Homework 2: Solution

**1. Data from Figure 1 of Fraser *NEJM* 1977**

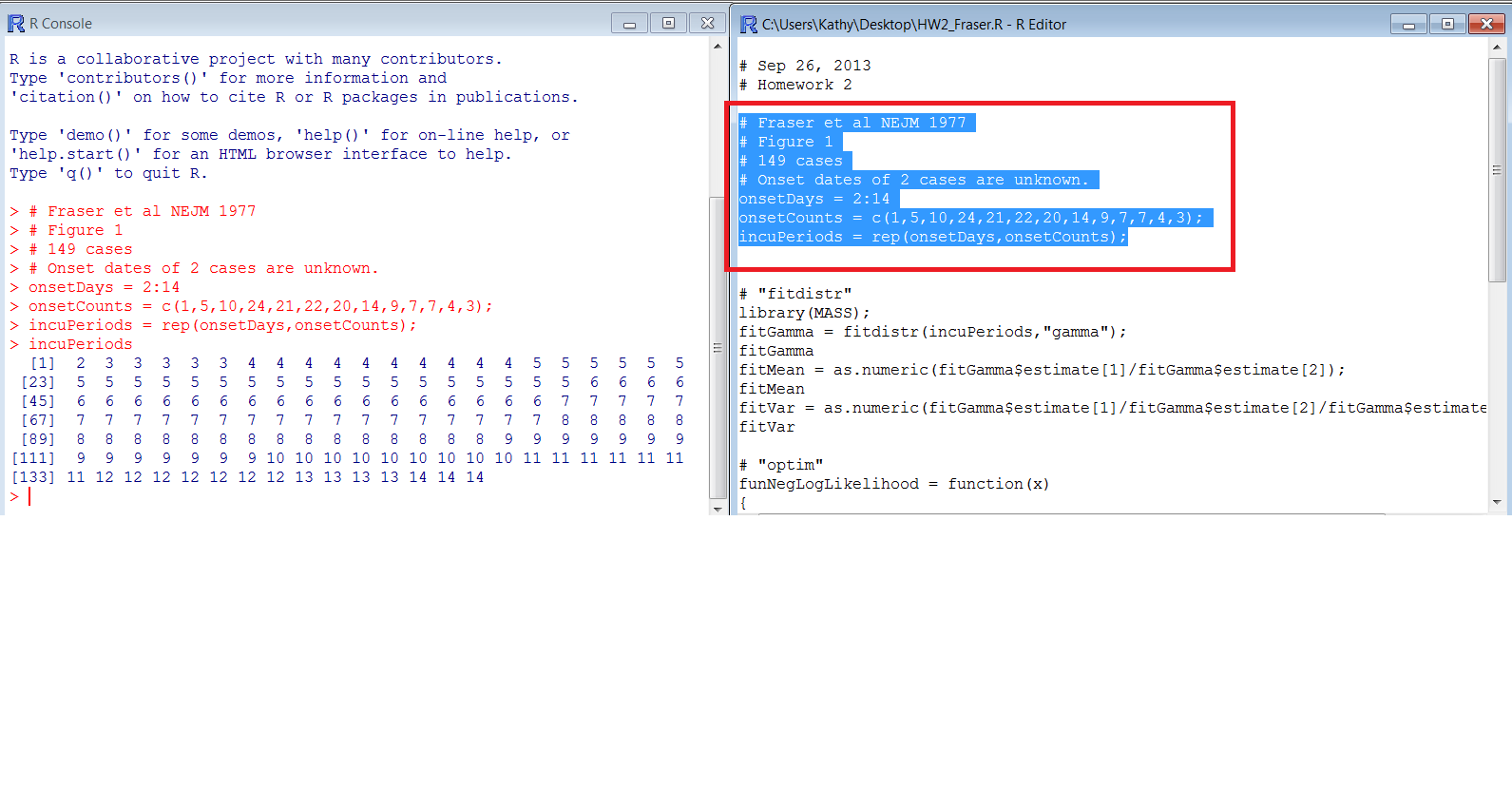
There were 149 cases among persons at the American Legion Convention during July 21-24 in 1976. But onset days of 2 cases were unknown. Assuming all cases were infected on July 20, 1976, we will use the information about the 147 cases shown in the following table to do the analysis.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Days from infection to onset | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Number of individuals | 1 | 5 | 10 | 24 | 21 | 22 | 20 | 14 | 9 | 7 | 7 | 4 | 3 |



