`rstanarm` - Exercise 2

Bayesian Inference - Lab Sessions

Marika D'Agostini
`marika.dagostini2@unibo.it`

University of Bologna

November-December 2023

Exercise 2: Radon data (I)

- $n = 919$ observations of the indoor radon concentration in the state of Minnesota
- A normal model is specified for the response variable that is the *logarithmic transformation of the radon measurement*
- As auxiliary information the following covariates are provided:
  - *log of uranium*: transformation of the uranium concentration (county-level information).
  - *floor*: dichotomous variable (measurement at basement level or first floor)
  - *county*: name of the county

$\rightarrow$ A **random effect** might be included in the model to account for the county of the measurement: a **Bayesian hierarchical model** is estimated.

Exercise 2: Radon data (II)

The total number of observations $n = 919$ are partitioned with respect to the county: in each one of the $j = 1, ..., 85$ counties there are $i = 1, ..., n_j$ measurements.

This notation allow us to consider different Bayesian hierarchical models:

a) **Random intercept model**
b) **Model with random intercepts and covariates**
c) **Model with random intercepts, covariates and random slopes**

a) Random intercept model

Likelihood:

$$y_{ij}|\mu_{ij}, \sigma^2 \sim \mathcal{N}(\mu_{ij}, \sigma^2);$$
$$\mu_{ij}|\beta_{[.]} = \beta_{0[j]}; \quad j = 1, ..., 85; \quad i = 1, .., n_j.$$

Priors:

$$\sigma \sim \pi(\sigma)$$
$$\beta_{0[j]}|\beta_0, \sigma^2_{\beta_0} \sim \mathcal{N}(\beta_0, \sigma^2_{\beta_0});$$

Hyperprior:

$$\sigma_{\beta_0} \sim \pi(\sigma_{\beta_0}).$$

```r
data2 <- read.csv("Data_Ex_2.csv")

mod_ex2a <- stan_lmer(log_radon~(1|county),
                                data = data2)
prior_summary(mod_ex2a)
```

```
Priors for model 'mod_ex2a'
------
Intercept (after predictors centered)
  Specified prior:
    ~ normal(location = 1.3, scale = 2.5)
  Adjusted prior:
    ~ normal(location = 1.3, scale = 2)

Auxiliary (sigma)
  Specified prior:
    ~ exponential(rate = 1)
  Adjusted prior:
    ~ exponential(rate = 1.2)

Covariance
 ~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)
------
```

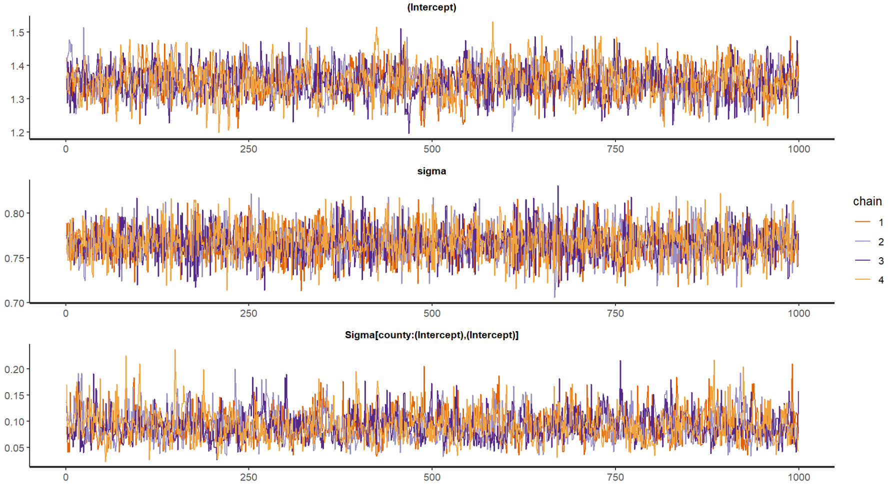
```
summary(mod_ex2a)
```

```
MCMC diagnostics
                                      mcse Rhat n_eff
(Intercept)                           0.0  1.0  1648
b[(Intercept) county:AITKIN]          0.0  1.0  4958
b[(Intercept) county:ANOKA]           0.0  1.0  3777
b[(Intercept) county:BECKER]          0.0  1.0  5373
```

