DATA 252 / DATA 551: Homework 6

- This homework is due by Monday 23, 2020 at the beginning of class. **You need to submit your answers on Moodle in a pdf document**. In addition, there will be a short quiz at the beginning of class on Monday 23, which might contain contents from this homework, including the assigned reading, in addition to contents from the most recent lecture.

1. In class, we used the `nCoV simple` dataset and fit normal, lognormal and gamma models. Fit a Weibull model to the same dataset and answer the following questions.

   (1) What are the parameter estimates of the Weibull model?

   (2) What is the log likelihood of the Weibull model? Also, write down the log likelihood of the normal, lognormal, and gamma models. Based on the log likelihood, which model is the best fit?

   (3) Using the Weibull model and the estimated parameters, what is the probability that the incubation period is shorter than 14 days? What is the probability that the incubation period is longer than 5 days?

   (4) Provide your code here (either by copy-pasting or using a screenshot).

2. Read the incubation period paper (posted on Moodle) and answer the following questions.

   (1) Based on results of this paper, what is the estimated median incubation period? Why do you think the the paper uses the median instead of the mean?

   (2) What is the sample size of the dataset used? Discuss at least one reason that the dataset might be biased and not a representative sample.

   (3) For the main results of the paper, what distribution is used to model the incubation period? Name another model mentioned as a comparison to the main model.

3. All of the data and code used in the paper are published online. On Moodle, you can find `nCoV-IDD-data-dictionary` and `nCoV-IDD-traveler-data` (both are csv files that you should be able to open in excel). Play with these two files and answer the following questions.

   (1) Write down what these variable names stand for: EL, ER, SL, SR, PL, PR.

   (2) Find the case whose UID is U0021. When did he arrive in Wuhan? When did he leave Wuhan? If you had to make one estimate of his incubation period, what would it be? Briefly explain.

   (3) Find the case whose UID is U0001. Why would it be very hard to determine this person's incubation period?

   (4) Find another case whose incubation period would be hard to determine. Discuss why.