**HW 2: Using models to estimate effects**

**Please read this publication and use the supplemental table provided by the authors:** <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003743#sec019>

**Question 1 : Logistic and Log binomial models**

**1.1** What type of regression model was used to estimate the effect of malaria on nutritional status?

List the independent variables and how they were coded to create the results in Table 3.

What variables from the study dataset were used as outcomes?

**1.2** Estimate the risk of anemia at study exit (outcome: **anemia\_saida**) among children who experienced a malaria infection at any time during the study (primary exposure of interest, malariasn1).

**1.3** How many children had anemia at study exit? How many were exposed to malaria infection during the study (defined by **malariasn1**)?

**1.4** Calculate the crude odds ratio for the association between malaria infection (**malariasn1**) and anemia (**anemia\_saida**) by hand using a 2x2 table:

**1.5** Estimate the crude odds ratio and 95% CI for the association between any malaria infection (**malariasn1**) and anemia at study exit (**anemia\_saida**) in R using glm():

**1.6** Estimate an adjusted odds ratio and 95% CI (for all variables) including these covariates as potential confounders: **anemia\_entrada** (y/n); **sexo** (males as reference); **socioeconomico** (rich/middle/poor):

Did the effect estimate for malaria change much between your crude and adjusted model?

What variables were statistically significant in your adjusted logistic model?

**Question 2 : Use Poisson regression to estimate the incidence of first malaria infection.**

**2.1** The **time\_to\_first\_malaria** variable has the number of days (365 days=no malaria). First create a tabular dataset to transform our individual-level dataset to reflect the count of malaria cases and cumulative exposure time to first malaria by variables of interest. In this example, we will explore **sexo**, **socioeconomico** and **helmintosentrada** status as covariates. (assume **helmintosentrada** = 2 should be coded as “no”).

**2.2** Calculate by hand a crude incidence rate of first malaria infection for the total study population (use **malariasn1** to define malaria outcomes. Report rate as number of cases per person-day.

**2.3** Use glm() to fit a Poisson regression to estimate the incidence rate. This model would have no predictors. What is the incidence rate?

**2.4** Add the covariates **sexo,**  **helmintosentrada** and **socioeconomico** to the model as predictors. Report these as incidence rate ratios with their confidence intervals. Are any of these IRRs statistically significant?

**2.5** Explore what happens when the length of person-time contributed is shortened (differential with respect to a given variable). Reduce the total person time among all subjects for which **helmnitosentrada**=yes by 30%.

What is the new IRR for **helmintosentrada** (include **sexo** in the model as another covariate)?

How did it change and is this what you expected? Please provide a rationale for why this IRR changed in the direction you observed.