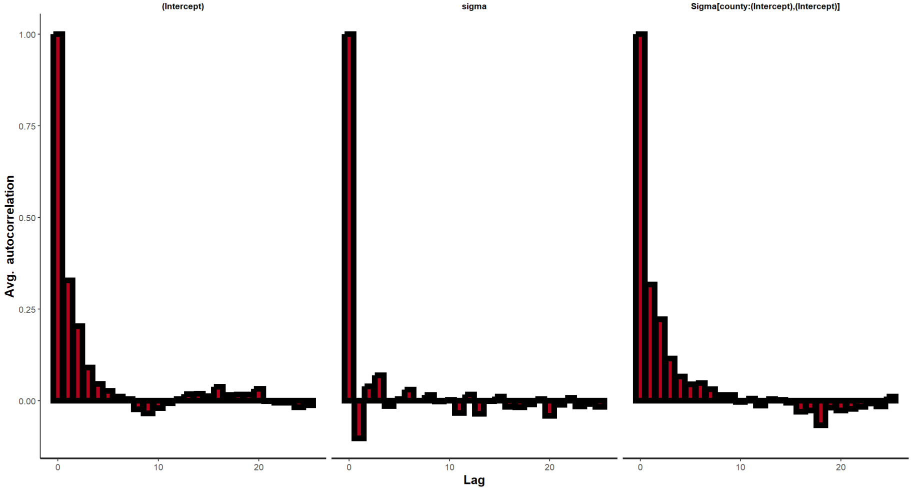
[...]

```
b[(Intercept) county:WASECA]          0.0  1.0  4266
b[(Intercept) county:WASHINGTON]      0.0  1.0  4219
b[(Intercept) county:WATONWAN]        0.0  1.0  4693
b[(Intercept) county:WILKIN]          0.0  1.0  5409
b[(Intercept) county:WINONA]          0.0  1.0  5353
b[(Intercept) county:WRIGHT]          0.0  1.0  5176
b[(Intercept) county:YELLOWMEDICINE]  0.0  1.0  5333
sigma                                 0.0  1.0  3927
Sigma[county:(Intercept),(Intercept)] 0.0  1.0  1454
mean_PPD                              0.0  1.0  4806
```

```r
stan_trace(mod_ex2a, pars = c("(Intercept)",
                  "sigma",
                  "Sigma[county:(Intercept),(Intercept)]"),
                  nrow = 3, ncol = 1)
```



```r
stan_ac(mod_ex2a, pars = c("(Intercept)",
                           "sigma",
                           "Sigma[county:(Intercept),(Intercept)]")
```



b) Model with random intercepts and covariates

Likelihood:

$$y_{ij}|\mu_{ij},\sigma^2 \sim \mathcal{N}(\mu_{ij},\sigma^2);$$
$$\mu_{ij}|\boldsymbol{\beta} = \beta_{0[j]} + \beta_1 \texttt{log\_uranium}_{ij} + \beta_2 \texttt{floor}_{ij};\ j = 1,...,85;\ i = 1,..,n_j.$$

Priors:

$$\begin{aligned} &\sigma \sim \pi(\sigma) \\ &\beta_{0[j]}|\beta_0,\sigma^2_{\beta_0} \sim \mathcal{N}(\beta_0,\sigma^2_{\beta_0}); \\ &\beta_k \sim \mathcal{N}(0,c) \qquad k = 1,2 \end{aligned}$$

