**Bayesian Statistics Homework 4**

**Stefanela Stevanović**

## Introduction

In this homework, we build different models that estimate how various features influence happiness, using the World Happiness Report dataset. Our goal is to evaluate the quality of these models, gaining insights into how these models compare against each other and informing decisions about potential model combinations.

## Methods

### Data

For this homework, we used data sourced from the official World Happiness Report website: *https://worldhappiness.report/data/*. We extracted data for seven years, from 2016 to 2022, removed all rows with NA values, and excluded the *healthy life expectancy* feature due to its high correlation with *economy* and *social support* features. The *happiness score*, representing the happiness of a country's residents, where a higher score indicates greater happiness, was modeled using the following features:

- *economy*: the natural log of GDP per capita.
- *social support*: the national average of the binary responses (either 0 or 1) to the Gallup World Poll (GWP) question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?".
- *freedom*: the national average of binary responses to the GWP question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?".
- *perceived corruption*: the national average of binary answers to two GWP questions: "Is corruption widespread throughout the government or not?" and "Is corruption widespread within businesses or not?"
- *generosity*: a function of the national average of GWP responses to the question "Have you donated money to a charity in the past month?" on GDP per capita.

Furthermore, min-max normalization was employed to scale each feature in a data set between a [0, 1] range. The distributions of the values for all features and the target variable are displayed in Figure 1.

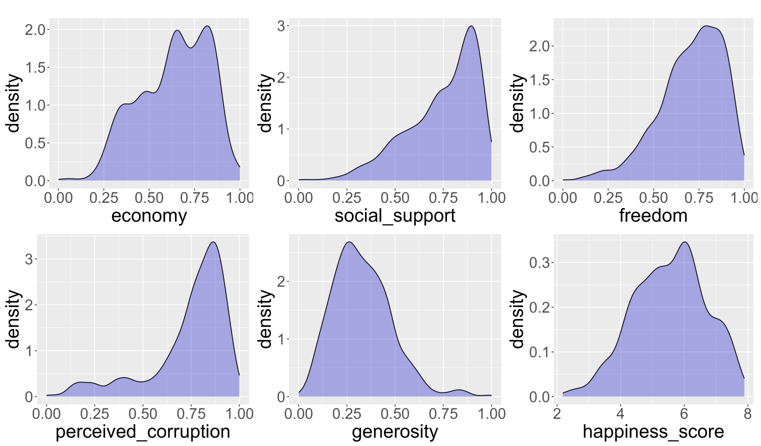


**Figure 1.** Distributions of *economy*, *social support*, *freedom*, *perceived corruption*, *generosity* features and *happiness score* target variable.

### Models

We built 10 different normal models $y_m \sim \mathcal{N}(\mu_m, \sigma_m), \forall m \in \{1, 2, ...10\}$, with their link functions shown in Table 1, using the default Stan priors. The beta values represent the coefficients of the models, while features are indicated by x. Each x is accompanied by a subscript indicating which feature it represents: $e$ for *economy*, $s$ for *social support*, $f$ for *freedom*, $c$ for *perceived corruption*, and $g$ for *generosity*. The first 5 models are simple linear models, where each successive model includes one additional feature. The remaining 5 models involve interactions among features.

**Table 1.** Model names and their corresponding link functions.

| Model | Link Function ($\mu$) |
|---|---|
| 1 | $\beta_0 + \beta_e x_e$ |
| 2 | $\beta_0 + \beta_e x_e + \beta_s x_s$ |
| 3 | $\beta_0 + \beta_e x_e + \beta_s x_s + \beta_f x_f$ |
| 4 | $\beta_0 + \beta_e x_e + \beta_s x_s + \beta_f x_f + \beta_c x_c$ |
| 5 | $\beta_0 + \beta_e x_e + \beta_s x_s + \beta_f x_f + \beta_c x_c + \beta_g x_g$ |
| 6 | $\beta_0 + \beta_e x_e + \beta_s x_s + \beta_f x_f + \beta_c x_c + \beta_{e-c} x_e x_c$ |
| 7 | $\beta_0 + \beta_e x_e + \beta_s x_s + \beta_f x_f + \beta_c x_c + \beta_{e-f} x_e x_f$ |
| 8 | $\beta_0 + \sum_{i \in \{s,f,c,g\}} \beta_{e-i} x_e x_i$ |
| 9 | $\beta_0 + \sum_{i \in \{s,f,c,g\}} \beta_{e-i} x_e x_i + \sum_{j \in \{s,f,g\}} \beta_{c-j} x_c x_j$ |
| 10 | $\beta_0 + \sum_{i \in \{s,f,c,g\}} \beta_{e-i} x_e x_i + \sum_{j \in \{s,c,g\}} \beta_{f-j} x_f x_j$ |

The posterior distributions for each model were explored using 4 MCMC chains, each with 500 warm-up and 500 sampling iterations. MCMC diagnostics, including Gelman-Rubin (R-hat), trace plots, and effective sample size, all indicated satisfactory results, affirming that MCMC chains have effectively explored parameter space and converged to the posterior distribution.

