Project 3

The data set (data_project_3.csv) is based on the study of Eziefula et al (2014). This study aimed at evaluating the effect of artemether-lumefantrine alone or combined with different primaquine (PQ) doses on the treatment of Ugandan children infected with malaria-causing parasites. The data set contains the following variables:

- Treatment – indicates the PQ dose (0 - artemether-lumefantrine alone, 0.1 mg/kg, 0.4 mg/kg, and 0.75 mg/kg).
- Age – Age in years of the individuals
- Gender – 1 = male, 2 = female
- rs1050828, rs1050829 – two genetic markers located on the GP6D gene.
- Pf.d0 – parasite density per mL at day 0 (participation enrollment)
- Hb_d0 – hemoglobin concentration per g/L at the day 0 (participation enrollment)
- Hb_d7 – hemoglobin concentration per g/L at the day 7

The main objective is to evaluate whether there is an impact of rs1050828, rs1050829, and PQ dose on the hemoglobin concentration in these children. To address this objective:

1. Perform a descriptive analysis (e.g., summary statistics/plots) for each variable.
2. For each genetic marker, retrieve the information about the respective genetic information (chromosome, location, alleles) and the allele distribution of different populations from the 1000 Human Genome Project (1000HGP).
3. Calculate the allele distribution for each genetic marker in male and females separately and evaluate whether these distributions are different between males and females using an appropriate statistical test.
4. Calculate the genotype distribution of each genetic marker in females only. Evaluate whether the observed genotype distribution of each genetic marker agrees with the expectations from the Hardy-Weinberg Equilibrium.
5. Using “Hb_d0” as the outcome variable, construct an appropriate linear regression model that tests whether the genetic markers affect the hemoglobin concentration while adjusting for gender, parasite density at day 0, and age (additional covariates). Validate the model with a residual analysis by testing the normal

distribution of the respective residuals. Use a transformation (see, for example, Box-Cox transformation) of the outcome if needed

6. Repeat the previous analysis using “hb_d7” as the outcome variable but now also adjusting the genetic effects of the genetic markers for gender, quantity of parasites at day of enrollment, age, and PQ dose (covariates).

Important:

Prepare a 15-minute presentation with your main findings. There will be a penalty of 0.5 points in your project grade if you exceed the time of your presentation. Upload your R script/R Markdown for code verification. Also upload your presentation as a pdf file. Failure to upload these files before classroom evaluation leads to a penalty of 0.5 points in your project grade.

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:

Eziefula AC, Bousema T, Yeung S, Kamya M, Owaraganise A, Gabagaya G, Bradley J, Grignard L, Lanke KH, Wanzira H, Mpimbaza A, Nsobya S, White NJ, Webb EL, Staedke SG, Drakeley C. Single dose primaquine for clearance of Plasmodium falciparum gametocytes in children with uncomplicated malaria in Uganda: a randomised, controlled, double-blind, dose-ranging trial. Lancet Infect Dis. 2014 Feb;14(2):130-9. doi: 10.1016/S1473-3099(13)70268-8.