



# Bayesian Analysis of linear models: Classification effects models

In the classification effects model, the columns of the incidence matrix includes the value 0 or 1 to link an observation to a group that is represented by levels on a factor. This is the case of frequentist ANOVA (analysis of variance). A first issue to deal with in these models is that incidence matrix is not full column rank. To circumvent this problem we will set the effect of one level to zero.

## Data and Model

In Falconer & Makay (page 206) a dataset on weights of male and female flies from 10 matings is presented (see R file for the data).

We will analyze this dataset using the following model:

$$y_i = \mu + S_i + e_{ij}$$

Where  $\mu$  is the overall mean,  $S_i$  is the effects of the  $i$ th level of sex (two levels).

## Model assumptions

We will assume that  $e_{ij} \sim N(0, \sigma_e)$ . In addition we need to set up priors for the error variance, for the sex effect. First, we set restrictions:  $S_1 = 0$ , which means that the effects of the first level of each factor is fixed to zero. Note: to do this in R, we will have to assign a VERY informative prior (see below)

## Flat improper priors

Next, for the first level of sex, we specify the following prior:  $S_1 \sim N(0, 0.0001)$ , which is a very informative prior, reflecting our restriction  $S_1 = 0$ .

For the other level of sex effect, we specify flat improper prior,  $S_2 \sim U(-\infty, \infty)$ .

We also specify a flat improper prior for sigma:  $\sigma \sim U(0, \infty)$ .

## Quantities of interest

We want to make inference about the treatment effects. In particular, we want to know the following:

### a) Posterior distribution of each treatment mean.

This is obtained by summing the overall mean plus the level of each sex:

Mean of males (Sex=2) =  $\mu + S_2$

Mean of females (Sex=1) =  $\mu + S_1$  (because we set the effect of the first level to zero)

## b) Posterior distribution of the difference between sexes and tail area probability

This is just the posterior distribution of the difference between the two sex effects. We will look into this after we fit the models (see section below).

## c) Error variance

We will also look into the posterior distribution of the error variance.

## Alternative parameterization: Cell mean model

We can fit an alternative model, that does not include the intercept:  $y_i = S_i + e_{ij}$ , but with  $S_i \sim U(-\infty, \infty)$  for all levels of sex (cell\_mean.stan).

This model should produce the same results as the previous one. We just don't need to add the overall mean to the sex effect in order to obtain the mean of the of each sex class. We just obtain that from the sex effect.

## Expanded model: Add a Mating (block) effect

We can add a mating effect to either of the two models presented above, for instance:

$$y_i = \mu + S_i + M_j + e_{ij}$$

### Treating the Mating effect as a fixed effect

And we need to set a prior for the mating effect, for instance:  $M_j \sim U(-\infty, \infty)$  would be equivalent to treating the mating effect as a fixed effect in a frequentist analysis. However, if we use this prior, we may run into identifiability issues, this is, we can't estimate all levels of the mating effect. One solution is to add a zero restriction, such as:  $M_1 = 0$ , in the form of a very informative prior:  $M_1 \sim N(0, 0.0001)$ .

This is a possible solution, however, now when we want to estimate the treatment means, we need to add the mean of the mating effect to:

$$\text{Mean of sex } i = \mu + S_i + \frac{\sum_j M_j}{10}$$

This model results in slow convergence and high autocorrelation because we estimate many parameters from a small dataset.

### Treating the mating effect as a random effect

An alternative consists of treating the Mating effect as if it was a random effect in a frequentist analysis. For that, the Bayesian solution assigns a normal prior on the mating effect:  $M_j \sim N(0, \sigma_M)$  and on its standard deviation:  $\sigma_m \sim U(0, \infty)$ .

This model specification will estimate the mating variance and the treatment means. Moreover, because the prior for the mating effects is a normal centered at zero, we don't need to add the average of the mating effect to the estimation of the treatment mean:

$$\text{Mean of sex } i = \mu + S_i$$

**This model may result in even WORSE convergence properties. Including divergent transitions that could yield biased estimates. We will deal with this in the lab practice.**

## Programming in stan

For programming these models in Stan, we will need to introduce two separate variables:

1. A vector of length N (number of observations) that indicates which level of the sex (or mating) effect corresponds to each observation.
2. A vector representing the sex (or mating) effect itself. This vector will have the length equal to the number of levels in the factor in question (2 for sex and 10 for M)

Let's see an example, with the stan code for the first model (overall mean and sex effect):

```
data {
  int<lower=0> N;
  int<lower=0> nsex;
  int<lower=0> sex[N];
  real Y[N];
}

parameters {
  real sexeff[nsex];
  real mu;
  real<lower=0> sigma;
}

model {
  real m[N];
  sexeff[1]~normal(0.0,0.001);

  for (i in 1:N)
    m[i] = mu+sexeff[sex[i]];
  Y ~ normal(m, sigma);
}



generated quantities{
  real S1;
  real S2;
  real <lower=0> vare;
  S1= mu + sexeff[1];
  S2= mu + sexeff[2];
  vare= sigma^2;
}
```

In this code, sex is a vector of length N=20, with values 1 or 2, representing the sex of the corresponding observation. while sexeff is a vector of length nsex=2 representing the effects of sex 1 and of sex 2.

## Posterior distribution of the difference between sexes and tail area probability

Once we understood the modeling of sex and mating effects, we will estimate the tail area probabilities. This is formally expressed as:  $\min(P(S_2 - S_1 > 0|\mathbf{y}), P(S_2 - S_1 < 0|\mathbf{y}))$ .

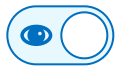
To obtain this quantity in stan, we will create two quantities: one is the estimated sex difference and the other one is an indicator variable comparing the sex difference to zero:

 Edit HTML Reflect in ePortfolio Download Print Alternative formats

## Activity Details

## Learning Objectives

## Completion Summary



Visibility

**Required: Automatic** ▼

View this topic to complete the activity

**Options**

Reflecting in ePortfolio is enabled

Bayesian Analysis of linear models Classifications effects models.html - Last Modified Oct 12, 2020 4:37 PM