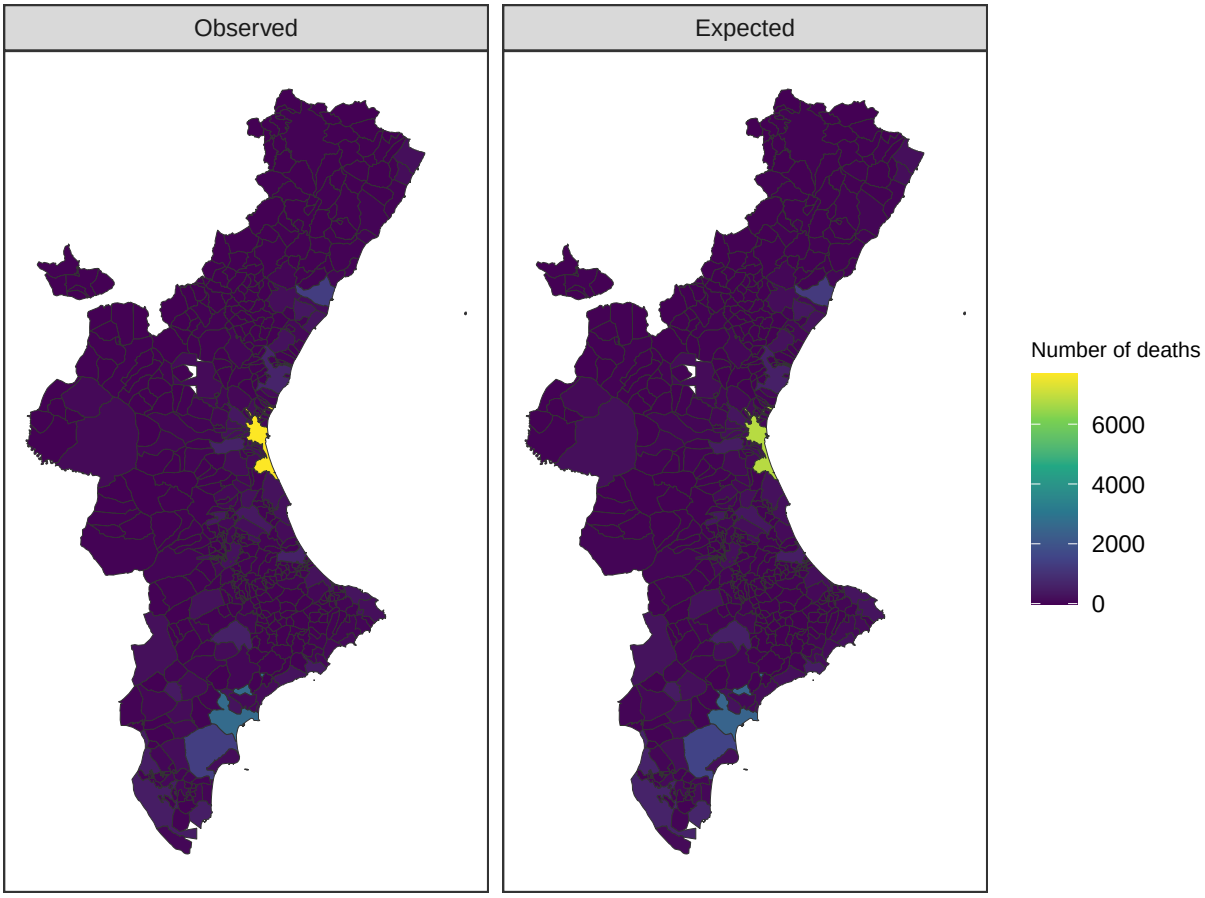

Figure 1: Lung cancer mortality among men in the Valencia region of Spain from 1991-2000.

