BIOST 555 Assignment 2

Benjamin Stan

February 2, 2022

Question 1
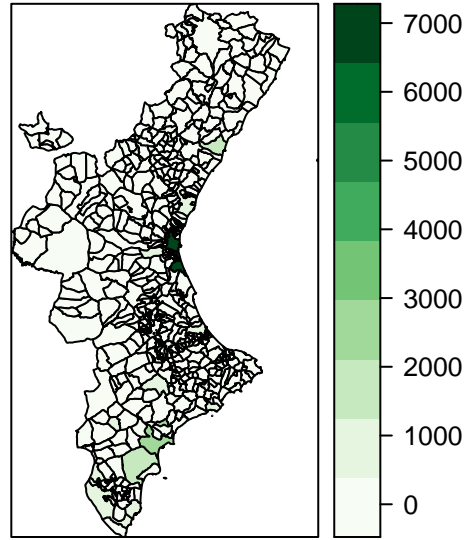
(a)

The plot of the observed counts (Y_i) is shown below:

Plot of observed counts
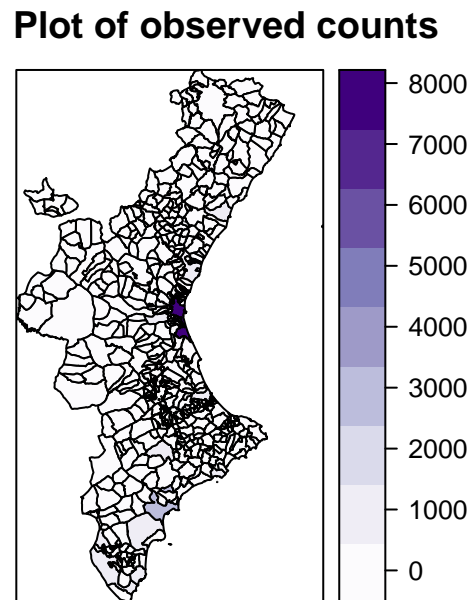
(b)

The plot of the expected counts (E_i) is shown below:
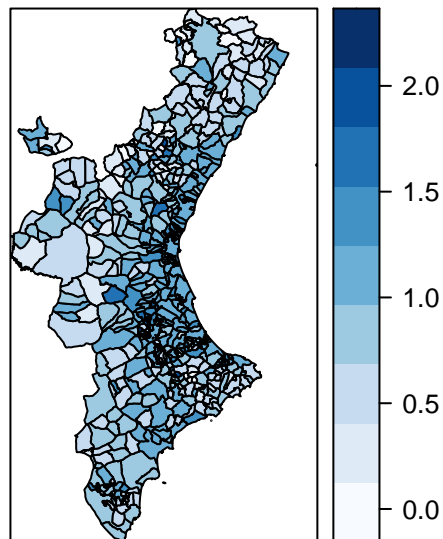
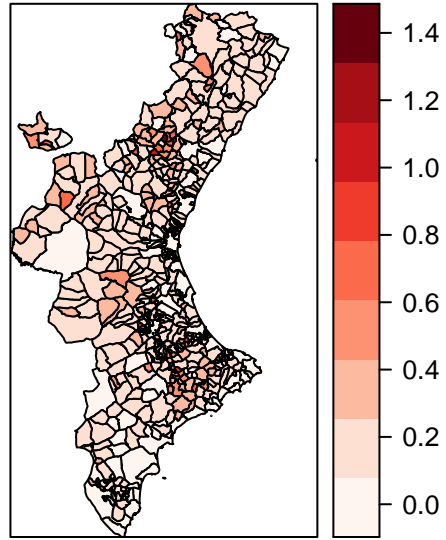# Plot of expected counts

(c)

The plot of observed SMRs is shown below:

Plot of SMRs

The computed SMRs, defined as Y_i/E_i, have a range of 0 to 2.21. The standard errors range from 0 to 1.39 (see plot below), indicating that there may be a few areas with a low expected count and high standard error relative to the SMR.

