Logistic regression 2

In this exercise we analyze the fracture data with logistic regression in R. This is a study concerning neurological deficit in thoracolumbar burst fractures. Between January 1981 and January 1991, 139 patients with thoracolumbar or lumbar burst fractures were treated in the Department of Orthopaedics and the Department of Neurosurgery of the Academic Hospital Dijkzigt of the Erasmus University of Rotterdam (Fontijne WPJ et al., J.of Bone & Joint Surgery, in press).

The following variables were observed:

- `ID`: Subject number
- `DEF`: the occurrence of neurological deficit ('yes' or 'no')
- `LOC`: the localization of the fracture ('lumbar', 'thoracolumbar' or 'thoracic')
- `STEN`: percentage spinal canal stenosis

Question 1

Display a couple of records from the data set to see what it looks like. You could use the `head` function. Make a contingency table of the relation between the location of the fracture and the occurrence of neurological deficit. Is this relation statistically significant?

Question 2

Next look at the relationship between percentage stenosis and localization, describe and test it. How can you test which groups are significantly different?

Question 3

Next look at the relationship between percentage stenosis and localization, describe and test it. How can you test which groups are significantly different?

Question 4

Now we will use logistic regression to examine the effect of percentage stenosis and localization on the probability of neurological deficit simultaneously. Specify a model with the covariates LOC, STEN and the interaction between these two. Is it justified to fit a model without the interaction term?

Question 5

Now fit a logistic model without the interaction term. Save the predicted probabilities. We will use this variable later. Looking at the results answer the following questions: - Does the percentage stenosis have a significant effect on neurological deficit, adjusted for localization? Describe the estimated effect. - Is localization significantly associated with the probability of neurological deficit (adjusted for percentage stenosis)? What is the p-value? - What are the odds ratios of the localization categories thoracic and thoracolumbar versus lumbar?

Question 6

Now fit the model again after changing the coding of the dummies for LOC such that thoracic is used as reference. In `glm` we can do this by including `c(LOC, base=3)` as a model term (3 because thoracic is the third level of this variable which you can check using the `levels` function). Before you run this model, try to find out what will be the estimates for the Constant and for the dummies. When you have run the new model, check whether the estimates are as expected.

Question 7

Next switch to the 'Deviation coding' contrast for LOC instead, by running the command: `contrasts(fracture$LOC) <- contr.sum(3)` In this case the effect of category 3 of LOC (the last category) is the negative sum of the effects of the categories 1 and 2. Fit the model using this coding and compare the estimated coefficients again.

Question 8

What is the predicted probability of neurological deficit for a patient with a thoracic fracture and a stenosis of 60%? (Using the model from Ex 5.)

Question 9

The functions cooks.distance and dfbetas gives you the Cook's distances and the dfbetas of the observations. Use: `coo_1 <- cooks.distance(glm35)` And try to find the patients with large value for Cook's distance. You may also sort the cases by the variable (in descending order) and print the cases with a high value. head(fracture[order(coo_1,decreasing = TRUE),], 10)

Question 10

Plot the area under the AUC curve for the model you estimated in question 5 (which we have called `glm35`) using the `ROC` function of the `Epi` package. Also compute Nagelkerke's R2 measure. For this you can use the `PseudoR2` function in the `DescTools` package. The packages `Epi` and `DescTools` are already loaded.