`rstanarm` - Exercise 3

Bayesian Inference - Lab Sessions

Marika D'Agostini
marika.dagostini2@unibo.it

University of Bologna

November-December 2023

Exercise 3: Logistic Regression (I)

If the interest is in modelling a dichotomous variable, the **Logistic Regression Model** is the most common choice.

It is a GLM with the *Bernoulli (or binomial) distribution* assumed for data and the linear predictor (function of the covariate pattern $\mathbf{x}_i$) is specified for a suitable transformation of the probability.

In particular, the *logit function* is used:

$$y_i | p_i \sim \mathbf{Ber}(p_i),$$
$$\log\left(\frac{p_i}{1-p_i}\right) | \boldsymbol{\beta} = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, ..., n.$$

Exercise 3: Logistic Regression (II)

- Data from a survey about the vote during the 2000 US Presidential elections
- The response variable = 1 if the subject voted for Bush, 0 otherwise.
- As auxiliary information the gender (1=female, 0=male), the race (1=black, 0=other) and the state are included in the study.

The objective of this exercise is to fit two models with different linear predictions:

- **(a) Simple Logistic Regression Model**:

$$\log\left(\frac{p_i}{1-p_i}\right) |\boldsymbol{\beta} = \beta_0 + \beta_1 \texttt{race}_i + \beta_2 \texttt{gender}_i, \quad i = 1, .., n$$

And considering the j=1,...,49 states:

- **(b) Logistic Regression Model with random intercept**:

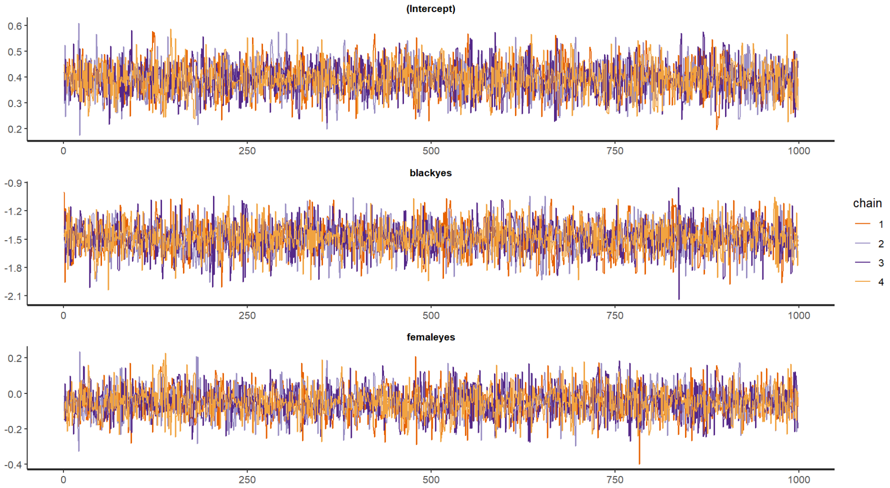
$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) |\boldsymbol{\beta} = \beta_{0[j]} + \beta_1 \texttt{race}_{ij} + \beta_2 \texttt{gender}_{ij}, \quad j = 1, ..., 49, \; i = 1, ..., n_j.$$