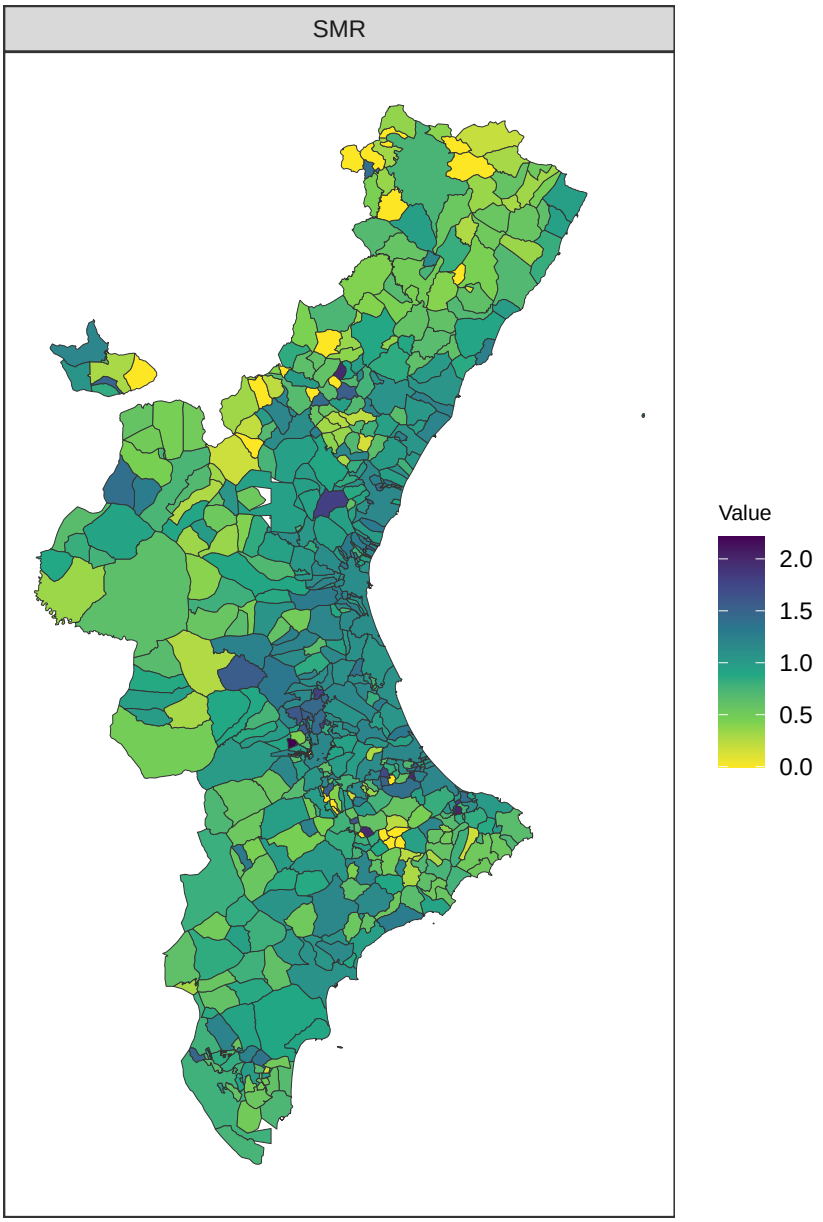

Figure 2: Standardized mortality rate of lung cancer mortality.

