# Homework Assignment 5

Due Date: Monday, April 08, 2019 at 5P

Total Points: 120

## 1 Case study in Poisson regression: effects of particulate air pollution on daily non-accidental mortality in Chicago

The Chicago mortality data is available through the R package: dlnm.

```
>library("dlnm")
>data(chicagoNMMAPS)
>str(chicagoNMMAPS)
```

More information about the data can be found at the link below
`https://rdrr.io/cran/dlnm/man/chicagoNMMAPS.html` and the paper:

Peng, Roger D.; Welty, Leah J.; and McDermott, Aidan, "The National Morbidity, Mortality, and Air Pollution Study Database in R"(June 2004). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 44. `https://biostats.bepress.com/jhubiostat/paper44`

Use the Chicago mortality data for this problem. The scientific objective is to estimate the relative risk of death for each 10 microgram per cubic meter increase in PM10, controlling for temperature and time trends. There are at least three complications.

1. Mortality depends on temperature and pollution exposures over the last week or more, not just on a single day.
2. Mortality depends on temperature in a non-linear way, with increasing mortality at low and high temperatures.
3. There are time trends in mortality that are unlikely caused by trends in PM10. For example, there is seasonality in mortality caused by influenza and other infectious disease processes and in PM10 because of seasonal power generation and use of vehicles.

To move forward, use current daily temperature and average temperature over the 3 previous days as predictors. For each, allow for possible non-linearity by including linear splines with breaks at 60 and 75 degrees F. Use PM10 on the prior days as the pollution variable. Include dummy variables for year, season, or month as detailed below.

1. Display the daily number of deaths against date, temperature and PM10 (separate plots). Work to make the displays as useful as possible for seeing patterns that your model must take into account. Comment on the relative size of the "effects" of time, temperature and PM10 on mortality, apparent in your figures.

2. Create the following integer time-trend variables: year, season, and month. Each variable should be an integer sequence (0, 1, 2, 3, 4, 5) that increments with each new year, season or month. For example, the month variable starts at 0 for January, 1987, because 1 in February, 1987 and end with 167 in December, 2000.

   Create linear splines (3 df) for the current and prior temperature variables.

3. Us log-linear regression to express expected number of deaths as a function of
   PM10
   Same day temperature (3 df)
   Average of prior 3 days temperature (3 df)
   Indicators of year (13 df)
   Enter the PM10 coefficient and it standard error in the table below. Also enter the Deviance and AIC for this model. Be sure to correct for over-dispersion in your calculation of standard error, CI, Deviance and AIC.

| Time Scale | PM10 coef | Standard error | 95% CI | Model df | Deviance | AIC |
|---|---|---|---|---|---|---|
| Year | | | | | | |
| Season | | | | | | |
| Month | | | | | | |

4. Now repeat the analysis replacing the indicators of year with indicators for season. Repeat the analysis again replacing the indicators for season with indicators for month. Describe how your PM10 relative risk changes as you control for an ever more "wiggly" time trend by including more degrees of freedom for trend. Explain in scientific terms why the changes you see may be occurring.

5. For the model with monthly indicators, check the main assumptions: independence, average Pearson residual equal 0 at each level of predicted value; few highly influential points.

6. Summarize your finding as if for a project for job interview.