Hyperprior:

$$\sigma_{\beta_0} \sim \pi(\sigma_{\beta_0}).$$

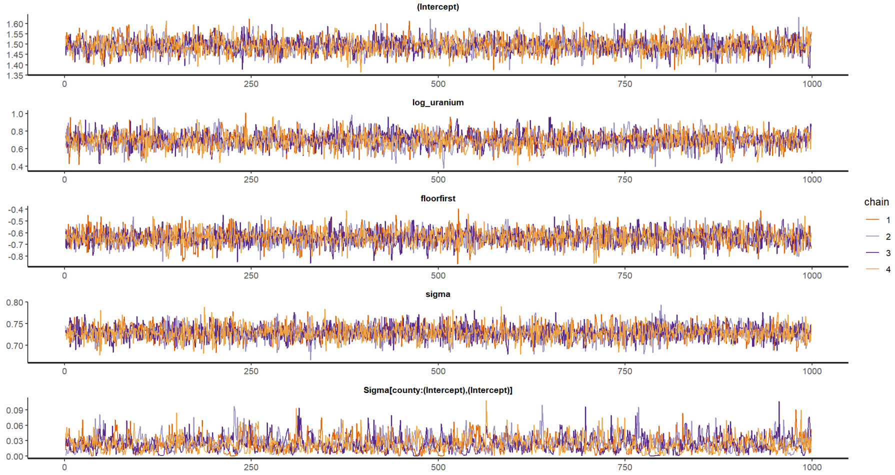
```
mod_ex2b <- stan_lmer(log_radon ~ log_uranium +
                      floor + (1|county),
                      data = data2)
summary(mod_ex2b)
```

```
MCMC diagnostics
                              mcse Rhat n_eff
(Intercept)                   0.0  1.0  2931
log_uranium                   0.0  1.0  2839
floorfirst                    0.0  1.0  4762
b[(Intercept) county:AITKIN]  0.0  1.0  4797
b[(Intercept) county:ANOKA]   0.0  1.0  3442
b[(Intercept) county:BECKER]  0.0  1.0  4421
```

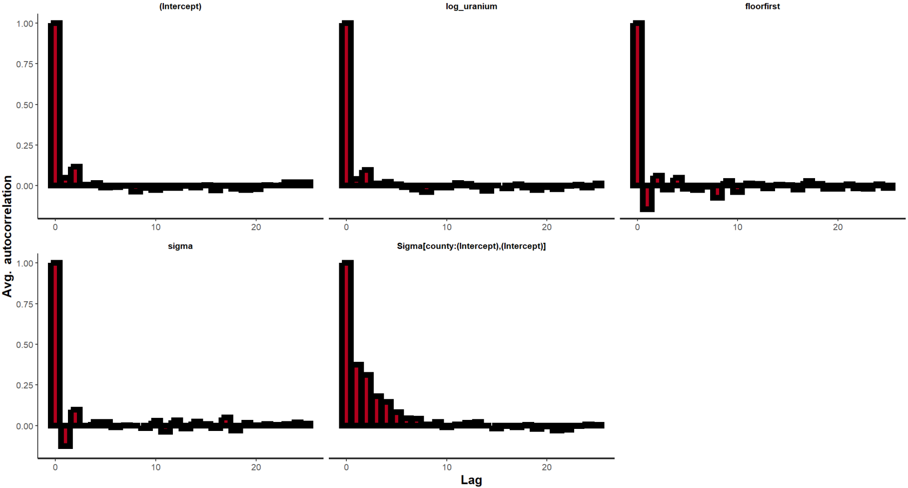
[...]

```
b[(Intercept) county:WINONA]            0.0  1.0  4581
b[(Intercept) county:WRIGHT]            0.0  1.0  3839
b[(Intercept) county:YELLOWMEDICINE]    0.0  1.0  4018
sigma                                   0.0  1.0  3814
Sigma[county:(Intercept),(Intercept)]   0.0  1.0  1107
mean_PPD                                0.0  1.0  4443
```

```r
stan_trace(mod_ex2b, pars = c("(Intercept)",
                    "sigma", "log_uranium", "floorfirst",
                    "Sigma[county:(Intercept),(Intercept)]"),
                    nrow = 5, ncol = 1)
```



```
stan_acf(mod_ex2b, pars = c("(Intercept)",
                   "sigma", "log_uranium", "floorfirst",
                   "Sigma[county:(Intercept),(Intercept)]"))
```



