

# Varying Intercepts Models

## General Principles

To model the relationship between a dependent variable and an independent variable while allowing for different intercepts across groups or clusters, we can use a *Varying Intercepts* model. This approach is particularly useful when data is grouped (e.g., by subject, location, or time period) and we expect the baseline level of the outcome to vary across these groups.

## Considerations

### Caution

- We have the same considerations as for [Regression for a continuous variable](#).
- The main idea of varying intercepts is to generate an intercept for each group, allowing each group to start at different levels. Thus, the intercept  $\alpha_k$  is defined based on the  $k$  declared groups.
- Each intercept has its own prior - i.e., a hyper-prior .
- In the code below, the *hyper-prior* is `a_bar`.

## Example

Below is an example code snippet demonstrating Bayesian regression with varying intercepts using the Bayesian Inference (BI) package. The data consists of a dependent variable representing individuals' survival (*surv*) and an independent categorical variable (*tank*), which indicates the tank where the individual was born, with a total of 48 tanks. This example is based on McElreath (2018).

## Python

```
from BI import bi
import numpy as np

# Setup device-----
m = bi(platform='cpu')

# Import Data & Data Manipulation -----
# Import
from importlib.resources import files
data_path = files('BI.resources.data') / 'reedfrogs.csv'
m.data(data_path, sep=';')
# Manipulate
m.df["tank"] = np.arange(m.df.shape[0])

# Define model -----
def model(tank, surv, density):
    sigma = m.dist.exponential( 1, name = 'sigma')
    a_bar = m.dist.normal( 0., 1.5, name = 'a_bar')
    alpha = m.dist.normal( a_bar, sigma, shape= tank.shape, name = 'alpha')
    p = alpha[tank]
    m.dist.binomial(total_count = density, logits = p, obs=surv)

# Run sampler -----
m.fit(model)

# Diagnostic -----
m.summary()
```

## R

```
library(BI)

# setup platform-----
m=importbi(platform='cpu')

# Import data -----
m$data(paste(system.file(package = "BI"),"/data/reedfrogs.csv", sep = ''), sep=';')
m$df$tank = c(0:(nrow(m$df)-1)) # Manipulate
m$data_to_model(list('tank', 'surv', 'density')) # Manipulate
```

```

m$data_on_model$tank = m$data_on_model$tank$astype(jnp$int32) # Manipulate
m$data_on_model$surv = m$data_on_model$surv$astype(jnp$int32) # Manipulate

# Define model -----
model <- function(tank, surv, density){
  # Parameter prior distributions
  sigma = bi.dist.exponential( 1, name = 'sigma',shape=c(1))
  a_bar = bi.dist.normal(0, 1.5, name='a_bar',shape=c(1))
  alpha = bi.dist.normal(a_bar, sigma, name='alpha', shape =c(48))
  p = alpha[tank]
  # Likelihood
  m$binomial(total_count = density, logits = p, obs=surv)
}

# Run MCMC -----
m$fit(model) # Optimize model parameters through MCMC sampling

# Summary -----
m$summary() # Get posterior distribution

```

## Mathematical Details

### *Frequentist formulation*

We model the relationship between the independent variable  $X$  and the outcome variable  $Y$  with varying intercepts  $\alpha$  for each group  $k$  using the following equation:

$$Y_{ik} = \alpha_k + \beta X_{ik} + \sigma$$

Where:

- $Y_{ik}$  is the outcome variable for observation  $i$  in group  $k$ .
- $\alpha_k$  is the varying intercept for group  $k$ .
- $X_{ik}$  is the independent variable for observation  $i$  in group  $k$ .
- $\beta$  is the regression coefficient.
- $\sigma$  is the error term, typically assumed to be normally distributed and positive.

## Bayesian Model

We can express the Bayesian regression model accounting for priors distributions as follows:

$$Y_{ik} \sim \text{Normal}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \alpha_k + \beta X_{ik}$$

$$\alpha_k \sim \text{Normal}(\mu_{\alpha_k}, \sigma_{\alpha_k})$$

$$\beta \sim \text{Normal}(0, 1)$$

$$\sigma \sim \text{Exponential}(1)$$

$$\mu_{\alpha_k} \sim \text{Normal}(0, 1)$$

$$\sigma_{\alpha_k} \sim \text{Exponential}(1)$$

Where:

- $Y_{ik}$  is the outcome variable for observation  $i$  in group  $k$ .
- $\alpha_k$  is the varying intercept for group  $k$ .
- $\mu_{\alpha_k}$  is the overall mean intercept.
- $\sigma_{\alpha_k}$  is the variance of the intercepts across groups.
- $\beta$  is the regression coefficient.
- $\sigma$  is a standard deviation parameter, which here has a *Exponential* prior that constrains it to be positive.

## Notes

### Note

- We can apply multiple variables similarly to [Chapter 2](#).
- We can apply interaction terms similarly to [Chapter 3](#).
- We can apply categorical variables similarly to [Chapter 4](#).
- We can apply varying intercepts with any distribution developed in previous chapters.

## Reference(s)

McElreath, Richard. 2018. *Statistical Rethinking: A Bayesian Course with Examples in r and Stan*. Chapman; Hall/CRC.