# Exercises 4

1. Consider the dose-response model

$$g(\mu) = \beta_0 + \beta_1 x.$$

Under the hypothesis that the response probability at $x_0$ is equal to $\mu_0$, show that the model reduces to

$$g(\mu) = \beta_0(1 - x/x_0) + g(\pi_0)x/x_0.$$

How would you fit this model using R?

2. This example is from Clayton & Hills (1993, p. 229-232). The data are from a case-control study of BCG vaccination (`http://en.wikipedia.org/wiki/Bacillus_Calmette-Guerin`) and leprosy (this vaccination is for tuberculosis, but is supposed to have a small protective effect on leprosy). The study design is a rather unusual example of a case-control study in which the controls were obtained from a 100% cross-sectional survey of the study base. The data are given below as frequency records:

| Cases | Total | Scar | Age |
|---|---|---|---|
| 1 | 7594 | 0 | 0 |
| 1 | 11720 | 1 | 0 |
| 11 | 7154 | 0 | 1 |
| 14 | 10198 | 1 | 1 |
| 28 | 5639 | 0 | 2 |
| 22 | 7583 | 1 | 2 |
| 16 | 2224 | 0 | 3 |
| 28 | 8145 | 1 | 3 |
| 20 | 2458 | 0 | 4 |
| 19 | 5607 | 1 | 4 |
| 36 | 4392 | 0 | 5 |
| 11 | 1636 | 1 | 5 |
| 47 | 5292 | 0 | 6 |
| 6 | 1240 | 1 | 6 |

The columns are number of leprosy cases, total number of subjects, BCG vaccination scar (1=present, 0=absent) and agegroup (0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34).

Using the `glm` function, fit a logistic regression model to obtain an estimate for the effect of the vaccination on the disease, adjusting for age (with agegroup as a categorical variable). What is the estimated odds ratio of being a case if vaccinated compared to not being vaccinated (adjusted for age) and the 95 % confidence interval? for BCG. Why is logistic model suitable for this situation?

3. (BCG example continued) Is it possible to make conclusions on the effect of the age to the risk of leprosy on the basis of this study? Try two different parameterizations for age, first by setting one of the agegroups as the reference group and second, estimating separate parameter for each age group. Obtain the variance-covariance matrix of the parameter estimates in the two situations. How do the two compare?

4. (BCG example continued) Write an R function which returns the log-likelihood of the above logistic regression model and obtain maximum likelihood estimates for the parameters using the `optim` function of R, utilizing the BFGS hill-climbing algorithm (`http://en.wikipedia.org/wiki/BFGS`). Obtain the numerically differentiated Hessian matrix at the maximum likelihood point and use it for calculating the standard errors of the parameter estimates. Compare the results to those obtained with `glm`.