Plot of standard errors

(d)

The plot of SMR versus the estimated standard error is shown below. As might be expected, the highest standard errors are associated with higher SMR values.
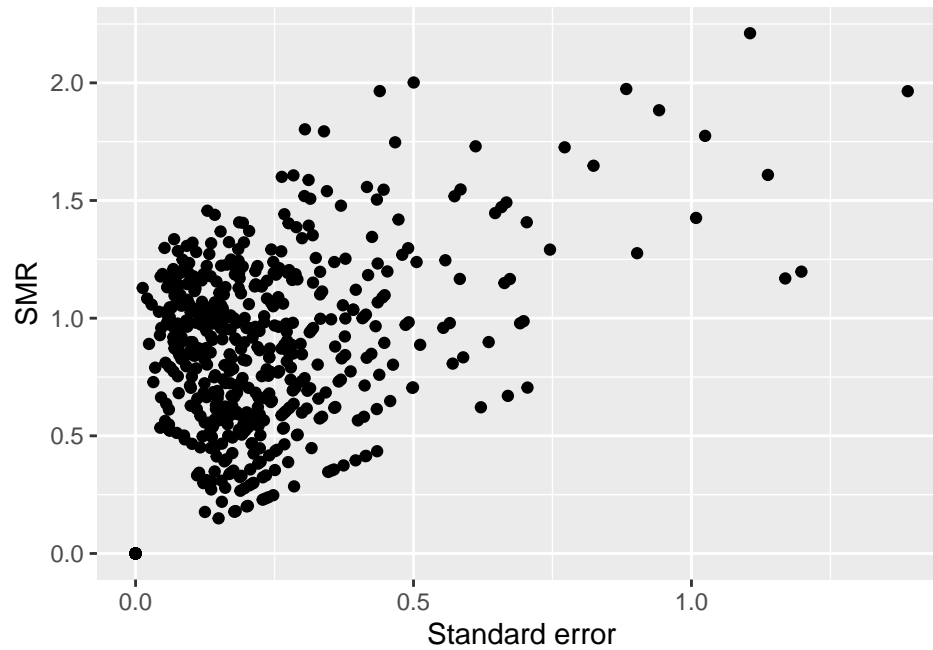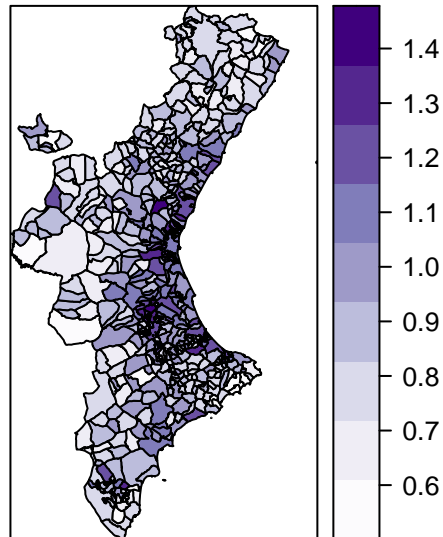
Question 2

(a)

The disease mapping Poisson-Lognormal model was fit using the `inla` function, and the estimates for the two parameters are shown below.

Parameter	Posterior Median	95% Interval Lower	95% Interval Upper
β_0	-0.133	-0.166	-0.103
σ_e	0.253	0.225	0.283

(b)

## Plot of posterior medians of relative risk

(c)