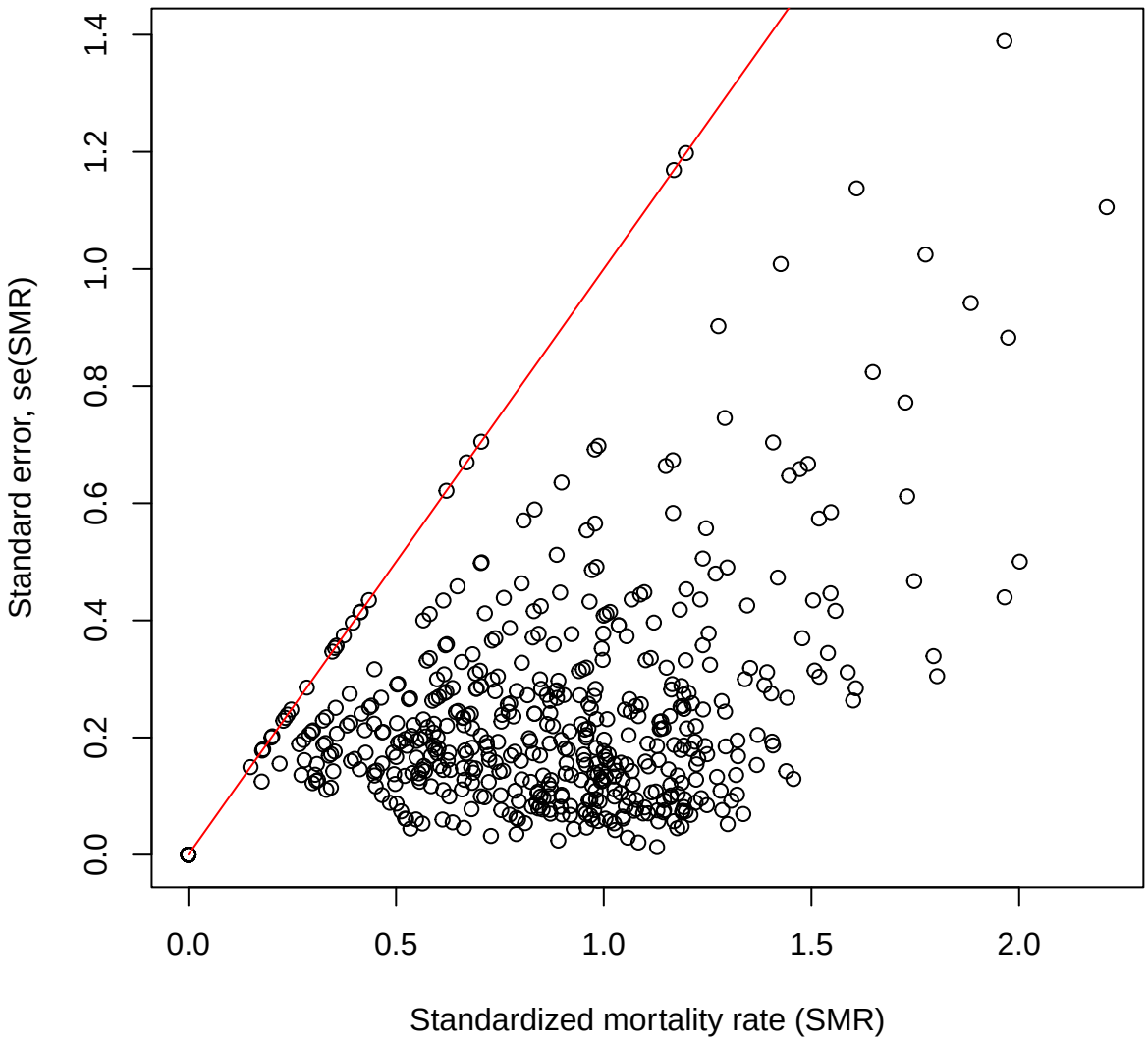

Figure 3: Standardized mortality rate versus estimate standard error.

We fit a Poisson-lognormal model of lung cancer mortality among these men, that smooths risk estimates toward a global value, $exp(\beta_0)$. We estimate β_0 to be -0.13 (95% interval: -0.17 to -0.1). A hyper-parameter that characterizes the extent of smoothing, σ_e, is estimated to be 0.25 (95% interval: 0.23-0.28).

(b) Extract the posterior medians of the relative risk (RR) estimates and provide a map of these.

See Figure 4.