c) Model random intercepts, covariates and random slopes

For example, let's assume a random effect for the variable `floor`

Likelihood:

$$y_{ij}|\mu_{ij},\sigma^2 \sim \mathcal{N}(\mu_{ij},\sigma^2);$$
$$\mu_{ij}|\boldsymbol{\beta} = \beta_{0[j]} + \beta_1 \texttt{log\_uranium}_{ij} + \beta_{2[j]}\texttt{floor}_{ij};\ j = 1, ..., 85;\ i = 1, .., n_j.$$

Priors:

$$\begin{aligned}
&\sigma \sim \pi(\sigma)\\
&\beta_{0[j]}|\beta_0,\sigma^2_{\beta_0} \sim \mathcal{N}(\beta_0,\sigma^2_{\beta_0});\\
&\beta_1 \sim \mathcal{N}(0,c)\\
&\beta_{2[j]}|\beta_2,\sigma^2_{\beta_2} \sim \mathcal{N}\left(\beta_2,\sigma^2_{\beta_2}\right)
\end{aligned}$$

Hyperprior: (prior for a matrix of random effect variances)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{\beta_0} & \\ \sigma_{\beta_0,\beta_2} & \sigma^2_{\beta_2} \end{bmatrix} \sim \pi(\boldsymbol{\Sigma})$$

```
mod_ex2c<-stan_lmer(log_radon ~ log_uranium + floor +
                (1+floor|county), data = data2)
summary(mod_ex2c)
```

```
MCMC diagnostics
                                       mcse Rhat n_eff
(Intercept)                            0.0  1.0   4525
log_uranium                            0.0  1.0   3799
floorfirst                             0.0  1.0   4913
b[(Intercept) county:AITKIN]           0.0  1.0   7304
```

...

```
sigma                                  0.0  1.0   4692
Sigma[county:(Intercept),(Intercept)] 0.0  1.0   1798
Sigma[county:floorfirst,(Intercept)]  0.0  1.0   1316
Sigma[county:floorfirst,floorfirst]   0.0  1.0   1080
mean_PPD                               0.0  1.0   4489
```

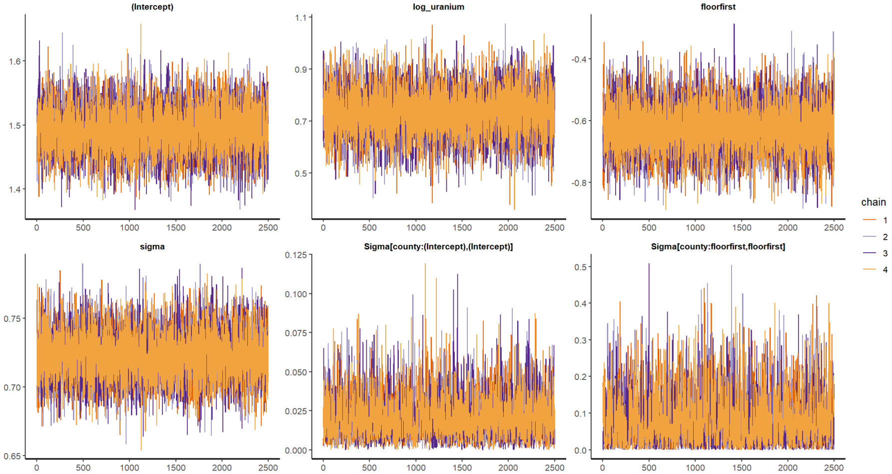
```
mod_ex2c <- update(mod_ex2c, iter=5000)

summary(mod_ex2c)
```

