

Logistic regression 3

Here we will examine the Vaccination data set (from Mead (1988) the Design of experiments: statistical principles for practical application). In the study from which the data set was taken a vaccination study was carried out using different vaccination methods. Additionally there are different batches of vaccine and different areas from which the subjects were sampled. The data set is loaded and named `vaccination`. All subjects were tested if the vaccination succeeded (a positive outcome) or not (a negative outcome).

Question 1

First plot the data set. What is the number of subjects that are tested, and how many tested positive and negative? Created a variable `Nneg` that contains the number of patients that tested negative.

Question 2

Estimate a logistic regression model using only the `Method` as predictor. Keep the data in aggregated form. Call the model `glm1`. Also compute odds ratios and their confidence intervals.

Question 3

Now correct for `Area` and `Vaccine.batch`. Call the model `glm2`. Again compute odds ratios and their confidence intervals. Perform a likelihood ratio test on the effect of the method.

Question 4

Now perform a deviance test to check the overall goodness-of-fit of `glm2`.

Question 5

Use the function `influencePlot` to check if there are unusual observations. This function is defined in the `car` package which we have already loaded.