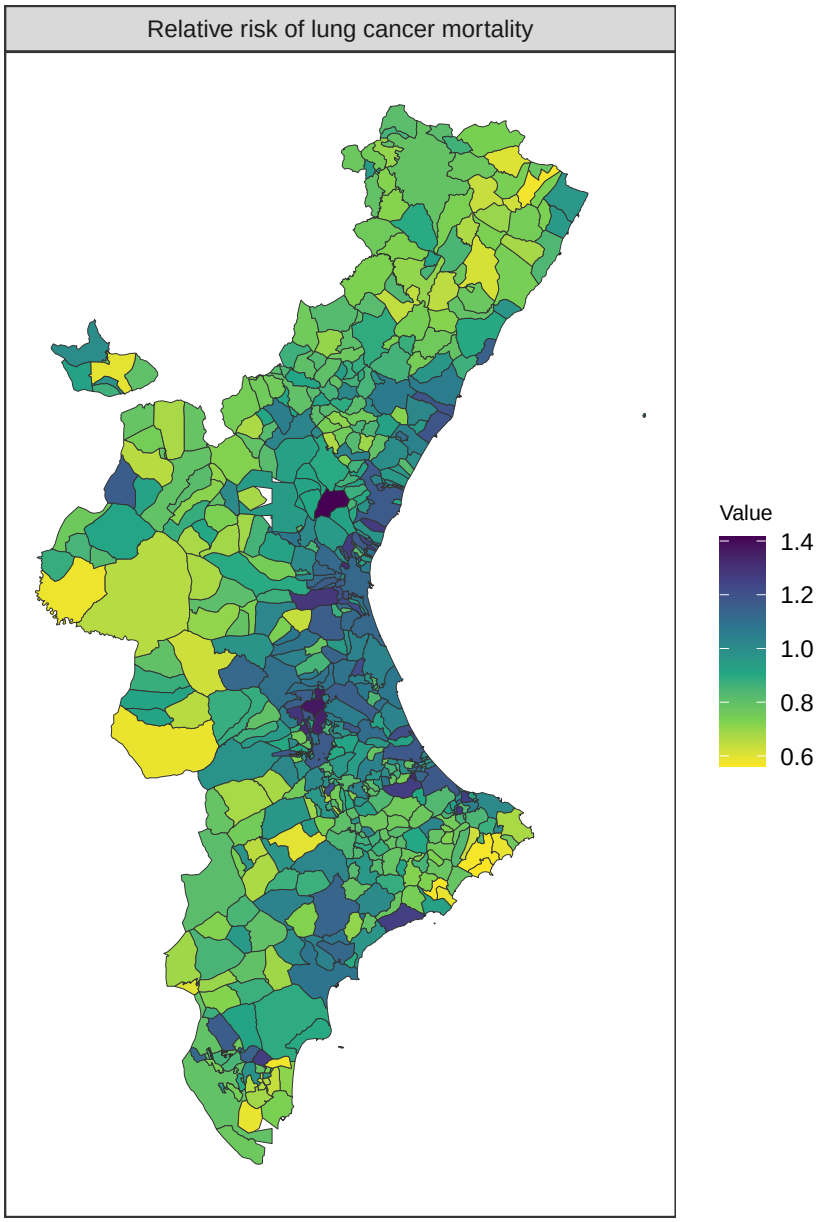

Figure 4: Posterior medians of relative risk estimates. Estimates are generated from a Poisson-Lognormal model with nonspatial smoothing.

(c) Plot these posterior RR estimates against the SMRs, and comment.