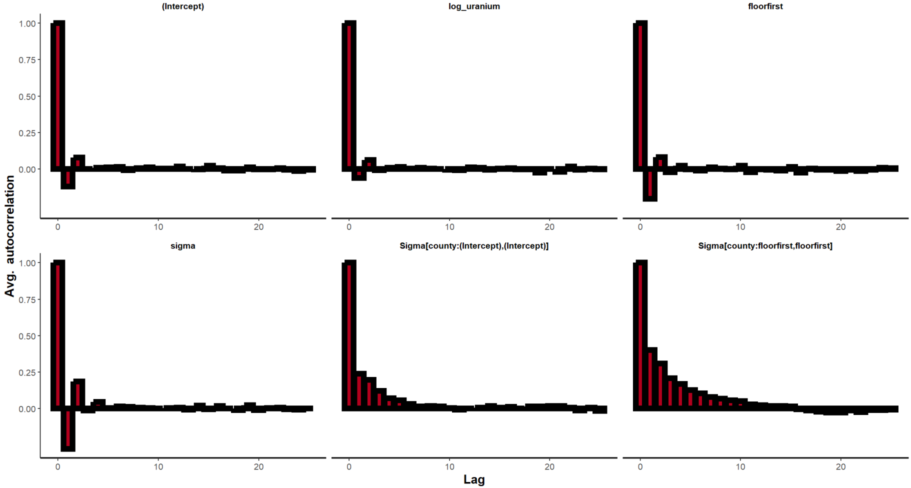
...

```
b[floorfirst county:YELLOWMEDICINE]   0.0  1.0  18462
sigma                                 0.0  1.0  10133
Sigma[county:(Intercept),(Intercept)] 0.0  1.0   3935
Sigma[county:floorfirst,(Intercept)]  0.0  1.0   3345
Sigma[county:floorfirst,floorfirst]   0.0  1.0   2280
mean_PPD                              0.0  1.0  10882
```

```
stan_trace(mod_ex2c, pars = c("(Intercept)",
        "sigma", "log_uranium", "floorfirst",
        "Sigma[county:(Intercept),(Intercept)]",
        "Sigma[county:floorfirst,floorfirst]"))
```



```
stan_acf(mod_ex2c, pars = c("(Intercept)",
        "sigma", "log_uranium", "floorfirst",
        "Sigma[county:(Intercept),(Intercept)]",
        "Sigma[county:floorfirst,floorfirst]"))
```



Model Choice

```
waic(mod_ex2a)
waic(mod_ex2b)
waic(mod_ex2c)
```

WAIC		
Model	Estimate	SE
mod_ex2a	2166.7	57.5
mod_ex2b	2054.7	57.7
mod_ex2c	2057.1	59.2

Summary of the better model

```
main_pars <- c("(Intercept)", "log_uranium", "floorfirst
       "Sigma[county:(Intercept),(Intercept)]")
summary(mod_ex2b, pars = main_pars, digits = 3)
```

```
Model Info:
 function:     stan_lmer
 family:       gaussian [identity]
 formula:      log_radon ~ log_uranium + floor + (1 | county)
 algorithm:    sampling
 sample:       4000 (posterior sample size)
 priors:       see help('prior_summary')
 observations: 919
 groups:       county (85)

Estimates:
                                        mean   sd     10%    50%    90%
(Intercept)                            1.494  0.037  1.448  1.494  1.541
log_uranium                            0.701  0.087  0.588  0.702  0.809
floorfirst                            -0.638  0.067 -0.722 -0.638 -0.554
sigma                                  0.729  0.018  0.706  0.729  0.752
Sigma[county:(Intercept),(Intercept)]  0.023  0.014  0.007  0.021  0.042

MCMC diagnostics
                                      mcse  Rhat  n_eff
(Intercept)                           0.001 1.001 2658
log_uranium                           0.002 1.003 2776
floorfirst                            0.001 1.000 5392
sigma                                 0.000 1.000 3440
Sigma[county:(Intercept),(Intercept)] 0.000 1.002 1187
```