The plot of the posterior medians of the relative risk (RR) estimates versus the SMRs is shown below. Due to the shrinkage imposed by the model, the posterior median RR values are concentrated between 0.5 and 1.5 across all values of the SMR. For SMR values below 0.88, the RR median is larger. For SMR values above 0.88, the RR median is smaller.
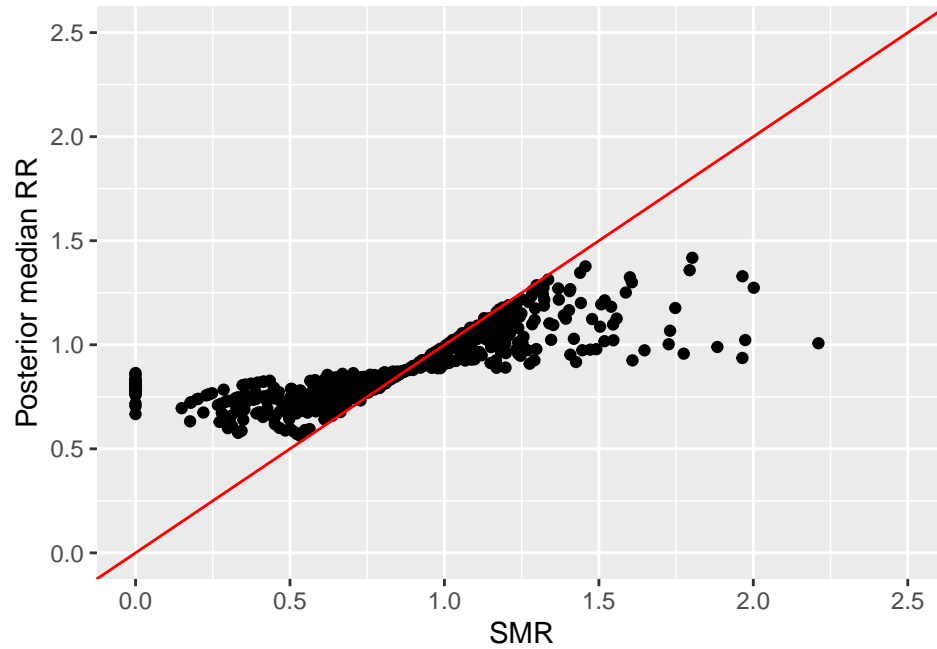
(d)

The plot of the posterior standard deviations of the RRs versus the standard errors of the SMRs is shown below. For 26 of the 540 values, the RR standard deviation was greater than the standard error of the SMR. However, for the remaining values, the RR SD was lower. As the SMR SE value increases, the difference between it and the RR SD also increases. The maximum value for SMR SE is 1.39, while the maximum value for RR SD is 0.26. The Poisson-Lognormal model reduces the variance in each estimate, and the effect is particularly seen in the areas with larger SMR standard error.
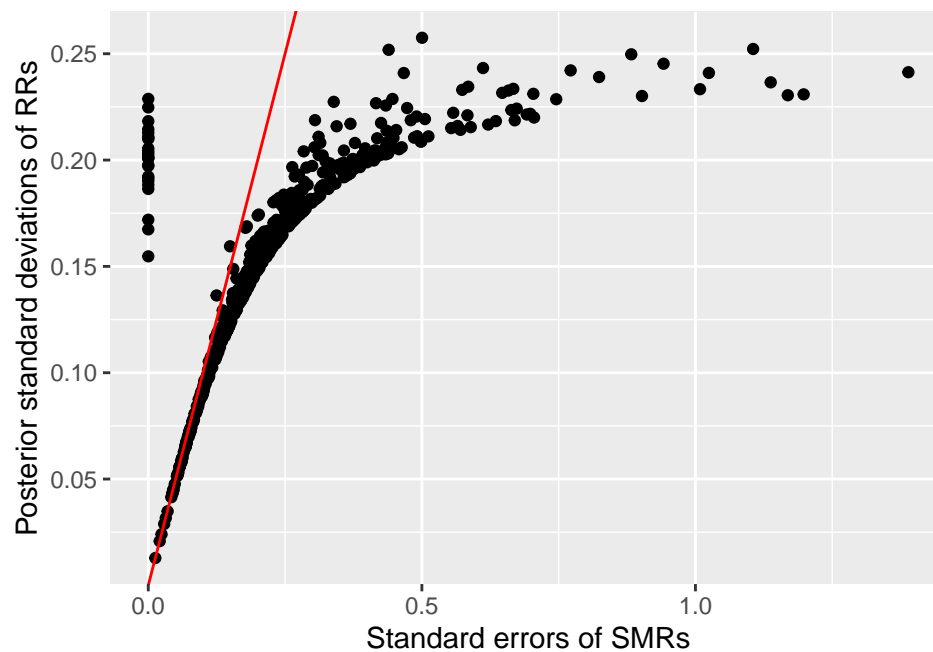
Question 3