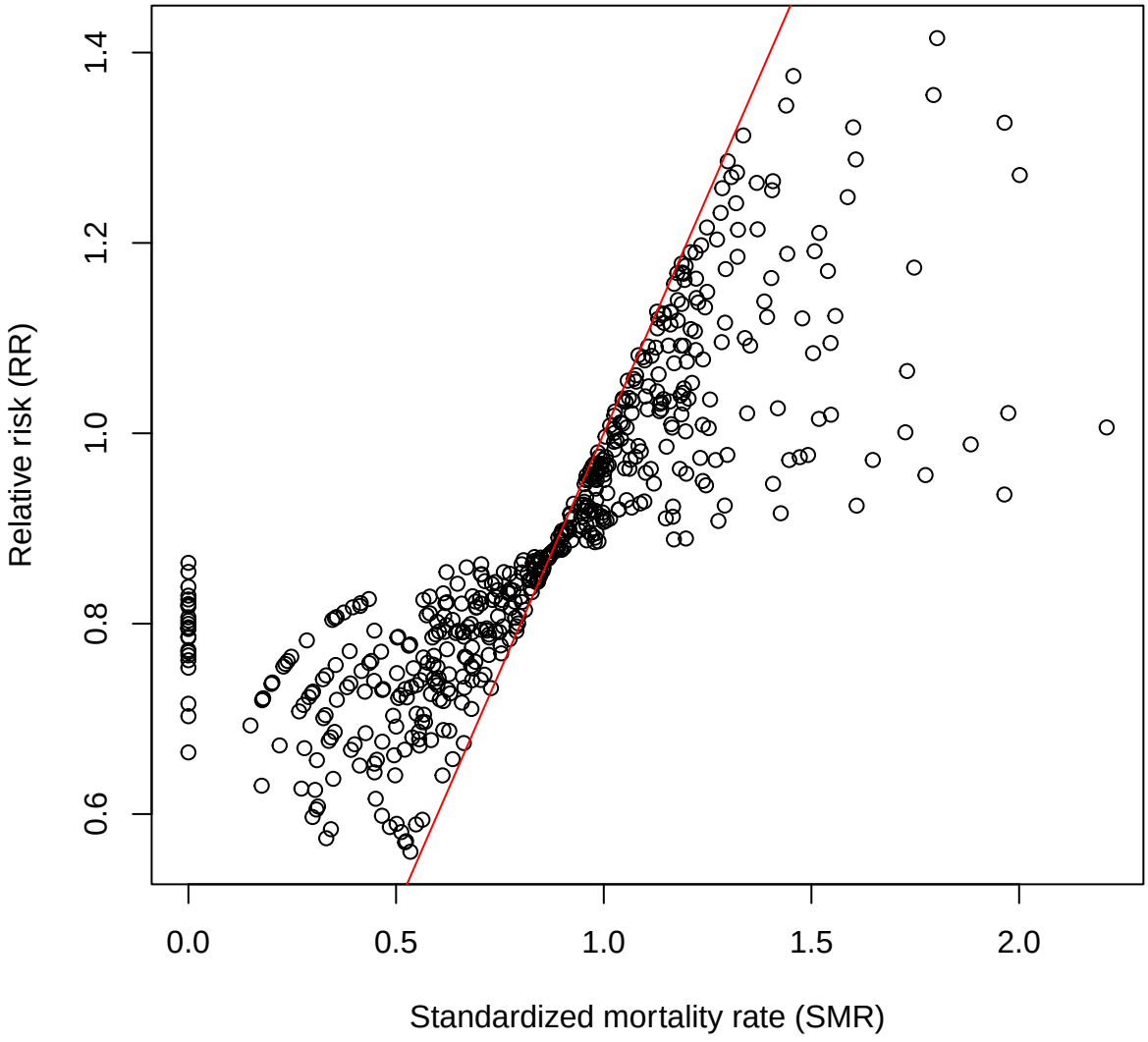

Figure 5: Standardized mortality rates versus posterior median relative risks. Relative risk estimates are generated from a Poisson-Lognormal model with nonspatial smoothing.

See Figure 5. SMR and RR estimates are generally drawn to be equivalent, both focused about $exp(\beta_0) \approx 0.88$.

