# Exercise 2: Distributional Regression

### Exercise 1 (Childhood Malnutrition in Zambia)

Reconsider the Zambia data already analyzed in a previous exercise.

1. In some situations, results may be sensitive with respect to the hyperprior for $\tau$. Compare the simple regression model

$$\text{stunting}_i \sim \mathcal{N}(\beta_0 + f(\text{agechild}_i), \sigma^2)$$

   for different specifications of the hyperprior $p(\tau)$. The prior can be specified by `s(agechild,xt=list(prior="ig",a=0.001,b=0.001)`.

2. Compare the additive regression model

$$\begin{aligned}\text{stunting}_i &\sim \mathcal{N}(\mu_i, \sigma^2)\\ \mu_i &= \beta_0 + f_1(\text{mbmi}_i) + f_2(\text{agechild}_i) + f_{\text{spat}}(\text{district}_i)\end{aligned}$$

   with the distributional regression model

$$\begin{aligned}\text{stunting}_i &\sim \mathcal{N}(\mu_i, \sigma_i^2)\\ \mu_i &= \beta_0^\mu + f_1^\mu(\text{mbmi}_i) + f_2^\mu(\text{agechild}_i) + f_{\text{spat}}^\mu(\text{district}_i)\\ \log\sigma_i &= \beta_0^\sigma + f_1^\sigma(\text{mbmi}_i) + f_2^\sigma(\text{agechild}_i) + f_{\text{spat}}^\sigma(\text{district}_i)\end{aligned}$$

   by plotting the quantile residuals and by DIC.

**Exercise 2 (Federal Elections in Germany)**

The data set `election_germany_2017.csv` contains proportions of electorates in 437 regions in Germany at the elections of the federal parliament in 2017. The 7 parties are CDU/CSU, SPD, FDP, AfD, Greens, Left and Others.

1. Estimate a Dirichlet regression model with nonlinear effects for the covariables `AQ`, `BIPpEW`, and `Wahlbeteiligung`, and a spatial effect for `nuts` with markov random field prior. The neighbourmatrix can be derived from the map data `germany_nuts.bnd`.
2. Check MCMC convergence. If necessary, increase the number of iterations and re-estimate your model.
3. Consider the location with `Kreisnummer` = 11000, which is in Berlin. Compare the predicted proportion of the votes with the observed proportions. `AQ` is a measure for the unemployment in the region. How does the prediction change, when `AQ` is increased?
4. Predict the expected proportions of electorates in each nuts region by setting all other covariates to their mean values. Plot separate maps for each party visualizing the expected proportions.
5. How could you evaluate the fit of the model?

**Exercise 3 (Munich Rent Index)**

The dataset `rent.dat` contains information on the rent index for the City of Munich. The data set comprises the following variables:

| Variable | Explanation |
|---|---|
| rent | net rent per month (in Euro) |
| rentsqm | net rent per month per square meter (in Euro) |
| area | living area in square meters |
| yearc | year of construction |
| district | district the flat is located in |
| location | quality of location according to an expert assessment<br>1 = average location<br>2 = good location<br>3 = top location |

Often, it is assumed that (conditionally on covariates) the rents are normally distributed with homoscedastic errors. In this exercise, we investigate whether alternatives to the normal distribution may yield an improved fit. The response variable considered is `rentsqm` and `munich.bnd` contains spatial information.

The dataset `rent_test.dat` contains hold-out data for out of sample comparison. Try out different distributional regression specifications. Use only the data from `rent.dat` for training and evaluate the models via the log-likelihood on the hold-out test data `rent_test.dat`. Find the best-performing model according to this metric.