

# RSTAN - Linear Regression

## Simple Linear Regression

### Load the packages

```
library(rstan)
```

### Model

$$y \sim \text{Normal}(\alpha + \beta * X, \sigma^2)$$

### Data

r에 내장되어있는 cars 데이터 사용.

```
#data
data1 <- cars

#list_data
data1 <- list(
  N = nrow(data1),
  y = data1$dist,
  x = data1$speed
)
```

### STAN Code

```
data {
  int<lower=0> N;
  vector[N] x; //predictor
  vector[N] y; //response
}

parameters {
  real alpha;
  real beta;
  real<lower=0> sigma;
}

model {
  y ~ normal(alpha + beta*x, sigma);
}
```

## Scatter Plot

```
plot(data1$x, data1$y, xlab = 'speed', ylab='dist')
```

## Fitting

```
fit1 <- stan(file = 'simple_linear.stan', data = data1)
```

## Trace Plot

```
traceplot(fit1, inc_warmup = TRUE)
```

## Estimated Parameters

```
plot(fit1, pars = 'alpha', show_density=TRUE)
plot(fit1, pars = 'beta', show_density=TRUE)
plot(fit1, pars = 'sigma', show_density=TRUE)
```

## Multiple Linear Regression

---

### No Categorical Data Case

- 모든 데이터의 형태가 Numerical

#### Data

R에 내장되어있는 attitude 데이터 사용

```
#data
data2 <- attitude

#listdata
data2 <- list(
  N = nrow(data2),
  x = data2[, -1],
  K = ncol(data2)-1,
  y = data2$rating
)
```

## STAN Code

```
data {  
  int<lower=0> N; // number of obs  
  int<lower=0> K; // number of predictors  
  matrix[N, K] x; // predictor matrix  
  vector[N] y; // outcome vector  
}  
  
parameters {  
  real alpha; // intercept  
  vector[K] beta; // coefficients for predictors  
  real<lower=0> sigma; // error scale  
}  
  
model {  
  y ~ normal(x*beta + alpha, sigma); // likelihood  
}
```

## Fitting

```
fit2 <- stan(file = 'multiple_linear.stan', data = data2)
```

## Trace Plot

```
traceplot(fit2, inc_warmup = TRUE)
```

## Estimated Parameters

```
plot(fit2, pars = 'alpha', show_density=TRUE)  
plot(fit2, pars = 'beta[1]', show_density = TRUE)  
plot(fit2, pars = 'beta[2]', show_density = TRUE)  
plot(fit2, pars = 'sigma', show_density = TRUE)
```

## Categorical Data Included Case

### Data

MASS Package 안에 있는 birthwt 데이터 이용.

```
#data  
library(MASS)  
str(birthwt)  
birthwt$race <- as.factor(birthwt$race)  
  
str(birthwt)  
design_matrix <- model.matrix(bwt ~ ., data = birthwt)
```

```
#listdata
data3 <- list(
  N = nrow(design_matrix),
  x = design_matrix[,-1],
  K = ncol(design_matrix) - 1,
  y = birthwt$bwt
)
```

## STAN Code

위의 케이스와 동일

## Fitting

```
fit3 <- stan(file = 'multiple_linear.stan', data = data3)
```

## Trace Plot

```
traceplot(fit3, inc_warmup = TRUE)
```

## Estimated Parameters

```
plot(fit3, pars = 'alpha', show_density = TRUE)
plot(fit3, pars = 'beta[1]', show_density = TRUE)
plot(fit3, pars = 'sigma', show_density=TRUE)
```