Homework 5

Question 1 (60 points)

A novel pandemic influenza virus has just started to spread in the population. No individuals have antibodies against the new virus at the start of this epidemic. A serial cross-sectional serologic surveillance system has been set up to estimate the number of infections at 30 and 60 days after the epidemic has started. Reliable research studies show that 90% of individuals infected by this new pandemic virus become seropositive. Ignore the delay between infection and seropositivity. The seroprevalence data are as follows:

Day	Number of subjects tested	Number of subjects seropositive
30	195	30
60	175	86

a) Read the documentation of Binomial Distribution in R. Use the binomial distribution to provide a maximum likelihood estimate and the associated confidence intervals for seroprevalence on days 30 and 60.

Day	Seroprevalence	95% confidence interval
30	0.15	0.11, 0.21
60	0.49	0.42, 0.57

For the binomial distribution, there actually exists a closed-form analytic solution to the maximum likelihood estimate of θ, namely the number of subjects seropositive divided by the number of total subjects tested, or in this case, 30/195 = 0.154.

Or you may just use any mle calculator available on the web. For example, the link below computes exact confidence intervals for the binomial distribution.
http://statpages.info/confint.html

In R, we can use '*optim*' function to find the value of θ that maximizes the likelihood of the given data. We can compute the likelihood as:

$$\mathcal{L}(\theta|y) = p(y|\theta)$$

and for the binomial distribution of the seroporevalence,

$$p(seropositive|\theta) = \theta$$
$$p(seronegative|\theta) = (1-\theta)$$

The likelihood of the entire dataset is the product of the likelihoods of each individual data point in the dataset. For the ease of computation, we take the logarithm to convert the products into sums, and so instead of multiplying the likelihood of each observation in the data set, we add the log-likelihood of each data point. As optimizers like to minimize things and not maximize things, we multiply the sum of log-likelihoods by -1 to get the negative log likelihood ("NLL") of the dataset:

$$NLL = -\sum_{i=1}^{N} lnp(y_i|\theta)$$

See the R code for more details.

b) Use your answer in question 1(a) to estimate the (cumulative) infection attack rate on days 30 and 60.

Since 90% of infected individuals become seropositive, the total infection attack rate will be:

cIAR at Day 30 = 0.15/0.9 = 0.17
cIAR at Day 60 = 0.49/0.9 = 0.55

Day	Seroprevalence	95% confidence interval
30	0.17	0.12, 0.23
60	0.55	0.47, 0.62

c) Estimate the infection attack rate between days 30 and 60.

cIAR at Day 60 - cIAR at Day 30 = 0.49/0.9 - 0.15/0.9 = 0.38

Question 2 (40 points)

Read Riccardo et al Eurosurveillance 2011 and download the spreadsheet for this question.

a) Modify the algorithm in Riccardo et al by using 4-weekly moving averages and 95% confidence intervals as the threshold for triggering alerts. Describe and explain how this modified algorithm triggers syndromic surveillance alerts.

In the paper (Riccardo et al., 2011), the following algorithm was used:

1) The expected incidence (u) on day t was the moving average of the observed incidence (w) in the past 7 days, i.e.

$$u(t) = \frac{\sum_{i-7}^{i-1} w(i)}{7}$$

2) The threshold was the 99% CI of the observed incidence on day t, assuming that it follows Poisson distribution.
3) When the expected incidence $u(t)$ was below the threshold calculated by the observed incidence $w(t)$, an alert was issued.
4) Whenever alerts were issued on at least two consecutive days, an alarm was defined.

For the homework question, we will modify the algorithm as follows:

1) The expected incidence (u) on day t will be the moving average of the observed incidence (w) in the past 4 weeks, i.e.

