

Univariate Linear Regression

General Principles

To study relationships between a continuous independent variable and a continuous dependent variable (e.g., height and weight), we can use linear regression. Essentially, we draw a line that passes through the point cloud of the two variables being tested. For this, we need to have:

- 1) An intercept α , which represents the origin of the line, i.e., the expected value of the dependent variable (height) when the independent variable (weight) is equal to zero.
- 2) A coefficient β , which informs us about the slope of the line. In other words, it tells us how much Y (height) increases for each increment of the independent variable (weight).
- 3) A standard deviation term σ , which informs us about the spread of points around the line, i.e., the variance around the prediction.

Considerations

Note

- Bayesian models allow us to update our understanding of parameters conditional on an observed data set. This allows us to consider model parameter uncertainty , which quantifies our confidence or uncertainty in the parameters in the form of a posterior distribution . Therefore, we need to declare prior distributions for each model parameter, in this case for: α , β , and σ .
- Prior distributions are built following these considerations:
 - As the data are normalized (see introduction), we can use a Normal distribution for α and β , with a mean of 0 and a standard deviation of 1. This tends to be a weakly regularizing prior, and weaker priors like a $Normal(0, 10)$ are also possible.
 - Since σ must be strictly positive, we must use a distribution with support on the positive reals, such as the *Exponential* or *Folded-Normal* distribution.

- Gaussian regression deals directly with continuous outcomes, estimating a linear relationship between predictors and the outcome variable without depending on a non linear link function (see introduction). This simplifies interpretation, as coefficients represent direct changes in the outcome variable.

Example

Below is an example code snippet demonstrating *Bayesian linear regression* using the Bayesian Inference (**BI**) package. Data consist of two continuous variables (height and weight), and the goal is to estimate the effect of weight on height. This example is based on McElreath (2018).

Python

```
from BI import bi

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

# Import Data & Data Manipulation -----
# Import
from importlib.resources import files
data_path = m.load.howell1(only_path = True)
m.data(data_path, sep=';')
m.df = m.df[m.df.age > 18] # Subset data to adults
m.scale(['weight']) # Normalize

# Define model -----
def model(weight, height):
    a = m.dist.normal(178, 20, name = 'a')
    b = m.dist.log_normal(0, 1, name = 'b')
    s = m.dist.uniform(0, 50, name = 's')
    m.dist.normal(a + b * weight , s, obs = height)

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

# Summary -----
m.summary() # Get posterior distributions

jax.local_device_count 32
```

```

0%| 0/1000 [00:00<?, ?it/s] warmup: 0%| 1/1000 [00:00<06:33, 2.54i
arviz - WARNING - Shape validation failed: input_shape: (1, 500), minimum_shape: (chains=2, c

```

	mean	sd	hdi_5.5%	hdi_94.5%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
a	154.64	0.26	154.27	155.14	0.01	0.01	422.23	407.54	NaN
b	5.82	0.29	5.40	6.31	0.02	0.01	376.62	356.33	NaN
s	5.15	0.20	4.81	5.47	0.01	0.01	446.91	333.97	NaN

R

```

library(BayesianInference)
m <- importBI(platform = "cpu")

# Load csv file
m$data(m$load$howell1(only_path = T), sep = ";")

# Filter data frame
m$df <- m$df[m$df$age > 18, ] # Subset data to adults

# Scale
m$scale(list("weight")) # Normalize

# Convert data to JAX arrays
m$data_to_model(list("weight", "height"))

# Define model -----
model <- function(height, weight) {
  # Parameter prior distributions
  s <- bi.dist.uniform(0, 50, name = "s")
  a <- bi.dist.normal(178, 20, name = "a")
  b <- bi.dist.normal(0, 1, name = "b")

  # Likelihood
  bi.dist.normal(a + b * weight, s, obs = height)
}

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

```

```
# Summary -----
m$summary()
```

Julia

```
using BayesianInference

# Setup device-----
m = importBI(platform="cpu")

# Import Data & Data Manipulation -----
# Import
data_path = m.load.howell1(only_path = true)
m.data(data_path, sep=';')
m.df = m.df[m.df.age > 18] # Subset data to adults
m.scale(["weight"]) # Normalize

# Define model -----
@BI function model(weight, height)
    # Priors
    a = m.dist.normal(178, 20, name = 'a')
    b = m.dist.log_normal(0, 1, name = 'b')
    s = m.dist.uniform(0, 50, name = 's')
    m.dist.normal(a + b * weight , s, obs = height)
end

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

# Summary -----
m.summary() # Get posterior distributions
```

Mathematical Details

Frequentist Formulation

The following equation describe the frequentist formulation of linear regression:

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Where:

- Y_i is the dependent variable for observation i .
- α is the intercept term.
- β is the regression coefficient.
- X_i is the input variable for observation i .
- ϵ_i is the error term for observation i , and the vector of the error terms, ϵ , are assumed to be independent and identically distributed.

Bayesian Formulation

In the Bayesian formulation, we define each parameter with priors . We can express a Bayesian version of this regression model using the following model:

$$Y_i \sim \text{Normal}(\alpha + \beta X_i, \sigma)$$

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

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

$$\sigma \sim \text{Uniform}(0, 50)$$

Where:

- Y_i is the dependent variable for observation i .
- α and β are the intercept and regression coefficient, respectively.
- X_i is the independent variable for observation i .
- σ is the standard deviation of the Normal distribution, which describes the variance in the relationship between the dependent variable Y and the independent variable X .

Notes

Note

We observe a difference between the *Frequentist* and the *Bayesian* formulation regarding the error term. Indeed, in the *Frequentist* formulation, the error term ϵ represents residual fluctuations around the predicted values. This assumption leads to point estimates for α and β . In contrast, the *Bayesian* formulation treats σ as a parameter with its own prior distribution. This allows us to incorporate our uncertainty about the error term into the model.

Reference(s)

McElreath, Richard. 2018. *Statistical Rethinking: A Bayesian course with examples in R and Stan*. Chapman; Hall/CRC.