

Generalized Linear Model

Data Analysis Report

Consider the *Feeding.txt* dataset, available on Moodle. This dataset results from a study conducted in a hospital, where 135 expectant mothers were asked how they mainly intended to feed their babies. More specifically, the following variables are available:

- *breast*: main feeding method, either breastfeeding or bottle-feeding (coded as "Breast" and "Bottle", respectively)
- *pregnancy*: pregnancy stage, either beginning or end (coded as "Beginning" and "End", respectively)
- *howfed*: main feeding method that the woman experienced as a baby, either breastfeeding or bottle-feeding (coded as "Breast" and "Bottle", respectively)
- *howfedfr*: main feeding method that the woman's friends used with their babies, either breastfeeding or bottle-feeding (coded as "Breast" and "Bottle", respectively)
- *partner*: woman's relationship status, either with a partner or single (coded as "Partner" and "Single", respectively)
- *smokenow*: whether the woman is currently a smoker or not (coded as "Yes" and "No", respectively)
- *smokebf*: whether the woman ever was a smoker or not (coded as "Yes" and "No", respectively)
- *age*: woman's age (in years)
- *educat*: age at which the woman left full-time education (in years)
- *ethnic*: woman's ethnicity (coded as "White" and "Non-White")

The study's goal is to determine which factors affect the decision to try breastfeeding and to assess the magnitude of these factors' effects, in order to better target women with a lower probability of choosing breastfeeding in promotion campaigns. Please perform a data analysis to address this research question, keeping the following steps of statistical modelling in mind:

- Exploratory data analysis
- Model formulation
- Parameter estimation
- Residuals and model checking
- Inference and interpretation

Do not forget to conclude by giving your answer to the research question!

You must submit by Friday, April 21st (midnight!) via email to pierre-yves.deleamont@unine.ch:

1. A report containing your analysis (in a readable format)
2. An *R* script containing all the relevant code used for the analysis

The report should consist of at most 3 pages of text (12 pt font size) describing the analysis and at most 5 pages of graphs/tables/other relevant *R* outputs.

Note: I will reply to your email to acknowledge the good reception of your work. If you have not received a confirmation three days after submitting, please contact me.