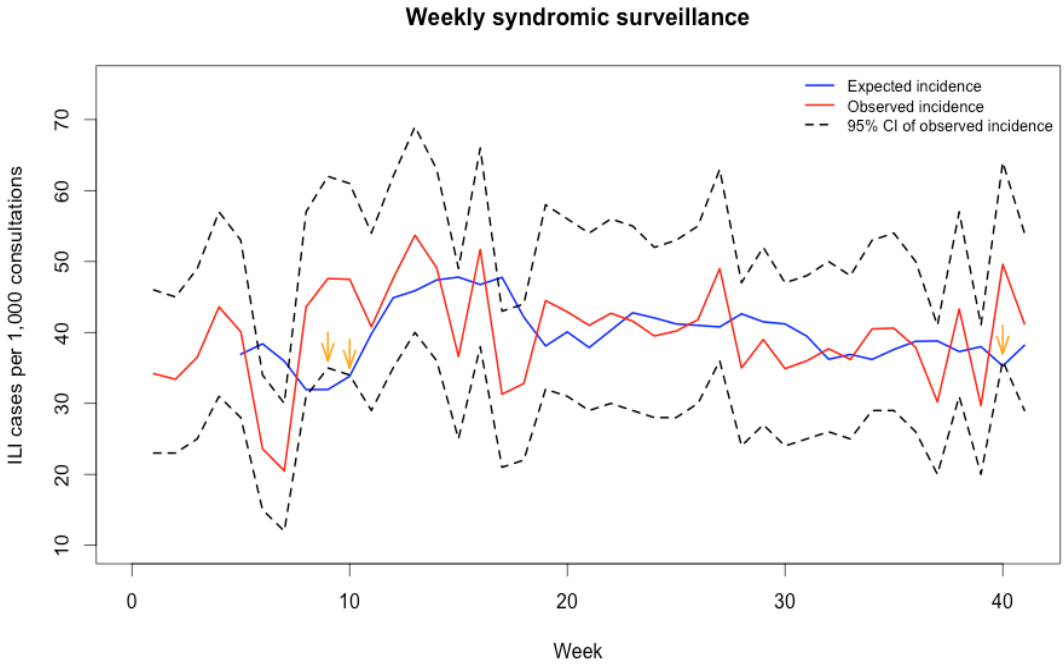
$$u(t) = \frac{\sum_{i-4}^{i-1} w(i)}{4}$$

2) The threshold will be the 95% CI of the observed incidence on day t, assuming that it follows Poisson distribution.
3) As done in the paper, when the expected incidence $u(t)$ is below the threshold calculated by the observed incidence $w(t)$, an alert should be issued.
4) Whenever alerts are issued on at least two consecutive days, an alarm will be issued.
5) With the 95% CIs, the threshold interval would become narrower and alerts (and alarms) would be issued more easily. Therefore, sensitivity would increase while specificity would decrease.

b) The spreadsheet contains weekly influenza-like-illness (ILI) surveillance data for a hypothetical population. Use the algorithm in question 2(a) to calculate the expected incidence and 95% confidence interval of the observed incidence and enter these

numbers into columns 3-5 of the spreadsheet, and submit the spreadsheet (in Excel) as well.
See the spreadsheet on page 4, and the R code

c) Plot the expected incidence, the observed incidence and the 95% confidence intervals of the observed incidence over time in the same figure.



d) In which weeks would an 'alert' be issued based on this syndromic surveillance algorithm?
Alert would be issued in week 9, 10, and 40. Alarm will be issued in week 10. See the next page for the table.

Question 2 (b). Observed incidence, expected incidence, and 95% confidence intervals of the observed incidence.

Week	Observed	Expected	95% Lower Limit	95% Upper Limit	Alert/Alarm
1	34.2	NA	23	46	
2	33.4	NA	23	45	
3	36.5	NA	25	49	
4	43.6	NA	31	57	
5	40.1	36.9	28	53	
6	23.6	38.4	15	34	
7	20.5	36.0	12	30	
8	43.6	32.0	31	57	
9*	47.6	32.0	35	62	Alert
10*	47.5	33.8	34	61	Alert - Alarm
11	40.8	39.8	29	54	
12	47.6	44.9	35	62	
13	53.7	45.9	40	69	
14	49.1	47.4	36	63	
15	36.6	47.8	25	49	
16	51.7	46.8	38	66	
17	31.3	47.8	21	43	
18	32.8	42.2	22	44	
19	44.5	38.1	32	58	
20	42.9	40.1	31	56	
21	41	37.9	29	54	
22	42.7	40.3	30	56	
23	41.6	42.8	29	55	
24	39.5	42.1	28	52	
25	40.2	41.2	28	53	
26	41.8	41.0	30	55	
27	49	40.8	36	63	
28	35	42.6	24	47	
29	39	41.5	27	52	
30	34.9	41.2	24	47	
31	36	39.5	25	48	
32	37.7	36.2	26	50	
33	36.2	36.9	25	48	
34	40.5	36.2	29	53	
35	40.6	37.6	29	54	
36	37.9	38.8	26	50	
37	30.2	38.8	20	41	
38	43.3	37.3	31	57	
39	29.7	38.0	20	41	
40*	49.6	35.3	36	64	Alert
41	41.2	38.2	29	54	