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

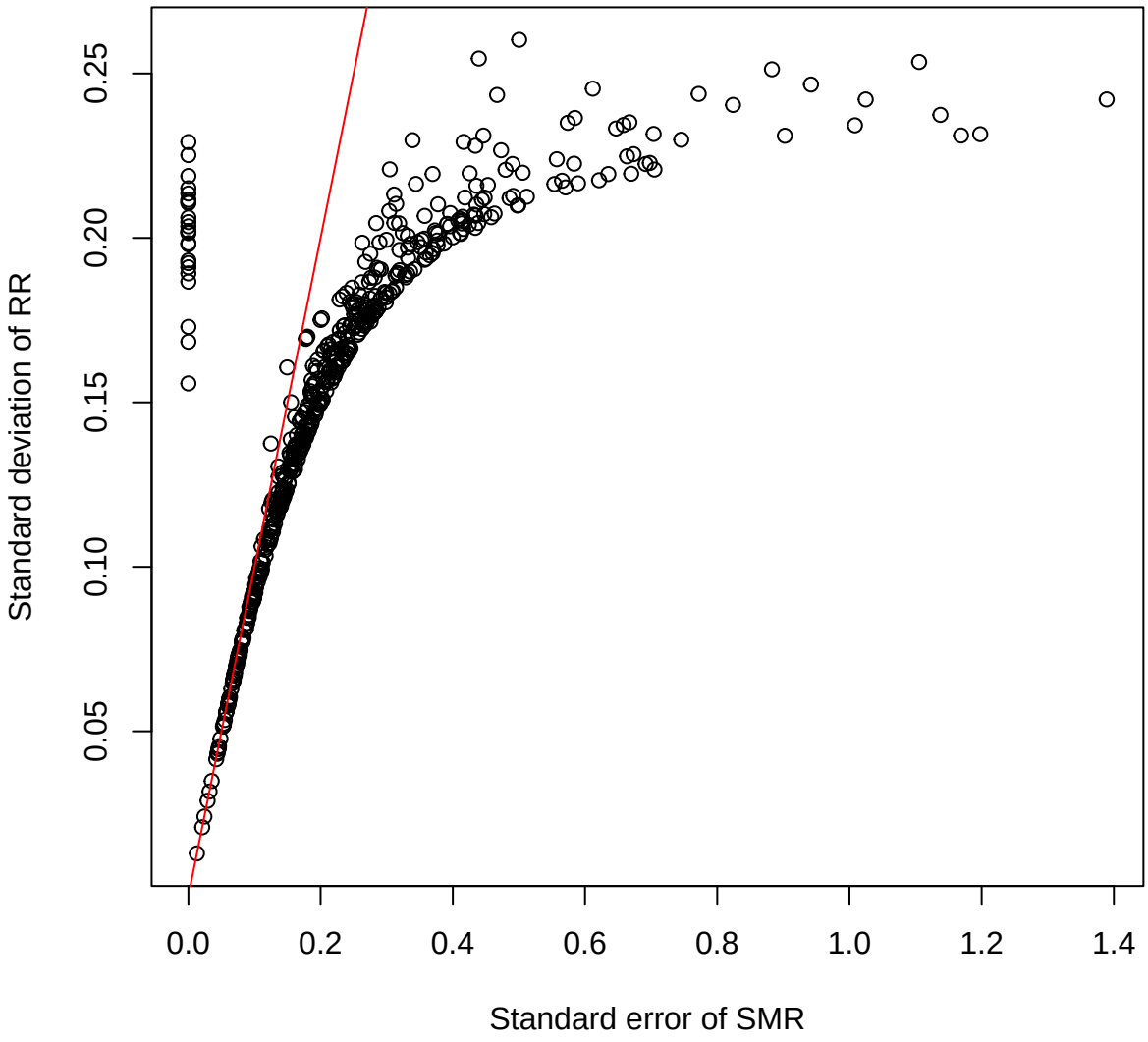

Figure 6: Standard error of standardized mortality rates versus posterior standard deviation of smoothed relative risks.

See Figure 6. Regions whose SMR have large standard errors have posterior RR standard deviations that are comparatively much smaller. This illustrates the instability of SMR estimates.

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

$Y_i|\beta_i, \epsilon_i \sim_{iid} Poisson(E_i e^{\beta_0} e^{S_i+\epsilon_i})$

$e_i|\sigma_\epsilon \sim_{iid} N(0, \sigma_\epsilon^2)$

$S_1, ..., S_n|\sigma_s^2 \sim ICAR(\sigma_s^2)$

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

(a) Using the `inla` function in R fit this model using the `bym2` model, with the default prior for β_0 and the provided prior specification for the spatial and non-spatial random effects. These 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.

We fit a Poisson-lognormal spatial model of lung cancer mortality among these men, that smooths risk estimates toward a global value, $exp(\beta_0)$, and toward values of its neighbors. We estimate β_0 to be -0.15 (95% interval: -0.18 to -0.13). We estimate the total variance of the random effects to be 1.03 (95% interval: 1.00-1.10) and the proportion of the total variance attributed to the spatial random effect to be 0.26 (95% interval: 0.22-0.29).

(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.

