Assignment 6, ST2304

Problem 1 The data set

```
malaria <- read.table("https://www.math.ntnu.no/~jarlet/statmod/malaria.dat")
```

contains a random sample of 100 children aged 3-15 years from a village in Ghana. The children were followed for a period of 8 months. At the beginning of the study, values of a particular antibody were assessed. Based on observations during the study period, the children were categorized into two groups: individuals with and without symptoms of malaria.

1. Using logistic regression, analyse the risk of malaria with age and $\log_{10}$-transformed antibody level as explanatory variables. Compute the estimated probability of developing malaria for a child of age 15 and with antibody level equal to 1000.

2. Fit a reduced model if any of the explanatory variables have a non-significant effect.

3. Make a plot of the relationship between the probability of developing malaria and the explanatory variable in your selected model. Add the observed data points to the plot.

4. Based on the fitted model, how does the odds of getting malaria change as a result of a ten-fold increase in antibody level, that is, what is the oddsratio associated with such a change in $\log_{10}$ ab?

5. Using the `confint` function, compute confidence intervals for the regression coefficients of the fitted model and for the above odds-ratio.

Problem 2 Suppose that $X_1, X_2, \ldots, X_k$ is multinomially distributed with parameters $p_1, p_2, \ldots, p_k$ and n. Show that

$$D = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \tag{1}$$

is approximately chi-square distributed with $k-1$ degrees of freedom in the case of $k = 2$ (two categories).

Hint: In the case of $k = 2$, p_2 may be substituted by $1 - p_1$ and X_2 by $n - X_1$. Then show that D can be rewritten to

$$\left[\frac{X_1 - np_1}{\sqrt{np_1(1-p_1)}}\right]^2 \tag{2}$$

Why is this statistic approximately chi-square distributed? Recall that a chi square distribution variable can be seen as a sum of squared standard normal variables.