Answer 1

The following assumptions can be made for the least square method, to find the best value for the parameter speed that best fits the real world data

Y_i: It is the actual covid-19 infection data/ data points in Portugal, which is a dependent variable.

X_i: It is the i^{th} day number that the model takes as an input, which is an independent variable.

f: It is the model that has been coded in netlogo.

β: It is a 4 valued vector, consisting of parameters density, infectioness , mortality and speed

m : It is the number of adjustable parameters in β, which is equal to 4.

i : It represents the number of datapoint pairs of xi and yi. In this case, it goes from 1 to 78 for the real datavalues. However, for some speeds, the number of data pairs available were either greater than or less than 78. Therefore, for those speed cases, I have not assumed any value, and marked it as “data not available” entry in the excel file.
In any case, the value of “i” that i have taken varies from 1 to minimum of entries in real data and the experimental data for different speeds.

Answer 2

Applying the least square strategy described in the question, we get the following value of squared sum of r_i :

Note:

$$S = \sum_{i=1}^{n} r_i^2$$

for different speed values.

S (speed = 1)	**S(speed=1.5)**	**S(speed=2)**	**S(speed=2.5)**	**S(speed=3)**	**S(speed=3.5)**	**S(speed=4)**
8699898.061	5000994.42	3888461.852	4508954.97	5506604.99	6691923.218	7484988.665

Therefore, it can be clearly seen that for speed = 2, the least squared error is the minimum, and therefore, speed = 2 is the value of speed that best fits the real-world data.