See Figure 7. All estimates are centered around 87%, with the SMR estimates having the largest spread. Relative risks generated with spatial smoothing have a slightly higher distribution than those generated with nonspatial smoothing. Visually, it is clear that estimates from the final model have undergone spatial smoothing.

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

Intentionally left blank.

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

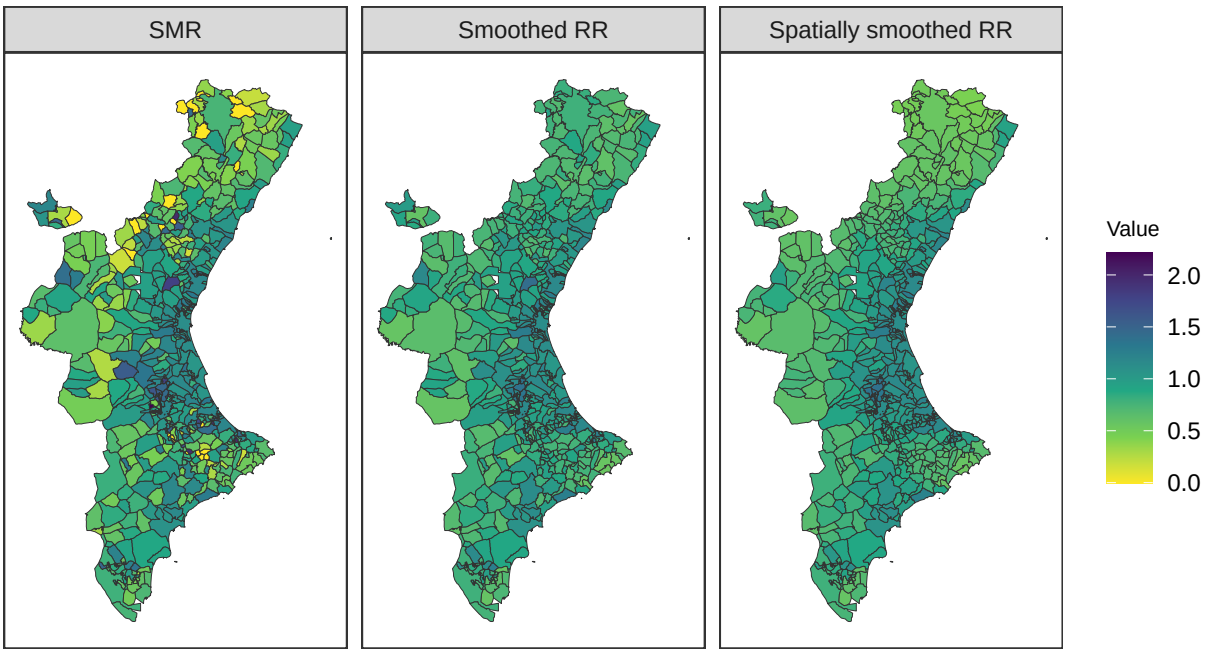

Figure 7: Estimated risk of lung cancer mortality among men near Valencia, Spain, 1991-2000. Relative risk estimates are generated from Poisson-Lognormal models.

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(SUMMER)  # spatial mapping
library(INLA)    # spatial modelling
library(dplyr)   # data manipulation

## 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
spain.ad1 <- data.frame(Region = rownames(Exp.mv3),
                  Count = data.frame(Obs.mv3)$Lung,
                  Expected = data.frame(Exp.mv3)$Lung)
# Compute SMR (standardized mortality ratio)
spain.ad1$SMR = with(spain.ad1, Count/Expected)
# Compute S.E. of SMR
spain.ad1$SMR.se = with(spain.ad1, sqrt(SMR/Expected))
head(spain.ad1)

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

# Map of observed and expected counts
SUMMER::mapPlot(data = spain.ad1, geo = geo,
                variables = c("Count","Expected"), values = c("Count","Expected"),
                labels = c("Observed", "Expected"),
                by.data = "Region", by.geo = "CODMUNI",
                legend.label = "Number of deaths", size = 0.01)

