Assignment #1 for Public Health Surveillance

**Due by 5pm February 21 (Monday)**. Submit your Word document (not .Rmd or .R files) via Canvas. **You should work on this assignment alone**. Feel free to consult the internet and class notes; do **not** discuss with your classmates. Any questions should be e-mailed to me or Jiye ([daniel.weinberger@yale.edu](mailto:daniel.weinberger@yale.edu); jiye.kwon@yale.edu), and I will respond to the whole class (via Canvas) for any issues requiring clarification. **Your answers should be brief** (1-2 sentences per question) and all answers should be typed in this Word document. This is scored out of 100 points.

Use the .Rmd file “assignment 1 hist limits and cusum.Rmd”

Part 1: Aberration detection in HIV hospitalizations (50 points)

Your goal is to set up an aberration detection system for hospitalizations associated with HIV in Chile. You have been provided with a subset of the larger database that includes ICD10 codes B20 and Z21, which are codes for HIV/AIDS-related causes of hospitalization.

1. Using an online ICD10 dictionary, what are the top 3 diagnosis codes used for HIV/AIDS in this database?
2. Using the historical limits method to detect aberrations in 2007 and 2008, do you see any unusual activity?
   1. Is there anything suspicious about the date in which the spike in observed cases is occurring?
   2. What is it about the analysis that is causing the alarm to signal for several weeks in a row in early 2008?
   3. The default settings evaluate 1 4-week period on either side of the current 4 week period (window size of 3) in each of the previous 5 years. How does the threshold change if you instead include 2 or 3 4-week periods on either side (window size of 5 or 7)? What would be an advantage or disadvantage of using a wider window with these data?
3. Repeating the analysis of 2007-2008 with Farrington,
   1. Is there a difference in the weeks flagged as unusual compared with the historical limits analysis?
   2. What would be the advantages and disadvantages of increasing the number of time points used to calculate the baseline (b) in any type of aberration detection analysis?
4. Now repeat both analyses but look for aberrations in 2011.
   1. Which method flags the most epidemics?
   2. Which method is most reliable, and why?

PART 2: Temporal aberration detection (50 pts)

It is 2016, and Zika has been sweeping across Latin America. The major concern with this epidemic is that pregnant mothers who are infected can give birth to babies that suffer from birth defects, most notably microcephaly. There haven’t been any cases of Zika in Chile, but as a precaution, you have been tasked with setting up surveillance for microcephaly in the country using the national hospitalization register. Microcephaly is coded as Q02 in ICD10. Using historical data from this database, you will determine which aberration detection algorithm would be most useful and suggest ways to fine tune this algorithm.

1. How many Q02 diagnoses are there in total? Do you expect this to be an accurate representation of the incidence of microcephaly?
2. There are some hospitalizations with a code of Q02 but where the age is >12 months. Should these children count for our case definition? Why?
3. There are a couple sets of analyses presented in the .Rmd file (historical limits, CUSUM/glrpois).
   1. Which of these algorithms do you want to select for your microcephaly aberration detection system?
   2. What are the strengths of this approach compared to the others, and what are the limitations of this approach?
4. On your chosen method, modify the parameters to fine tune the sensitivity of the system (ie so that it doesn’t alarm too often).
   1. Copy and paste the plot showing the observed values with the threshold and alarmed values in red.
   2. Explain which parameters you modified and justify your choices.
   3. What’s your best guess for how often might you expect to see a false alarm with these parameters?