(a)

Based on the Poisson-Lognormal-Spatial model fit, the estimated values are shown below. Note that $1/\tau_b$ corresponds to the the total variance of the random effects, and ϕ corresponds to the proportion of the total variance attributed to the spatial random effect.

Parameter	Posterior Median	95% Interval Lower	95% Interval Upper
β_0	-0.152	-0.177	-0.128
$1/\tau_b$	0.066	0.051	0.087
ϕ	0.965	0.861	0.996

(b)

The relative risk estimates using the Poison-Lognormal-Spatial model are shown below:

Plot of posterior medians of relative risk
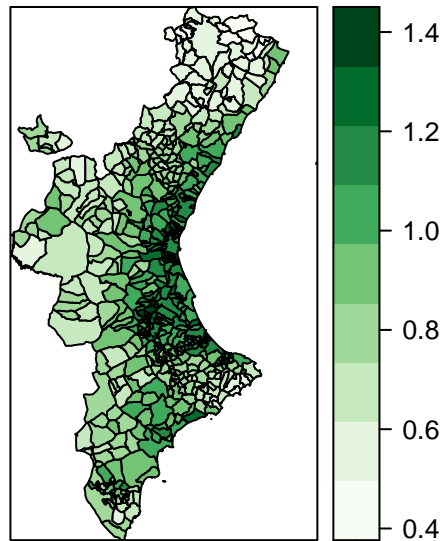
The plot of the posterior medians of the Poison-Lognormal-Spatial model versus the SMR values is shown below. Similar to the comparison of the Poison-Lognormal model to the SMR values, the