a) Simple Logistic Regression Model

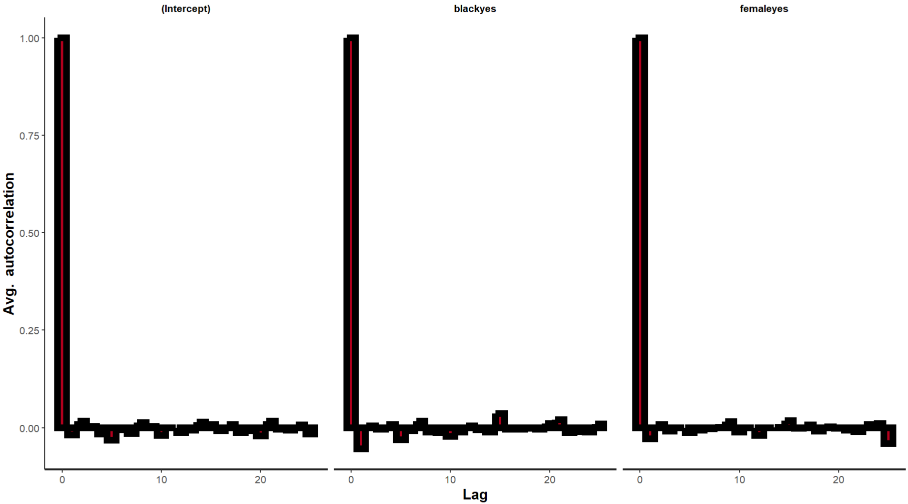
```
data3 <- read.csv("Data_Ex_3.csv")

mod_ex3a <- stan_glm(bush~black+female,
                     data = data3,
                     family = "binomial",
                     iter = 4000, warmup = 2000)

summary(mod_ex3a)
```

```
MCMC diagnostics
              mcse Rhat n_eff
(Intercept)   0.0  1.0  4007
blackyes      0.0  1.0  4478
femaleyes     0.0  1.0  4235
mean_PPD      0.0  1.0  3660
log-posterior 0.0  1.0  1657
```

```
stan_trace(mod_ex3a, nrow = 3, ncol = 1)
```



stan_ac(mod_ex3a)



b) Logistic Regression Model with Random Intercept

```r
mod_ex3b <- stan_glmer(bush~black + female + (1|state),
                       data = data3,
                       family = "binomial",
                       iter = 4000, warmup = 2000)
```

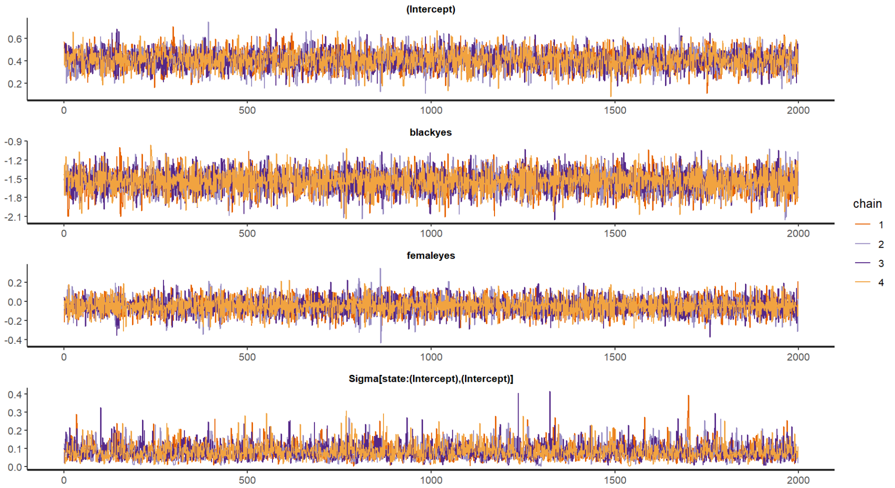
```r
summary(mod_ex3b)
```

```
MCMC diagnostics
                                       mcse Rhat n_eff
(Intercept)                            0.0  1.0   5738
blackyes                               0.0  1.0  12033
femaleyes                              0.0  1.0  12267
b[(Intercept) state:1]                 0.0  1.0  11922
b[(Intercept) state:3]                 0.0  1.0  13115
```