# Map of SMR
mapPlot(data = spain.ad1, geo = geo,
        variables = "SMR", values = "SMR",
        by.data = "Region", by.geo = "CODMUNI",
        size = 0.01, direction = -1)

# Plot SMR versus se(SMR)
plot(x = spain.ad1$SMR, y = spain.ad1$SMR.se,
     xlab = "Standardized mortality rate (SMR)",
     ylab = "Standard error, se(SMR)")
lines(0:10, 0:10, col = "red")
```

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

# Fit Poisson-lognormal model in INLA:
spain.fit <- INLA::inla(Count ~ 1 + f(Region, model = "iid"), data = spain.ad1,
                        family = "poisson", E = Expected)
# Get parameter estimates
beta0.est <- spain.fit$summary.fixed[4]
##              mean    sd      0.025quant  0.5quant  0.97quant
## (Intercept)  -0.13   0.016   -0.17       -0.13     -0.1
sigma.est <- sqrt(1/spain.fit$summary.hyperpar)[4]
##                    mean    sd      0.025quant  0.5quant  0.97quant
## sqrt(1/Precision)  0.25    0.73    0.28        0.25      0.23

# Extract posterior medians of relative risk (RR) estimates
spain.ad1$fit1RR <- spain.fit$summary.fitted.values[,4]

# Map median RR estimates from Poisson-lognormal smooth model
mapPlot(data = spain.ad1, geo = geo,
        variables = "fit1RR", values = "fit1RR",
        by.data = "Region", by.geo = "CODMUNI",
        size = 0.01, labels = "Relative risk of lung cancer mortality",
        direction = -1)

# Plot SMR versus RR
plot(x = spain.ad1$SMR, y = spain.ad1$fit1RR,
     xlab = "Standardized mortality rate (SMR)",
     ylab = "Relative risk (RR)")
lines(0:10, 0:10, col = "red")

# Extract posterior standard deviations of RR estimates
spain.ad1$fit1RR.sd <- spain.fit$summary.fitted.values[,2]

# Plot se(SMR) versus sd(RR)
plot(x = spain.ad1$SMR.se, y = spain.ad1$fit1RR.sd,
     xlab = "Standard error of SMR",
     ylab = "Standard deviation of RR")
lines(0:10, 0:10, col = "red")

####################
#### QUESTION 3 ####
####################

# The VR.graph file identifies regions by their index, 1:nrow(...)
# Create a new RegionIndex to match VR.graph
spain.ad1$RegionIndex <- 1:nrow(spain.ad1)

# Specify priors of BYM2 model
formula <- Count ~ 1 +
  f(RegionIndex, model="bym2", graph="../data/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)))
```

```r
# Fit Poisson-lognormal BYM2 model in INLA
spain.fit2 <- inla(formula, data=spain.ad1, family="poisson", E=Expected)

# Get parameter estimates
beta0.est <- spain.fit2$summary.fixed[4]
##              mean    sd      0.025quant  0.5quant  0.97quant
## (Intercept)  -0.15   0.013   -0.18       -0.15     -0.13
sigma2.est <- sqrt(1 / spain.fit2$summary.hyperpar)[4]
##                    mean    sd      0.025quant  0.5quant  0.97quant
## sqrt(1/Precision)  0.063   0.54    0.08        0.064     0.051

# Extract posterior medians of relative risk (RR) estimates
spain.ad1$fit2RR <- spain.fit2$summary.fitted.values[,4]

# Summarize distributions of SMR and RR's
spain.ad1[,c(4,6,9)] %>% colMeans
spain.ad1[,c(4,6,9)] %>% lapply(quantile)
spain.ad1[,c(4,6,9)] %>% lapply(IQR)

# Map median RR estimates from Poisson-Lognormal spatial and nonspatial model
mapPlot(data = spain.ad1, geo = geo,
        variables = c("SMR", "fit1RR", "fit2RR"),
        values = c("SMR", "fit1RR", "fit2RR"),
        by.data = "Region", by.geo = "CODMUNI",
        size = 0.01, direction = -1,
        labels = c("SMR", "Smoothed RR", "Spatially smoothed RR"))

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

End of document.