

Week 9: Hierarchical GLM

18/03/23

Lip cancer

Question 1

Explain a bit more what the `expect.i` variable is. For example, if a particular area has an expected deaths of 6, what does this mean?

Answer

Expected deaths is the implied number of lip cancer deaths for a particular region given that region's age structure and the national level age-specific mortality rates for lip cancer. For example, an expected number of deaths of 6 would mean that for that particular region, we would expect 6 lip cancer deaths if this region were to experience the same age specific mortality rates as at the national level.

Question 2

Run three different models in Stan with three different set-up's for estimating θ_i , that is the relative risk of lip cancer in each region:

1. Intercept α_i is same in each region $= \alpha$ (with covariate)
2. α_i is different in each region and modeled separately (with covariate)
3. α_i is different in each region and the intercept is modeled hierarchically (with covariate)

Answer

$$y_i | \theta_i \sim \text{Poisson}(\theta_i \cdot e_i)$$

Look at three models for $\log \theta_i$: Model 1:

$$\log \theta_i = \alpha + \beta x_i$$

Model 2:

$$\log \theta_i = \alpha_i + \beta x_i$$

Model 3:

$$\log \theta_i = \alpha_i + \beta x_i$$

with

$$\alpha_i \sim N(\mu, \sigma^2)$$

Model 1

```
library(here)
library(tidyverse)
library(rstan)
library(tidybayes)

stan_data=list(y=observe.i,
              log_e=log(expect.i),
              N=length(observe.i),
              x=aff.i-mean(aff.i))
mod1=stan(data=stan_data,file=here("lab9_1.stan"))
```

Model 2

```
mod2=stan(data=stan_data,file=here("lab9_2.stan"))
mod2
```

Model 3

```
mod3=stan(data=stan_data,file=here("lab9_3.stan"))
mod3
```

Question 3

Make two plots (appropriately labeled and described) that illustrate the differences in estimated θ_i 's across regions and the differences in θ s across models.

Answer

The first plot shows a histogram of log theta estimates for each model. The second plot shows the log estimates for each region with error bars, with colour indicating each model for the estimates and the size of the points indicating the number of observed deaths in each region.

```
res_mod1=mod1 %>%
  gather_draws(log_theta[i]) %>%
  median_qi() %>%
  rename(median_mod1=.value,
         lower_mod1=.lower,
         upper_mod1=.upper) %>%
  select(i,median_mod1:upper_mod1)

res_mod2=mod2 %>%
  gather_draws(log_theta[i]) %>%
  median_qi() %>%
  rename(median_mod2=.value,
         lower_mod2=.lower,
         upper_mod2=.upper) %>%
  select(i,median_mod2:upper_mod2)

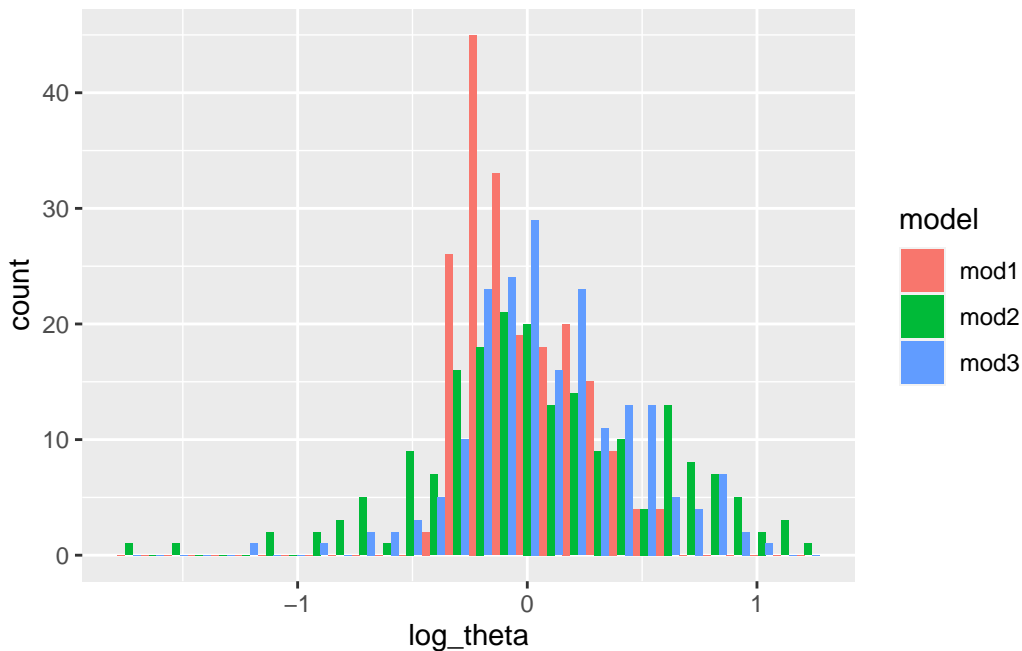
res_mod3=mod3 %>%
  gather_draws(log_theta[i]) %>%
  median_qi() %>%
  rename(median_mod3=.value,
         lower_mod3=.lower,
         upper_mod3=.upper) %>%
  select(i,median_mod3:upper_mod3)

res=res_mod1 %>%
  left_join(res_mod2) %>%
  left_join(res_mod3)
```

```

res %>%
  select(median_mod1,
         median_mod2,
         median_mod3) %>%
  pivot_longer(median_mod1:median_mod3,names_to="model",values_to="log_theta") %>%
  mutate(model=str_remove(model,"median_")) %>%
  ggplot(aes(log_theta,fill=model))+
  geom_histogram(position="dodge")

```



```

res %>%
  mutate(deaths=observe.i) %>%
  mutate(log_smr=log(observe.i/expect.i)) %>%
  ggplot(aes(log_smr,median_mod1,color="Model 1"))+
  geom_point(aes(size=deaths),alpha=0.6)+
  geom_errorbar(aes(ymin=lower_mod1,ymax=upper_mod1),alpha=0.6)+
  geom_abline(slope=1,intercept=0)+
  geom_point(aes(log_smr,median_mod2,color="Model 2"),alpha=0.6)+
  geom_errorbar(aes(ymin=lower_mod2,ymax=upper_mod2,color="Model 2"),alpha=0.6)+
  geom_point(aes(log_smr,median_mod3,color="Model 3"),alpha=0.6)+
  geom_errorbar(aes(ymin=lower_mod3,ymax=upper_mod3,color="Model 3"),alpha=0.6)

```