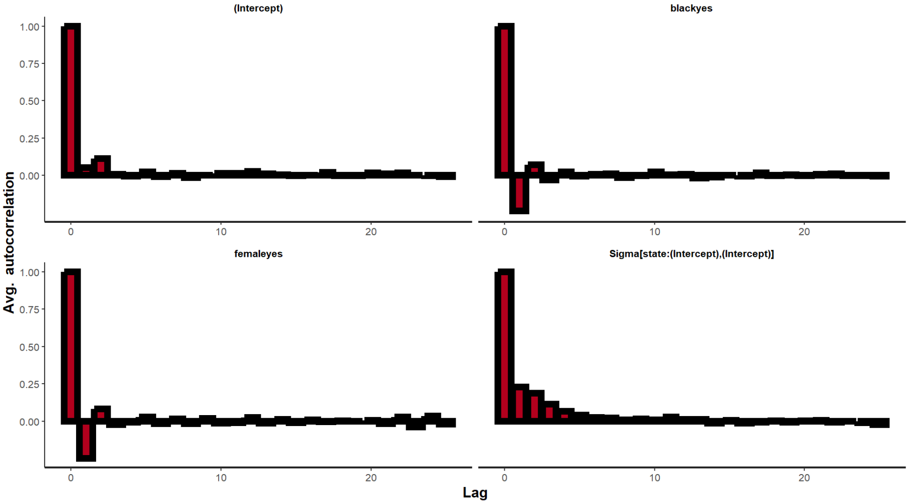
...

```
b[(Intercept) state:51]                0.0  1.0  12010
Sigma[state:(Intercept),(Intercept)]   0.0  1.0   3220
mean_PPD                               0.0  1.0   9876
```

```
stan_trace(mod_ex3b, nrow = 4, ncol = 1,
        pars = c("(Intercept)", "blackyes", "femaleyes",
        "Sigma[state:(Intercept),(Intercept)]"))
```



```
stan_ac(mod_ex3b, pars = c("(Intercept)", "blackyes",
"femaleyes", "Sigma[state:(Intercept),(Intercept)]"))
```



Model Comparison

```
waic(mod_ex3a)
waic(mod_ex3b)
```

WAIC		
Model	Estimate	SE
mod_ex3a	3464.0	22.7
mod_ex3b	3445.5	24.2

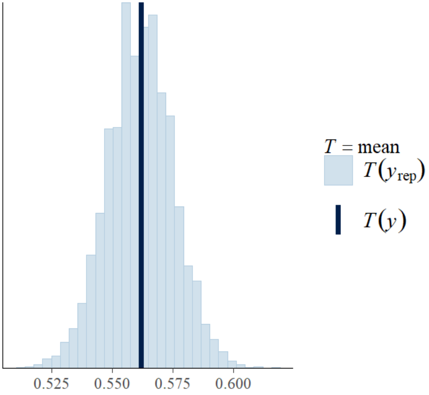
Summary of the better model

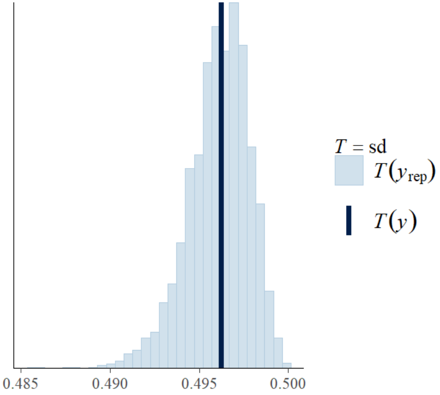
```
main_pars <- c("(Intercept)", "blackyes",
"femaleyes","Sigma[state:(Intercept),(Intercept)]")

summary(mod_ex3b, pars = main_pars, digits = 3)
```

```
Model Info:
 function:     stan_glmer
 family:       binomial [logit]
 formula:      bush ~ black + female + (1 | state)
 algorithm:    sampling
 sample:       8000 (posterior sample size)
 priors:       see help('prior_summary')
 observations: 2591
 groups:       state (49)

Estimates:
                                       mean   sd     10%    50%    90%
(Intercept)                           0.409  0.082  0.304  0.411  0.512
blackyes                             -1.542  0.170 -1.764 -1.538 -1.328
femaleyes                            -0.048  0.084 -0.156 -0.048  0.060
Sigma[state:(Intercept),(Intercept)]  0.083  0.042  0.037  0.076  0.138

MCMC diagnostics
                                     mcse  Rhat  n_eff
(Intercept)                          0.001 1.000  5738
blackyes                             0.002 1.000 12033
femaleyes                            0.001 1.000 12267
Sigma[state:(Intercept),(Intercept)] 0.001 1.000  3220
```