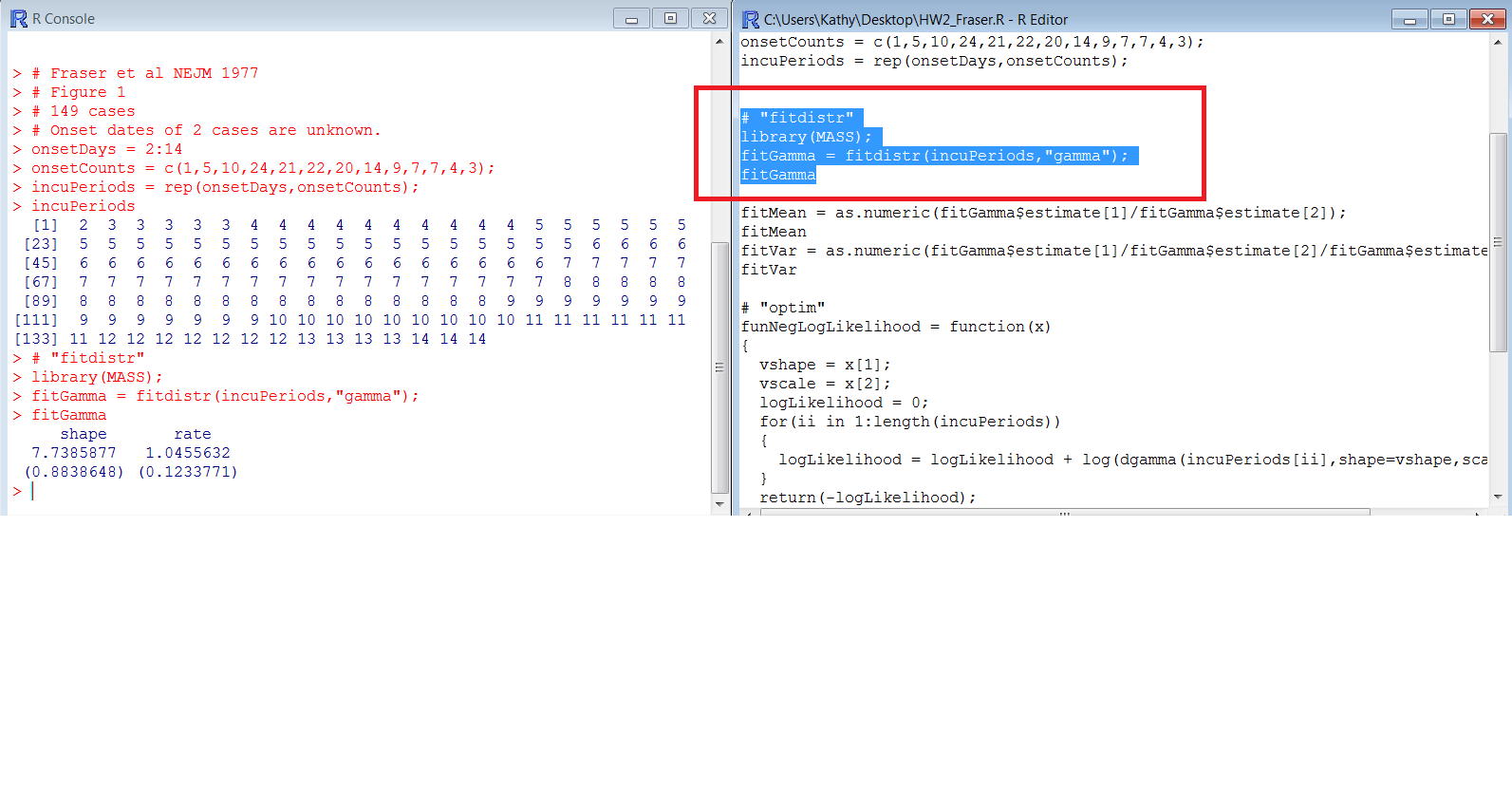
**2. Input data to R**

Use the function “rep” to input data in the above table to R. You can type “length(incuPeriods)” in the console window to check whether the number of cases is correct. In this case, the number of cases should be 147.



**METHOD 1: use the function “fitdistr” to estimate the incubation period distribution**

Since the function “fitdistr” belongs to the library “MASS”, we need to load the library before calling “fitdistr” by “library(MASS)”. Assuming the incubation period was gamma distributed, we specified “gamma” in the arguments passed to the function.