model values are far more concentrated than the SMR. While the SMR values range from 0 to 2.21, the fitted values are between 0.44 and 1.38. The model values tend to be greater than the SMR for SMR values below 0.5 and less than the SMR for SMR values above 1.25.
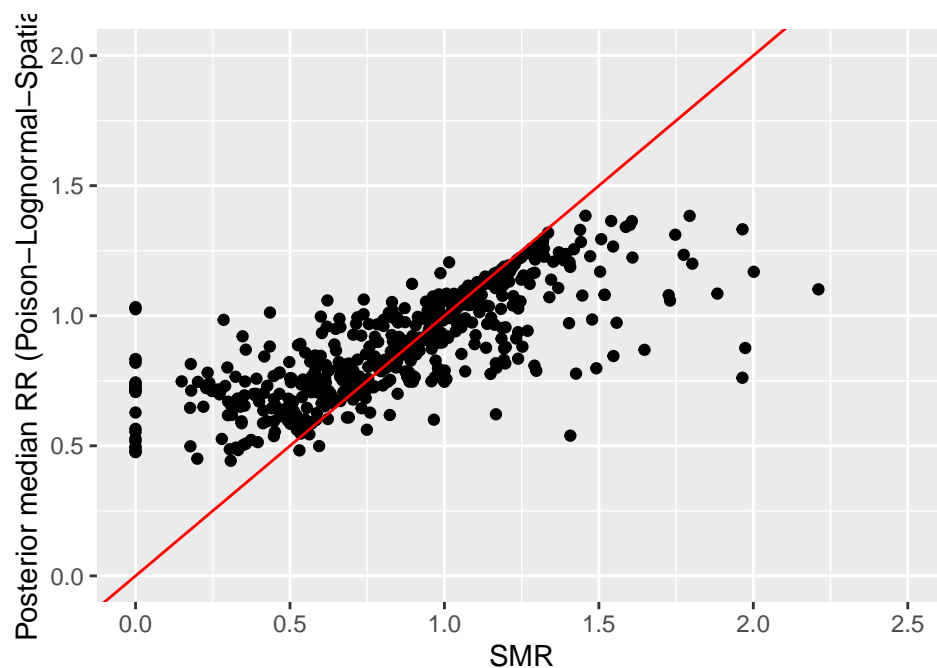
The plot of the posterior medians using the Poison-Lognormal-Spatial model (with specified prior distributions) versus the the posterior medians using the Poison-Lognormal model (with IID random effects) is shown below. The two model fits seem the show similar values that increase linearly with one another and have no consistent pattern of difference.
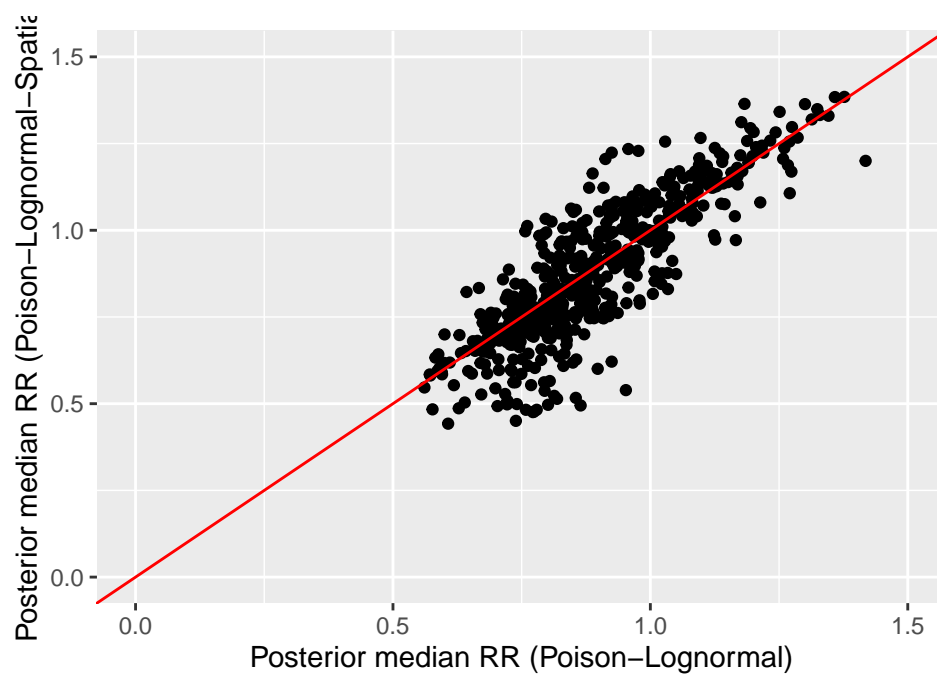

Question 4

The model to fit the log-SMRs used the following framework:

$$Z_i = log(\hat{\theta}_i) \sim N(\beta_0 + e_i, \sigma^{*2}_{\epsilon i})$$

This model was fit using `inla` with fixed error variances. The plot of estimated relative risks from this model versus the estimates from the Poisson-Lognormal model are shown below. The values appear to generally increase linearly with one another, but there is a greater difference at lower values of the Poisson-Lognormal estimate.
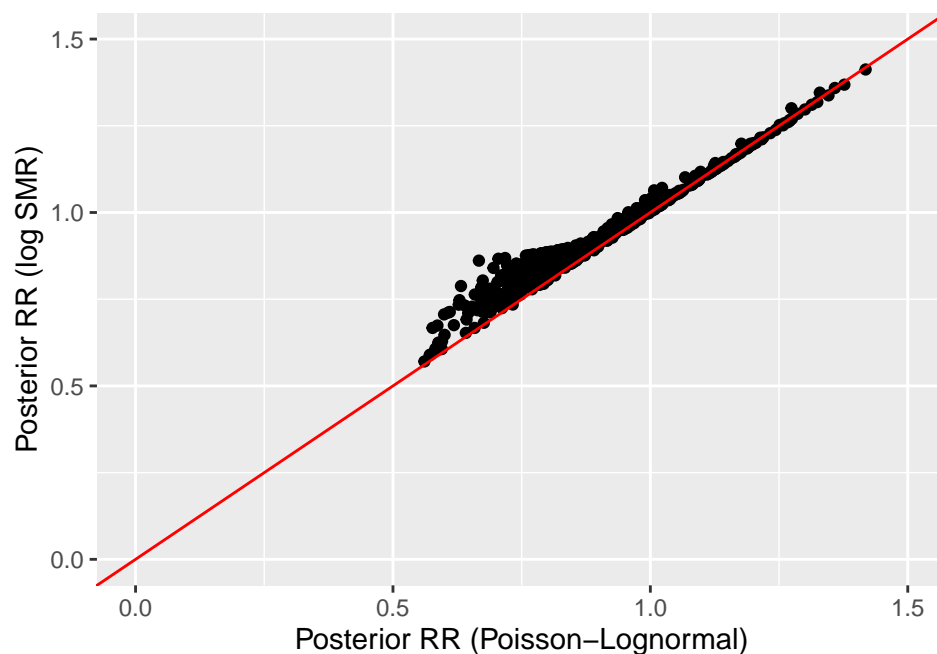
Appendix

```r
## Set working directory and load libraries
setwd("/Users/bstan/Documents/UW/Courses/BIOST 555")
rm(list = ls())
library(tidyverse)
library(tidyr)
library(tinytex)
library(knitr)
library(rgdal)
library(sp)
library(splancs)
library(spdep)
library(SpatialEpi)
library(RColorBrewer)
library(ggplot2)
library(ggridges)
library(INLA)

## Load data and create relevant data frames
load("HW2data.Rdata")
dat = data.frame(expected = Exp.mv3[,"Lung"],
```

```r
                       observed = Obs.mv3[,"Lung"])
dat$smr = dat$observed/dat$expected
dat$se = sqrt(dat$smr/dat$expected)
rownames(dat) = rownames(Exp.mv3)
dat$region = 1:nrow(dat) #as.numeric(rownames(Exp.mv3))
smap = SpatialPolygonsDataFrame(VR.cart, dat, match.ID = TRUE)

## Question 1a
spplot(smap,
       zcol = "observed",
       col.regions = brewer.pal(9, "Purples"),
       cuts = 8,
       main="Plot of observed counts")

## Question 1b
spplot(smap,
       zcol = "expected",
       col.regions = brewer.pal(9, "Greens"),
       cuts = 8,
       main="Plot of expected counts")

## Question 1c
spplot(smap,
       zcol = "smr",
       col.regions = brewer.pal(9, "Blues"),
       cuts = 8,
       main="Plot of SMRs")

spplot(smap,
       zcol = "se",
       col.regions = brewer.pal(9, "Reds"),
       cuts = 8,
       main="Plot of standard errors")

## Question 1d
ggplot(dat, aes(x = se, y = smr)) + geom_point() +
    labs(y = "SMR", x = "Standard error")

## Question 2a
fit0 <- inla(observed ~ 1 + f(region, model = "iid"),
             data = dat,
             family = "poisson",
             E = expected)
fit0$summary.fixed[, 1:5]
fit0$summary.hyper[, 1:5]
(fit0$summary.hyper[, 1:5])^(-0.5)

## Question 2b
dat$fit0fitted = fit0$summary.fitted.values$`0.5quant`
smap = SpatialPolygonsDataFrame(VR.cart, dat, match.ID = TRUE)
spplot(smap,
       zcol = "fit0fitted",
       col.regions = brewer.pal(9, "Purples"),
```

```r
        cuts = 8,
        main="Plot of posterior medians of relative risk")

## Question 2c
ggplot(data = dat, aes(y = fit0fitted, x = smr)) + #, size = se)) +
  geom_point() +
  labs(y = "Posterior median RR", x = "SMR") +
  geom_abline(intercept = 0,slope = 1, color = "red")   +
  xlim(0, 2.5) +