Posterior Intervals

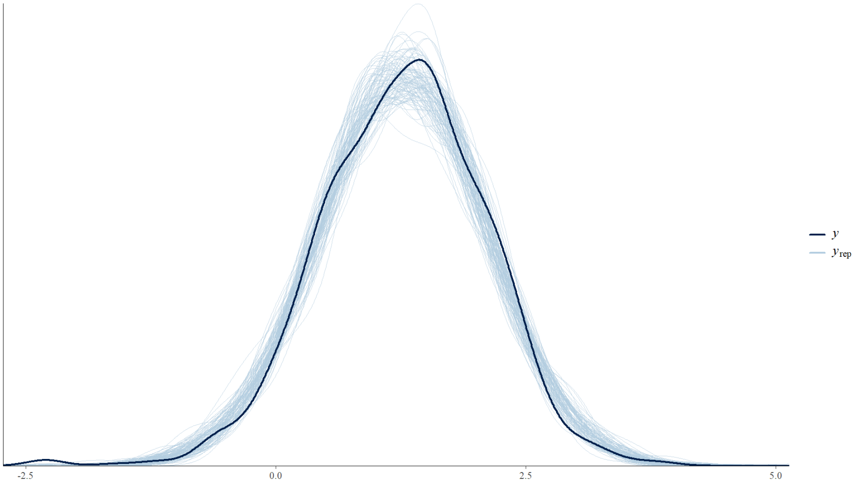
```
posterior_interval(mod_ex2b, prob = 0.8,
                   pars = "log_uranium")

posterior_interval(mod_ex2b, prob = 0.9,
                   pars = "log_uranium")
```

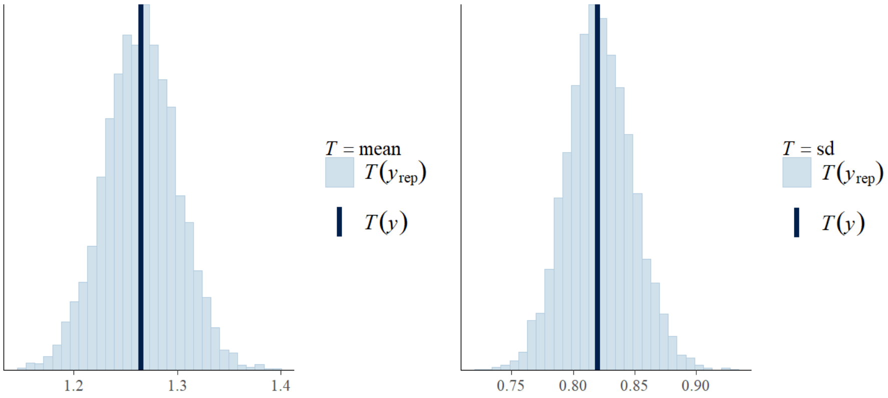
```
> posterior_interval(mod_ex2b, prob = 0.8, pars = "log_uranium")
                 10%       90%
log_uranium 0.587984 0.8085578
> posterior_interval(mod_ex2b, prob = 0.9, pars = "log_uranium")
                   5%       95%
log_uranium 0.5592441 0.8413743
```

Posterior predictive

```
y_tilde <- posterior_predict(mod_ex2b)
ppc_dens_overlay(y = data2$log_radon,
                 yrep = y_tilde[1100:1200,])
```



```
ppc_stat(y = data2$log_radon, yrep = y_tilde,
                   stat = "mean")
ppc_stat(y = data2$log_radon, yrep = y_tilde,
                   stat = "sd")
```



```
ppc_stat(y = data2$log_radon, yrep = y_tilde,
                stat = "max")
ppc_stat(y = data2$log_radon, yrep = y_tilde,
                stat = "min")
```