Project 1

The data set is based on the study of Beshir et al (2017) in malaria-infected individual from Kenya. The data set contains the following variables:

- HRP2/HRP3 – indicates the presence (1) or the absence of HRP2/HRP3 in the parasites of an infected individual
- RDTpos – indicates whether an infected individual could be diagnosed positive according to a given rapid diagnostic tests (RDT)
- Parasite.Density – indicates the estimated parasite density per mL present in an infected individual.

The objective of this project is to study the impact of the HPR2/HRP2 gene deletions on the performance of the RDT in individuals truly infected with malaria-causing parasites. To address this objective:

1. Perform a descriptive analysis (e.g., summary statistics and plots) for each variable.
2. Estimate the overall sensitivity, the prevalence of HRP2 deletions, and the prevalence of HRP3 deletion. Calculate the respective confidence intervals.
3. Imagine that an epidemiological study was conducted in the same area using the same RDT for malaria parasite detection. It was found 53 infections among 176 individuals. Use the findings from 2. to estimate the true prevalence of infection. For this estimation, assume that the RDT has perfect specificity.
4. Test the association between HRP2 and HRP3. Draw your conclusions.
5. Test the association between HRP2 and RDTpos with an appropriate statistical test. Do the same for for HRP3 versus RDTpos. Draw your conclusions.
6. Construct an appropriate statistical model to evaluate the impact of “HRP2”, “HRP3”, and “Parasite.Density” on “RDTpos”. What is the accuracy of the model constructed and which covariates are statistically significant?
7. Using the estimated model, construct a similar plot (without the boxplots) to Figure 3 of Beshir et al (2017) for the following situations: (i) the presence of both HRP2 and HRP3, the presence of HRP2 but the absence of HRP3, the presence of HRP3 but the absence of HRP2, the absence of both HRP2 and HRP3. Calculate the cutoff in the parasite density that ensure a sensitivity of at least 90% for each of the four situations. Draw your conclusions considering the main objective of the study.

Prepare a presentation with all your findings. Provide the R scripts or the R markdown of the analysis.

Note:
The original data set was modified for the purpose of this project and, therefore, the published results should not be used as guidance.

Reference:
Beshir KB, Sepúlveda N, Bharmal J, Robinson A, Mwanguzi J, Busula AO, de Boer JG, Sutherland C, Cunningham J, Hopkins H. Plasmodium falciparum parasites with histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in two endemic regions of Kenya. Sci Rep. 2017 Nov 7;7(1):14718. doi: 10.1038/s41598-017-15031-2.