3rd assignment
Poisson Regression

Consider the "Air pollution" data set (pollution.dta). Stata dataset has 12 variables all of which have labels (and value labels where necessary) attached. Data have been collected in Athens during a 4 years period. The main aim of this study was to investigate the association between air pollution, using as indicator the SO_2 levels, and daily all-cause mortality.

1. Produce two graphs showing a) Number of deaths over time and b) SO_2 levels over time for the whole time period (4 years). Comment on those two graphs focusing on the seasonality patterns.

2. Fit Poisson models using the daily all-cause mortality as dependent variable and the SO_2 levels as independent variable. Check the dose-response relationship between SO_2 and mortality graphically* and consider an appropriate transformation on the pollutant's levels if necessary

3. Try to control for seasonal trend in mortality using either a 4-level nominal variable for calendar season (or a 12-level nominal variable for calendar month) or a sinusoid curve of 1-year period**. Check also if a long-term trend is present in these data (consider calendar year as a continuous or 4-level nominal variable). Comment on the results.

4. Investigate the relationship between temperature or relative humidity and mortality using graphical methods (see above). Check if these two factors have a significant effect on daily all-cause mortality using Poisson models. Take into account your findings in all previous steps of the analysis.

5. Check also the effect of the "holidays" and "day of the week" variables.

6. Try to adjust your final results for overdispersion (if present in your model).

7. Check the fit of your final model and try to improve it if necessary.

8. Present the results in a tabular form and interpret your main findings

* Consider smoothing functions (i.e. `ksm var1 var2, lowess` for a lowess kernel smoother) or alternatively plot the mean of SO_2 levels and of corresponding number of deaths (log-transformed) for every 20 consecutive days.

** For a sinusoid term generate two new variables:

```
gen sin1=sin(2*3.14*time/365.25)

gen cos1=cos(2*3.14*time/365.25)
```

Both variables should be included in the model.