

# GLM: Choose the Probability Distribution (student version)

Dr. H.G.J. van Mil (revision by Dr. F.J. Rodenburg)

September 2021

Save this file and “Day8.Rdata” in the same location. Then, after opening this file, go to **Session > Set Working Directory > To Source File Location**.

Then run this:

```
load("Day12.Rdata")
```

## Species Diversity and pH

Is there an effect of pH and biomass on diversity (**Species**)? These data are stored in DF1.

1. Fit an appropriate model to the data. Explain your choice.
2. Is there an interaction between pH and biomass?
3. Write a short conclusion in terms of which variables affect the outcome.

## Insecticide

We will use the Bliss insecticide data from the **faraway** package (install this package if you don't have it).

```
library("faraway")  
data(bliss)
```

## Cluster

Is there an effect of distance from nuclear plant and frequency of cancer? These data are stored in DF2.

- **Cancers:** Number of listed cancer occurrences;
  - **Distance:** Distance from a nuclear plant.
1. Fit an appropriate model to the data. Explain your choice.
  2. Is there a relation between distance and frequency of cancer?

## Parasitic Infection

Is parasitic infection in a certain animal species more common in males or females? What about older or heavier subjects? These data are stored in DF3.

1. Fit an appropriate model to the data (the outcome must be **infection**). Explain your choice.
2. Which variables affect the probability of infection?

## Galapagos Data (hard)

We will use the Galapagos data from the `faraway` package (install this package if you don't have it).

```
library("faraway")
data(gala)
```

1. Go to the help file of the Galapagos data by running `?gala` in the console.
2. The interest of this study lies in the number of endemic species, and how this is related to any of the other variables in the data set. The total number of species is also given. Can you think of an appropriate model? Explain.
3. Fit the model you chose below and report which explanatory variables have a significant effect.

## Blood Clotting Times (hard)

We will use the blood clotting data from the `faraway` package (install this package if you don't have it).

The clotting times of blood for plasma diluted with nine different percentage concentrations with prothrombin-free plasma

- `time`: time in seconds to clot;
- `conc`: concentration in percent;
- `lot`: lot number - either one or two.

```
library("faraway")
data(clot)
```

1. Fit an appropriate model to the data. Explain your choice.
2. Write a short conclusion in terms of which variables affect the outcome.