AS.280.347 CLASS 2.4

- Review your work!

- Your projects!

Module 2: Particulate air pollution and mortality

- Question 2.1 (Q2.1): How does the daily risk of death depend upon air pollution level in American cities?

- Question 2.2 (Q2.2): Is the estimate of the pollution effect sensitive to assumptions about seasonal or weather effects?

- **Question 2.3 (Q2.3): How do you pool PM effect (log relative rate) estimates from multiple cities taking account of both natural geographic variability in the true effects and statistical errors that might differ among cities?**

- We will answer these questions using data from the National Morbidity and Mortality Air Pollution Study (NMMAPS)

Assignment 2.3

- **For each single city:**
  - Update your time series display of PM10, temperature, and total mortality versus date.
  - Regress mortality on PM10 using a Poisson model with different indicator variables for time: (a) nothing (0 df); (b) season (4-1 df); (c) month (12-1 df); (d) season by year (4x19-1 df); (e) month by year (12x19-1 df)
  - Repeat the regressions with natural splines to give a smooth relationship between mortality and time using ns(time, df) for df = 0, 3, 11, 75, 227
  - Display the 10 PM10 coefficients with confidence intervals in a table or graph to see the effect of the method of control for seasonality.
  - Choose your preferred model from among the 10 above. Explain your choice. **Interpret the log relative rate from this chosen model.**
- **Across the cities:**
  - Using your chosen seasonal adjustment – the same for each city – pool the city-specific estimates of log relative rate from your team to estimate an average value taking account of statistical and city-specific variation in the log relative rate estimates.
  - Write an R function to do the pooling calculation

- Work together in groups!
- Submit assignment in R markdown through Blackboard by **Sunday, March 24th** @ midnight.

	Log RMR (pm10)	p	95% CI
Mortality ~ pm10 (NY)	-0.00002	0.90870	(-0.00038, 0.00034)
Mortality ~ pm10 + season (NY)	0.00119	0.00000	(-0.00061, -0.00008)
Mortality ~ pm10 + month (NY)	0.00123	0.00000	(0.00051, 0.00123)
Mortality ~ pm10 + season x year (NY)	0.00133	0.00000	(0.00082, 0.00155)
Mortality ~ pm10 + month x year (NY)	0.00124	0.00000	(-0.00010, 0.00044)
Mortality ~ pm10 (LA)	-0.00035	0.01004	(-0.00023, 0.00052)
Mortality ~ pm10 + season (LA)	0.00017	0.22028	(0.00085, 0.00161)
Mortality ~ pm10 + month (LA)	0.00032	0.02268	(0.00004, 0.00059)
Mortality ~ pm10 + season x year (LA)	0.00020	0.16701	(-0.00027, 0.00049)
Mortality ~ pm10 + month x year (LA)	0.00038	0.01209	(0.00095, 0.00170)
Mortality ~ pm10 (SEA)	0.00087	0.00000	(-0.00008, 0.00047)
Mortality ~ pm10 + season (SEA)	0.00015	0.43603	(-0.00025, 0.00053)
Mortality ~ pm10 + month (SEA)	0.00011	0.58705	(0.00082, 0.00166)
Mortality ~ pm10 + season x year (SEA)	0.00014	0.46830	(0.00008, 0.00068)
Mortality ~ pm10 + month x year (SEA)	0.00015	0.48014	(-0.00026, 0.00055)

AIC Values for Model Generated

	Cities		
	NY	**LA**	**SEA**
Model A	12319.057	12949.97	27555.22
Model B	11544.072	12034.80	27173.90
Model C	11242.832	11557.80	27117.83
Model D	9842.436	11834.77	26961.87
Model E	9483.671	11351.71	26982.88
Model A, NS	12319.057	12949.97	27555.22
Model B, NS	10749.964	12851.35	27348.83
Model C, NS	10674.878	12831.17	27346.31
Model D, NS	9606.354	11464.54	26899.31
Model E, NS	9468.832	11252.82	26955.88

Pooling the City-Specific Estimates

City	Estimate of Log Rel Rate	SE $_{Estimate}$	Total Variance $(V=t^2+ SE^2)$	V^{-1}	V^{-1}/Sum(V^{-1})	Weighted Average

$$\hat{\beta}_i = \beta_0 + \eta + \varepsilon_i$$

$$Var(\hat{\beta}_i) = Var(\beta_0 + \eta + \varepsilon_i) = Var(\eta) + Var(\varepsilon_i) = \tau^2 + SE^2$$

$$\hat{\tau}^2 = Var(\hat{\beta}_i) - Average(SE^2)$$

Assignment 2.4

Finalize your report for Module 2 to answer Q2.1 – Q2.3.

- ***For each question, you should have a data display and a statistical analysis to address the question.***

- ***Provide a caption for your data display(s).***
  ***(http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtablefigs.html*)**

- ***Write up your results in a few paragraphs to answer these questions. In your write-up, you should refer to your data display(s) and your analysis results. Be numerate!***

- Work together in groups!
- Submit your assignment in R markdown through Blackboard by Sunday @ midnight.

Your Project Design

- Question:

- Data set and design
    - Outcome:
    - Predictor variables of primary interest:

    - Effect modifiers:

    - Confounders:

- DAG:

- Primary analysis to address question:

- Communicating results in tables, figures

Your project!

- Question:

- Data set and design
  - Outcome:

  - Predictor variables of primary interest:

  - Effect modifiers:

  - Confounders:

- Directed Acyclic Graph (DAG):

- Primary analysis to address question:

- Communicating results in tables and figures:

Your project!

For your final project, you will need:

- A research question of interest in public health
- A data source that you can use to answer this question

Framing a research question in public health:

- Start with a <u>general</u> area of public health in which you have interest, and then narrow to a <u>specific</u> question you'd like to answer.
- It can be helpful to frame your question in terms of investigating a relationship between a specific outcome variable (like "disease status" for our Module 1) and one or more primary predictor variables ("smoking status" for our Module 1.)
- Later you will need to think about the possibility of effect modifiers and possible confounders, but for now just think about that primary relationship of interest!

Locating data to answer this question:

- If you have a specific area of interest in mind, you can Google for data in that area
- Or explore the links below to see what type of data is available:
  https://www.healthdata.gov/
  http://guides.lib.berkeley.edu/publichealth/healthstatistics/rawdata
  http://www.datasciencecentral.com/profiles/blogs/10-great-healthcare-data-sets
  https://www.cdc.gov/nchs/data_access/ftp_data.htm
  https://catalog.data.gov/dataset?_organization_limit=0&organization=hhs-gov#topic=health_navigation

Types of regression analysis

- Linear regression

- Logistic regression

- Poisson regression (log-linear model)

Assignment 3.1

- Write a short introduction to your question of interest:
  - Question
  - Data source
  - Outcome variable
  - Primary predictor variable(s)
- If you have it available, read your data into Rstudio
- Examine and explore your data:
  - Summaries of your variables of interest
    - Is there missing data? Anything unusual or concerning?
  - Recode from numbers to factors
    - 1 -> "female", 0 -> "male", etc
  - Make a few basic exploratory plots to answer your question
- What type of basic analysis could you use to address your question?

Submit assignment in R markdown through Blackboard by Sunday @ midnight.