Interaction Terms in Regression

General Principles

To study the relationships between two independent continuous variables and their interaction effect on a dependent variable (e.g., temperature and humidity affecting energy consumption), we can use Regression Analysis with Interaction Terms. In this approach, we extend the simple linear regression model to include an interaction term (a multiplication) between the two continuous variables.

Considerations

> Caution
>
> - We have the same assumptions as for Regression for continuous variable.
> - We wish to model the relationship between dependent variable Y and independent variable X_1 to vary as a function of independent variable X_2. To do this, we explicitly model the hypothesis that the slope between Y and X_1 depends—is conditional—upon X_2.
> - For continuous interactions with scaled data, the intercept becomes the grand mean of the outcome variable.
> - The interpretation of estimates is more complex. The estimate for a non-interaction term reflects the expected change in Y when X_1 increases by one unit, holding X_2 constant at its average value. The estimate for the interaction term represents how the effect of X_1 on Y changes depending on the value of X_2, and vice versa, showing how the relationship between the two variables influences the outcome Y.
> - Triptych plots are very handy for understanding the impact of interactions, especially when more than two interactions are present.

Example

Below is example code demonstrating Bayesian regression with an interaction term between two continuous variables using the Bayesian Inference (BI) package. The data consist of three continuous variables (temperature, humidity, energy consumption), and the goal is to estimate the effect of the interaction between temperature and humidity on energy consumption. This example is based on McElreath (2018).

Python

```python
from BI import bi

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

# Import Data & Data Manipulation ------------------------------------------------
# Import
from importlib.resources import files
data_path = files('BI.resources.data') / 'tulips.csv'
m.data(data_path, sep=';')
m.scale(['blooms', 'water', 'shade']) # Scale


# Define model ------------------------------------------------
def model(blooms,shade, water):
    sigma = m.dist.exponential(1, name = 'sigma', shape = (1,))
    bws = m.dist.normal(0, 0.25, name = 'bws', shape = (1,))
    bs = m.dist.normal(0, 0.25, name = 'bs', shape = (1,))
    bw = m.dist.normal(0, 0.25, name = 'bw', shape = (1,))
    a = m.dist.normal(0.5, 0.25, name = 'a', shape = (1,))
    mu = a + bw*water + bs*shade + bws*water*shade
    m.normal(mu, sigma, obs=blooms)

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

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

R

```r
library(BI)
m=importbi(platform='cpu')

# Load csv file
m$data(paste(system.file(package = "BI"),"/data/tulips.csv", sep = ''), sep=';')
m$scale(list('blooms', 'water', 'shade')) # Scale
m$data_to_model(list('blooms', 'water', 'shade')) # Send to model (convert to jax array)

# Define model ------------------------------------------------
model <- function(blooms, water,shade){
  # Parameter prior distributions
  alpha = bi.dist.normal( 0.5, 0.25, name = 'a')
  beta1 = bi.dist.normal( 0,  0.25, name = 'b1')
  beta2 = bi.dist.normal(  0,  0.25, name = 'b2')
  beta_interaction_ = bi.dist.normal(  0, 0.25, name = 'bint')
  sigma = bi.dist.normal(0, 50, name = 's')
  # Likelihood
  m$normal(alpha + beta1*water + beta2*shade + beta_interaction_*water*shade, sigma, obs=bloo
}

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

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

Mathematical Details

Frequentist formulation

We model the relationship between the input features (X_1 and X_2) and the target variable (Y) using the following equation:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{interaction} X_{1i} X_{2i} + \sigma$$

Where:

- Y_i is the dependent variable for observation i.

- α is the intercept term.
- X_{1i} and X_{2i} are the two values of the independent continuous variables for observation i.
- β_1 and β_2 are the regression coefficients for X_1 and X_2, respectively.
- $\beta_{interaction}$ is the regression coefficient for the interaction term $(X_1 X_2)$.
- σ is the error term, assumed to be normally distributed.

In this context, the interaction term $X_{1i} * X_{2i}$ captures the joint effect of X_{1i} and X_{2i} on the target variable Y_i.

Bayesian formulation

In the Bayesian formulation, we define each parameter with priors . We can express the Bayesian regression model accounting for prior distributions as follows:

$$Y \sim Normal(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{interaction} X_{1i} X_{2i}, \sigma)$$

$$\alpha \sim Normal(0, 1)$$

$$\beta_1 \sim Normal(0, 1)$$

$$\beta_2 \sim Normal(0, 1)$$

$$\beta_{interaction} \sim Normal(0, 1)$$

$$\sigma \sim Exponential(1)$$

Where:

- Y_i is the dependent variable for observation i.
- α is the prior distribution for the intercept.
- β_1, β_2, and $\beta_{interaction}$ are the prior distributions for the regression coefficients.
- X_{1i} and X_{2i} are the two values of the independent continuous variables for observation i.
- σ is the prior distribution for the standard deviation, ensuring it is positive.

Reference(s)

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