### Evaluation Criteria

In evaluating the performance of our models, we used Leave-One-Out Information Criterion (LOOIC). LOOIC provides insights into the predictive accuracy and generalization capabilities of our models. A model with a lower LOOIC is considered better in terms of predictive accuracy. Furthermore, we used Akaike weights to determine how decisions made by multiple models should be mixed together (averaged).

## Results

The obtained results are presented in this section. The Figure 2. shows LOOIC values of 10 models, while Akaike weights of the models are presented in the Figure 3.

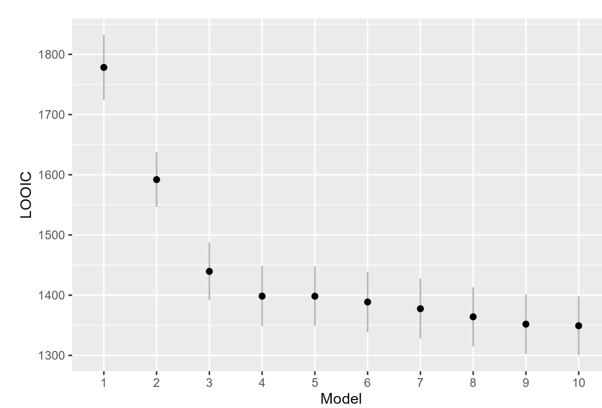


**Figure 2.** Leave one out information criterion (LOOIC) of 10 models.

If we observe the LOOIC values in Figure 2 for the first 4 models, we notice that by including one more feature in the model, its performance improves (lower LOOIC). Between the fourth and fifth model, we do not observe a significant difference, concluding that added the generosity feature in the fifth model does not significantly impact the performance. This suggests that this feature is less influential in estimating happiness score compared to others. Furthermore, incorporating interactions between features further improves model performance. Model 9, which includes interactions between the economy and other features, as well as perceived corruption and other features, and Model 10, which includes interactions between the economy and other features and freedom and other features, show the best performance.

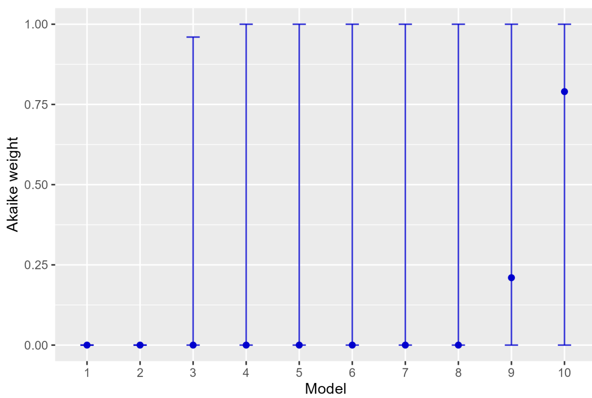


**Figure 3.** Akaike weights of 10 models.

In Figure 3, showing the Akaike weights of the models, we observe that in the case of model combination, we should combine models 9 and 10, as these are the only models with Akaike weights greater than 0. This suggests that these models significantly outperform the others, with Model 10 exhibiting the best performance. Akaike weights can be interpreted as estimates of the probability that a model will make the best prediction on new data. This means that Model 9 is expected to yield the best prediction in 21% of cases, while Model 10 is expected to do so in 79% of cases. Therefore, for weighing their decision in a combined model, these values should be utilized.

## Discussion

The results of this homework have shown that, in modeling the *happiness score* from the World Happiness Report, incorporating multiple features and interactions between features leads to improved model performance. The top-performing model was identified as Model 10, which includes interactions between the economy and all other features, as well as freedom and all other features. It was followed by Model 9, which includes interactions between the economy and all other features, along with perceived corruption and all other features. Furthermore, our analysis has shown that Model 10 is expected to provide the best predictions in 79% of cases, while Model 9 is expected to do so in 21% of cases. If we were interested in combining predictions from multiple models, it would be advisable to use Model 9 and Model 10, assigning weights to their predictions of 0.21 and 0.79, respectively.