**EXERCISE**

*Data set*

You are given a new set of data collected from rhesus macaques (*Macaca mulatta*) in India. The data come from 4 groups of monkeys and below you find information about the data:

|  |  |  |
| --- | --- | --- |
| Variable | Description | Type |
| Focal\_ID | ID of the monkey | Factor |
| RI | Monkey’s dominance rank | Continuous variable |
| HM\_agg\_rates | Rates of human-macaque aggression | Continuous variable |
| HM\_sub\_rates | Rates of human-macaque submission | Continuous variable |
| Prov\_rates | Rates of people feeding the monkeys | Continuous variable |
| Tot\_Gr | Number of grooming interactions | Count data |

*Analysis*

We want to assess whether rates of monkey’s grooming are significantly affected by monkey’s rank, human-macaque aggression, human-macaque submission and/or the frequency of people feeding the monkeys. The underlying hypothesis is that interactions with humans decrease macaques’ grooming time and that dominance rank affects grooming behaviour. Let’s test our hypothesis by running both linear regression and generalized linear model.

EXERCISE

1. Upload the data “hm\_data”
2. Generate a new column in which you calculate grooming rates (by dividing total number of grooming interactions by observation time).
3. Check for collinearity between the continuous ***independent*** variables. Did you find any collinearity?
4. Run linear regression to assess whether human-monkey interactions affect grooming rates, making sure you include variables that are not correlated.
5. Plot the relationship between grooming and rank, with the correct axis labels
6. Now run the analysis with Generalized Linear Model with Poisson distribution, rather than linear regression. This time the outcome (dependent) variable will be the Number of Grooming bouts (Tot\_Gr). Is the model overdispersed?
7. Run the analysis with Generalized Linear Model with Negative Binomial distribution.
8. Plot the relationship between total number of grooming bouts and Provisioning rates