Project 3

The data set is based on the study of Eziefula et al (2014). This study aimed at evaluating the effect of artemether-lumefantrine alone or in combination 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
- Pf.d0 – malaria parasite density (per mL) at the day 0 (participation enrollment)
- Hb_d0 – hemoglobin concentration (per g/L) at the day 0 of the study (participation enrollment)
- Hb_d7 – hemoglobin concentration (per g/L) at the day 7 of the study
- clear.inf.d7 – indicates whether infection is still present at day 7

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

1. Perform a descriptive analysis (e.g., summary statistics/plots) for each variable.
2. Use appropriate statistical tests to evaluate whether the hemoglobin concentration at day 0 varies significantly across children receiving different doses of PQ.
3. Use an appropriate statistical test to evaluate whether the parasite density at day 0 varies significantly across children receiving different doses of PQ.
4. Construct an appropriate linear regression model that quantifies the effect of PQ dose on “Hb_d0” (the outcome variable). Include gender, quantity of parasites at day 0 and age as 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
5. Do the same as point 5 but using “Hb_d0” as the outcome variable. Include gender, parasite density at day 0, age, and parasite as additional covariates.
6. Use an appropriate statistical test to evaluate whether the prevalence of infection at day 7 varies significantly across children receiving different doses of PQ.
7. Construct an appropriate regression model to evaluate the effect of PQ dose on the presence of infection at day 7 (the outcome variable). Include gender, parasite density at day 0 and age as additional covariates. What is the predictive accuracy of the model?

Prepare a presentation with all your findings. Provide the R scripts or 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:
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.