  ylim(0, 2.5)
# Find point that SMR and median RR are approx equal
dat %>% filter(fit0fitted > smr) %>% dplyr::select(smr) %>% max()

## Question 2d
dat$fit0sd = fit0$summary.fitted.values$sd
ggplot(data = dat, aes(y = fit0sd, x = se)) +
  geom_point() +
  geom_abline(intercept = 0,slope = 1, color = "red")   +
  labs(y = "Posterior standard deviations of RRs", x = "Standard errors of SMRs")
# Find number of RR SD values greater than SMR SE
dat %>% filter(fit0sd > se) %>% nrow()

## Question 3a
formula1 = observed ~ 1 + f(region, model="bym2", graph="VR.graph", scale.model=T, constr=T, hyper=list
fit1 = inla(formula1,
            data = dat,
            family = "poisson",
            E = expected,
            control.predictor = list(compute = TRUE),
            control.compute = list(config = TRUE))
fit1$summary.fixed[, 1:5]
fit1$summary.hyper[, 1:5]
(fit1$summary.hyper[, 1:5])^(-1)

## Question 3b
dat$fit1fitted = fit1$summary.fitted.values$`0.5quant`
smap = SpatialPolygonsDataFrame(VR.cart, dat, match.ID = TRUE)
spplot(smap,
       zcol = "fit1fitted",
       col.regions = brewer.pal(9, "Greens"),
       cuts = 8,
       main="Plot of posterior medians of relative risk")

# Compare to SMRs
ggplot(data = dat, aes(y = fit1fitted, x = smr)) +
  geom_point() +
  labs(y = "Posterior median RR (Poison-Lognormal-Spatial)",
       x = "SMR") +
  geom_abline(intercept = 0,slope = 1, color = "red")   +
  xlim(0, 2.5) +
  ylim(0, 2)

# Compare to Poison-Lognormal model
```

```r
ggplot(data = dat, aes(y = fit1fitted, x = fit0fitted)) + #, size = se)) +
  geom_point() +
  labs(y = "Posterior median RR (Poison-Lognormal-Spatial)",
       x = "Posterior median RR (Poison-Lognormal)") +
  geom_abline(intercept = 0,slope = 1, color = "red")  +
  xlim(0, 1.5) +
  ylim(0, 1.5)

## Question 4
dat$Ystar = ifelse(dat$observed == 0, dat$observed + 0.5, dat$observed)
dat$Estar = ifelse(dat$observed == 0, dat$expected + 0.5, dat$expected)
dat$thetastar = dat$Ystar/dat$Estar
dat$Z = log(dat$thetastar)
dat$varZ = 1/(dat$Estar * dat$thetastar)
dat$precZ = 1/dat$varZ

fit2 = inla(Z ~ 1 + f(region, model = "iid"),
            data = dat, family = "gaussian",
            control.predictor = list(compute = TRUE),
            control.family = list(hyper = list(prec = list(initial = log(1), fixed = TRUE))),
            scale = precZ)
dat$fit2fitted = exp(fit2$summary.fitted.values$`0.5quant`)
ggplot(data = dat, aes(y = fit2fitted, x = fit0fitted)) + #, size = se)) +
  geom_point() +
  labs(y = "Posterior RR (log SMR)", x = "Posterior RR (Poisson-Lognormal)") +
  geom_abline(intercept = 0,slope = 1, color = "red")  +
  xlim(0, 1.5) +
  ylim(0, 1.5)
```