

Hw3

1. INFANT MORTALITY: The file *InfantMortality0.csv* has information about the number of infants that die before they are one year old per thousand of births in a country in 1957, *infd*, the number of inhabitants for each medical doctor in that country, *phys*, the number of inhabitants per squared Km, *dens*, the number of inhabitants per thousand agricultural acres, *agds*, the percentage of adults that can read, *lit*, the number of university students per 100.000 inhabitants, *hied*, and the gross national product per capita in that country, *gnp*, all for 49 countries.

Explore the relationship between infant mortality and the other variables, and which explanatory variables are relevant.

2. GALAPAGOS: In *Galapagos0.csv* you can find the number of species and the number of endemic species in 29 islands of the Galapagos. The goal is to relate the biological diversity of an island, measured either through the number of *Species* or the number of *Endemic* species, to its *Area*, its maximum height, *Elevation*, its distance from Santa Cruz which is a measure indicating how peripheral that island is, *DistStCruz*, its distance to the nearest island, *DistNearIsl*, and to the area of the nearest island, *AreaNearestIsl*.

Model separately the number of species and the number of endemic species.

Is there any problem with using normal linear models given the fact that here the response is a count variable?

3. HAMILTON: Build two separate models for Y against $X1$ and Y against $X2$. Can you improve on them? Data can be found in *Hamilton0.csv*.