

## R Notebook

```
library(MASS)

# Loading Data
covid_data = read.csv("./nCoV_simple.csv")$days

# Function to fit dist and return parameter values
fit_distribution <- function(dist) {

  parameters = fitdistr(covid_data,dist)
  return(parameters)
}

# Function for Log_Likelihood
calculate_log_likelihood <- function(parameters,dist){
  p1 = as.numeric(parameters$estimate[1])
  p2 = as.numeric(parameters$estimate[2])
  print(dist)
  log(prod(get(dist)(covid_data,p1, p2)))
}
```

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?

```
parameters_ = fit_distribution("Weibull")
parameters_

##      shape      scale
## 1.8192496 4.6111525
## (0.1992127) (0.3768674)
```

(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?

```
calculate_log_likelihood(parameters = parameters_, "dweibull")

## [1] "dweibull"

## [1] -109.5789

# Normal
parameters_n = fit_distribution("normal")
calculate_log_likelihood(parameters = parameters_n, "dnorm")

## [1] "dnorm"
```

```
## [1] -112.9528

# Log Normal
parameters_ln = fit_distribution("lognormal")
calculate_log_likelihood(parameters = parameters_ln, "dlnorm")

## [1] "dlnorm"

## [1] -115.7273

# gamma
parameters_g = fit_distribution("gamma")

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

calculate_log_likelihood(parameters = parameters_g, "dgamma")

## [1] "dgamma"

## [1] -110.5826
```

Weibull fits better.

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

```
shape_w = as.numeric(parameters_$estimate[1])
scale_w = as.numeric(parameters_$estimate[2])
pweibull(14, shape_w, scale_w)

## [1] 0.9994694
```

What is the probability that the incubation period is longer than 5 days?

```
1-dweibull(5, shape_w, scale_w)

## [1] 0.8676643
```

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 paper uses the median instead of the mean?

- a) 5.1 days (95% CI, 4.5 to 5.8 days)
- b) 11.5 days (97.5% CI, 8.2 to 15.6 days)

Median because mean can be effected by outliers.

(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.

181 Cases sample. a) Publicly reported cases may overrepresent severe cases, the incubation period for which may differ from that of mild cases. b) Small sample size

- (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.

log-normal distribution is assumed based of previous similar diseases. Further more, compared with gamma, weibull and erlang distributions.

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.

```
# Loading Data
covid_data = read.csv("./nCoV_simple.csv")$days
```

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

EL => exposure left bound ER => exposure right bound SL => symptom (any) onset left bound SR => (YYYY-MM-DD) symptom (any) onset right bound PL => presentation of case to hospital left bound PR => presentation of case to hospital right bound

- (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.

1/10/2020 0:00 to 1/18/2020 23:59 About 8 days of stay.

The suspect traveled from Wuhan to Hongkon. it could have longer incubation period as compared to mainland china. As the person may have left without any symptoms Or may have not reported the symptoms that could be a possible case. In gernerl, according to report people from outside china are expected to be have little longer incubation period.

From my observarion, it clearly more then 8 days assuming that he had no sympthoms while the person was in wuahan.

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

This person travel Wuhan to Washington Snohomish County at 1/15/2020 11:59:00 PM. Assuming that he had no symptoms at the time of travel. We have no data when he arrived in wuhan. Because of that we cannot bound the possible time period of exposure in this case.

As stated in the report: "We have used conservative bounds of possible exposure and symptom onset where exact times were not known, but there may be further inaccuracy in these data that we have not considered"

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

U0003, U0004 and many more cases like this dont have left bound for potential exposure. Alternative appoches are used in these cases for exposure esitimation. Because of the same reason as mentioned in above question.

Other possible scenario could be, the people with high incubation period transmit the virus to the others, and we do not even know when exactly that happens. As the first person does not show any symptoms in the first place. This makes estimations harder.