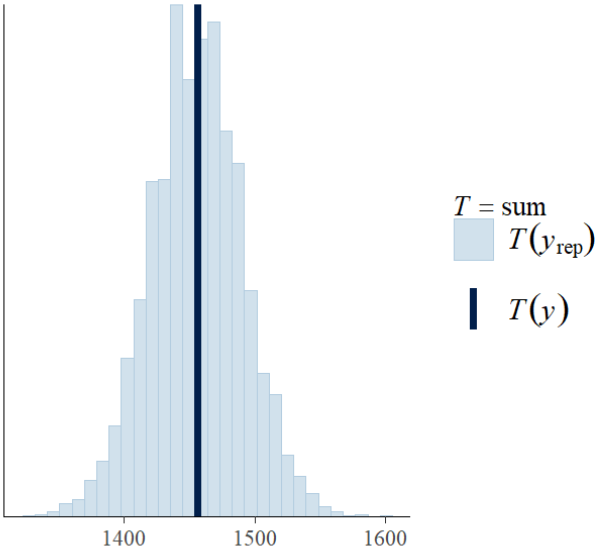
Posterior predictive checks

```
y_tilde <- posterior_predict(mod_ex3b)

ppc_stat(y = data3$bush, yrep = y_tilde, stat = "mean")
ppc_stat(y = data3$bush, yrep = y_tilde, stat = "sd")
```





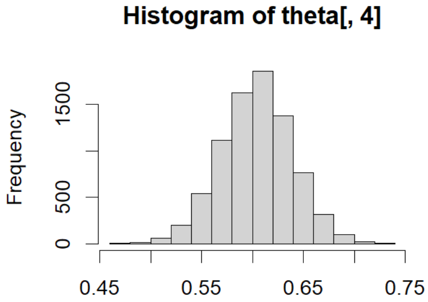
```
ppc_stat(y = data3$bush, yrep = y_tilde, stat = "sum")
```



Posterior Inference

Example: estimated posterior probability for subject 4

```r
theta <- posterior_linpred(mod_ex3b, transform = TRUE)

hist(theta[,4])
mean(theta[,4]); sd(theta[,4])
# 95% Credibility Interval
quantile(theta[,4], probs = c(0.025,0.5,0.975))
```




```
> hist(theta[,4])
> mean(theta[,4]);sd(theta[,4])
[1] 0.6041644
[1] 0.03486138
> quantile(theta[,4],
+          probs = c(0.025,0.5,0.975))
     2.5%       50%     97.5%
0.5347957 0.6049091 0.6715410
>
```

[Extra] Additional models

- **(c) Logistic Regression Model with random effect on variable** `race`:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right)|\boldsymbol{\beta} = \beta_0 + \beta_{1[j]}\texttt{race}_{ij} + \beta_2\texttt{gender}_{ij}, \quad j = 1, \ldots, 49, \; i = 1, \ldots, n_j.$$

```
mod_ex3c <- stan_glmer(bush~black + female +
                            (black|state),
                       data = data3,
                       family = "binomial")
```

- **(d) Logistic Regression Model with random effect on intercept and both variables**:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right)|\boldsymbol{\beta} = \beta_{0[j]} + \beta_{1[j]}\texttt{race}_{ij} + \beta_{2[j]}\texttt{gender}_{ij}, \; j = 1, \ldots, 49, \; i = 1, \ldots, n_j.$$

```
mod_ex3d <- stan_glmer(bush~black + female +
                            (1+black+female|state),
                       data = data3,
                       family = "binomial")
```