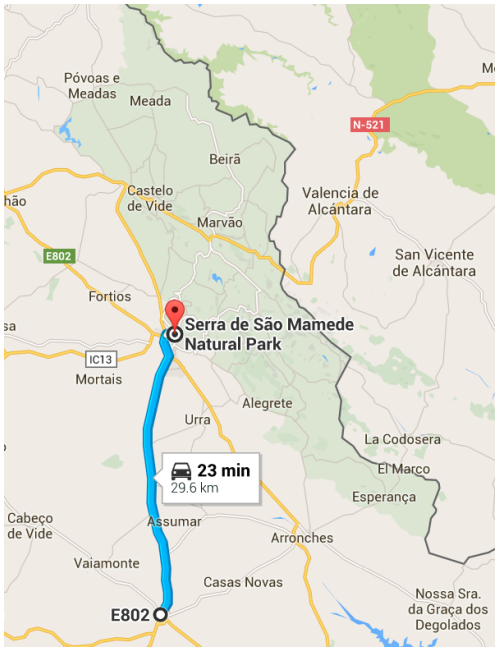
Homework 4
(each question 6 pts)



The file "RoadKills.txt" contains data on amphibian roadkills at 52 sites along a road in southern Portugal (each site is a 500m stretch of road). The dataset includes several amphibian variables, and associated environmental variables that could help explain the occurrence of roadkills (roadkills are more likely if the road crosses a migration route to and from spawning sites, or during juvenile dispersal). We will focus on TOT.N, the total number of roadkills observed during the sampling period; D.PARK, the distance to Serra de São Mamede National Park, which has good amphibian habitat (meters); URBAN, the area of urban habitat surrounding the site (hectares); and D.WAT.RES, the distance to a water reservoir. Let's explore which of these variables are good predictors of the total number of roadkills.

1. Plot TOT.N vs. the other variables I mentioned. URBAN is highly skewed. This creates issues because the few very large values will have a large effect on the fitted relationships. Make a columns where this predictors is square-root transformed, and plot the relationships with the new predictors. We'll discuss transforming predictors more in class.

Fit a Poisson GLM to test if total number of roadkills is predicted by D.PARK. Plot the raw relationship and plot the fitted curve on top of it. Plot residuals vs. the predictor: use both the raw (response) residuals and the deviance residuals. Are there any strong patterns? How do the raw and deviance residuals differ, and why? Do a likelihood ratio test to test for the significance of the predictor.

2. Fit the same model with quasipoisson, and with a negative binomial distribution. Do appropriate hypothesis tests on these models. How big is overdispersion based on the quasipoisson? What is the theta parameter for the negative binomial? Does overdispersion affect the conclusions you would draw from this analysis?

3. Now fit a Poisson model that adds in the other two predictors, URBAN (square-root transformed) and D.WAT.RES. What is effect size of the 3 different predictors, i.e. how much does # roadkills change as these predictors vary? How do residuals vs. fitted values and residuals vs. predictors look (you can just use deviance or pearson residuals, as they are more appropriate for GLMs)? Do appropriate (marginal) likelihood ratio tests for each of the three predictors. Plot the fitted effects using the 'effects' package. How do you interpret these results?

4. Now fit the same model with quasi-Poisson and negative binomial. Plot the fitted effects using the 'effects' package. How similar are the parameter estimates between Poisson, quasi-Poisson, and negative binomial model? How similar are hypothesis test results for the three models? Why do you think these results would differ from what you found for #3?

5. Let's go back to a model with just TOT.N vs D.PARK. Fit a standard linear regression for this relationship, i.e. with normally distributed error. Plot the data and the fitted relationship. Plot residuals vs. fitted values. Based on these plots, do you think this model is a good alternative for this data? Explain why you think yes/no.