The shape and rate parameter of the estimated gamma distribution are 7.74 and 1.05 respectively. The mean and variance of the estimated distribution are **7.40 days and 7.08 days2** respectively. To know more about gamma distribution, please click: <http://en.wikipedia.org/wiki/Gamma_distribution>

To find the 95th percentile of the resulting best-fit incubation period distribution, you can use

qgamma(p, shape, rate) function. When you input qgamma(.95, fitGamma$estimate[1], fitGamma$estimate[2]), you should have **12.25** for the 95th percentile.

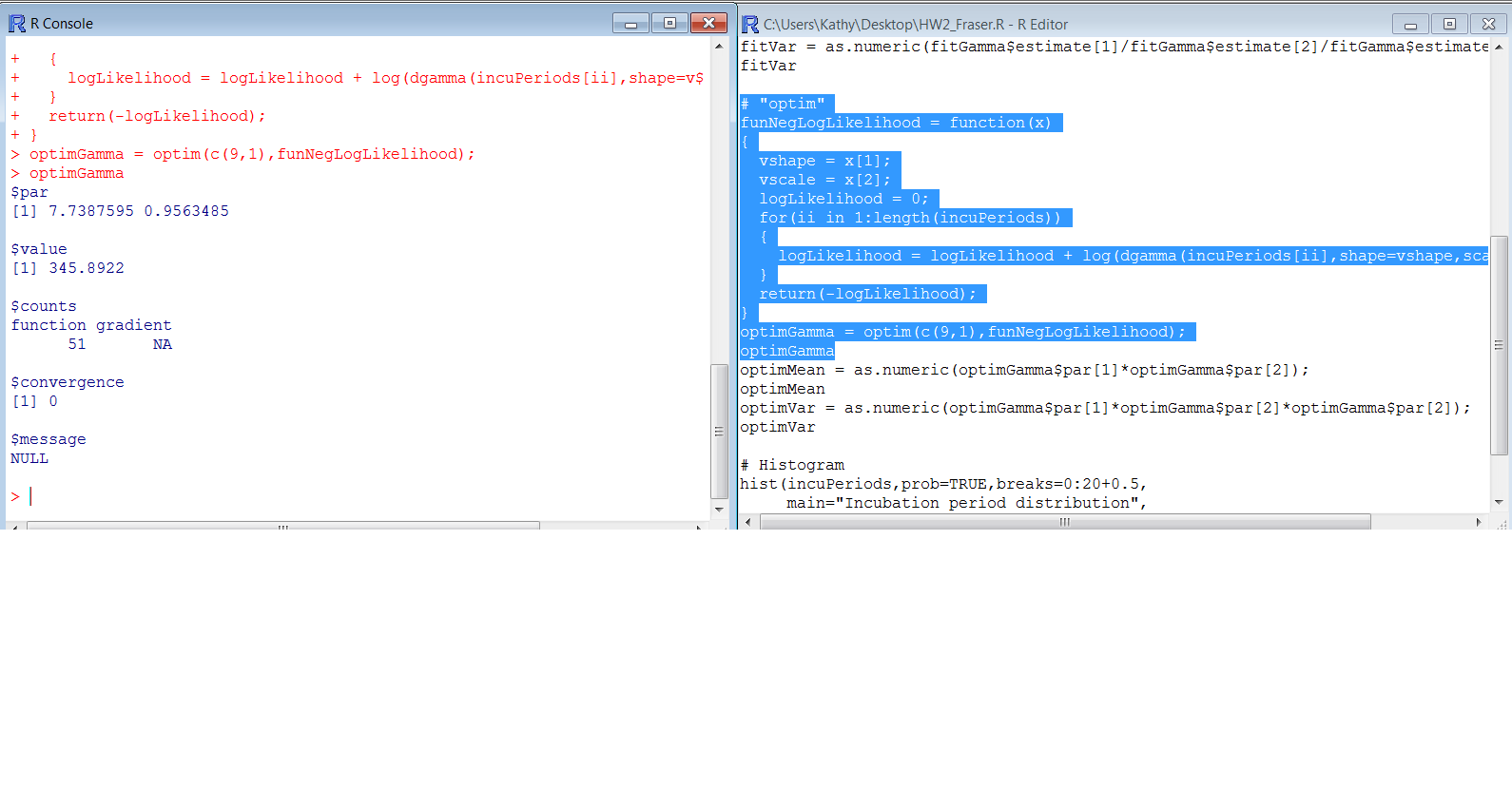
A histogram with fitted curve can be obtained by:

hist(incuPeriods, prob=TRUE, breaks=0:15+0.5, main="Incubation period distribution", xlab="Incubation period", ylab="Probability")

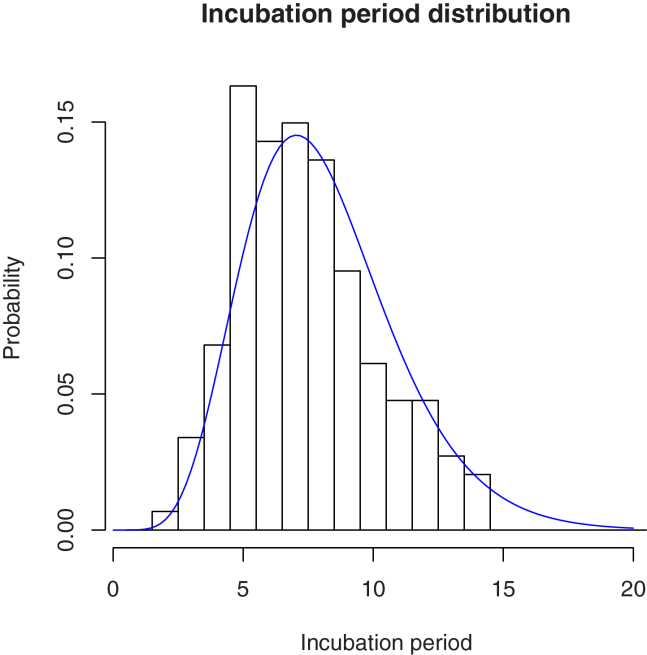
curve(dgamma(x,fitGamma$estimate[1],fitGamma$estimate[2]),add=TRUE,col="blue")

**METHOD 2: define the likelihood function and use “optim” to estimate the incubation period distribution**

Assuming that the incubation period is gamma distributed, we define our own function “funNegLogLikelihood” using gamma density function “dgamma”. We then use the function “optim” to minimize the target function to estimate the incubation period distribution. The estimated shape and scale parameter are 7.74 and 0.96 respectively. The mean and variance of the estimated distribution are 7.40 days and 7.08 days2 respectively too.

****

**5. Plotting the fitted distribution against the data**

****