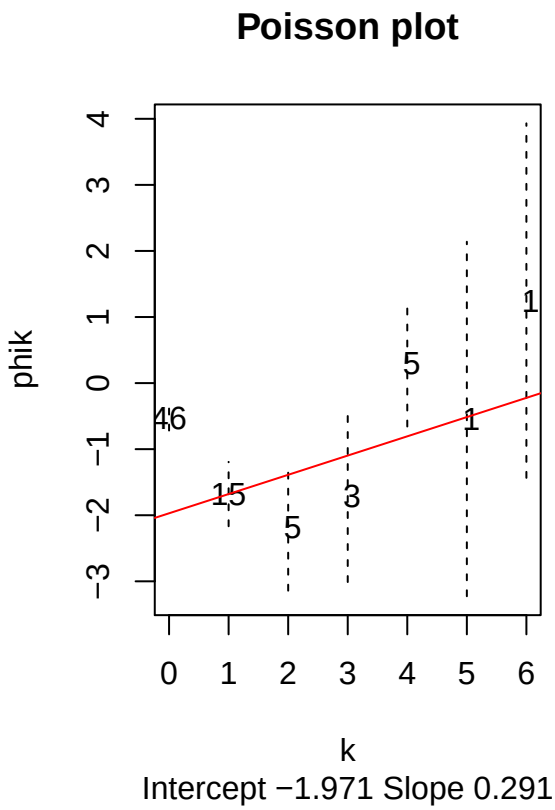
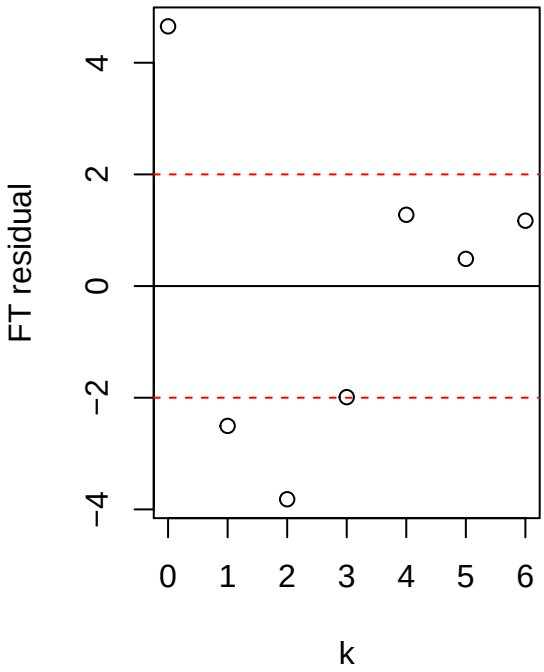
EDA Oct 31

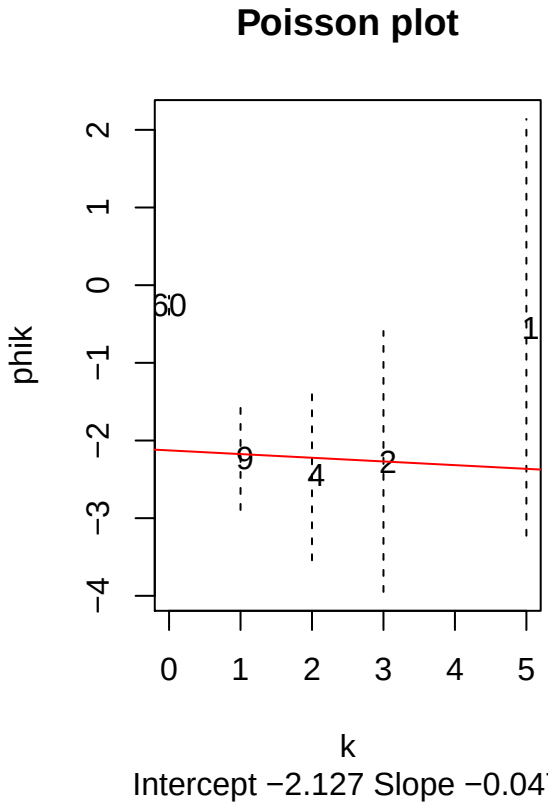
Ex1

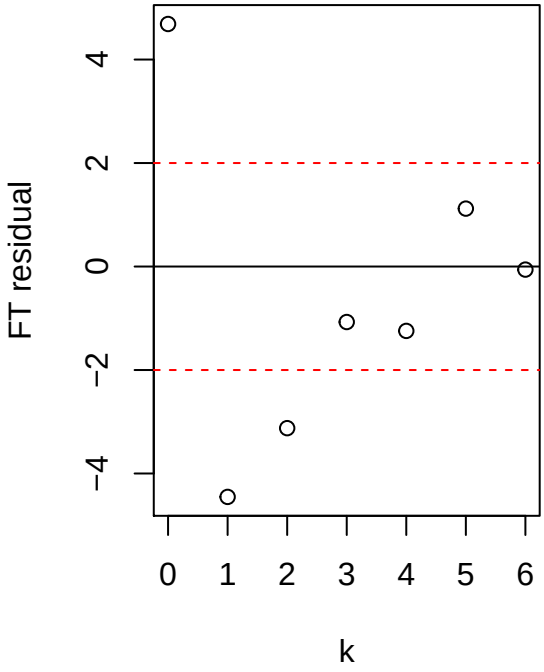
Jenkins & Johnson (RAND report) lists 65 incidents of international terrorism in Argentina and 28 in Turkey during the 76 months between Jan 1968 and Apr 1974. From each frequency distribution, calculate nk* and make the Poissonness plots. Does a negative binomial model seem to give a better fit? If so, choose a value of n for the negative binomial distribution and examine the plots.



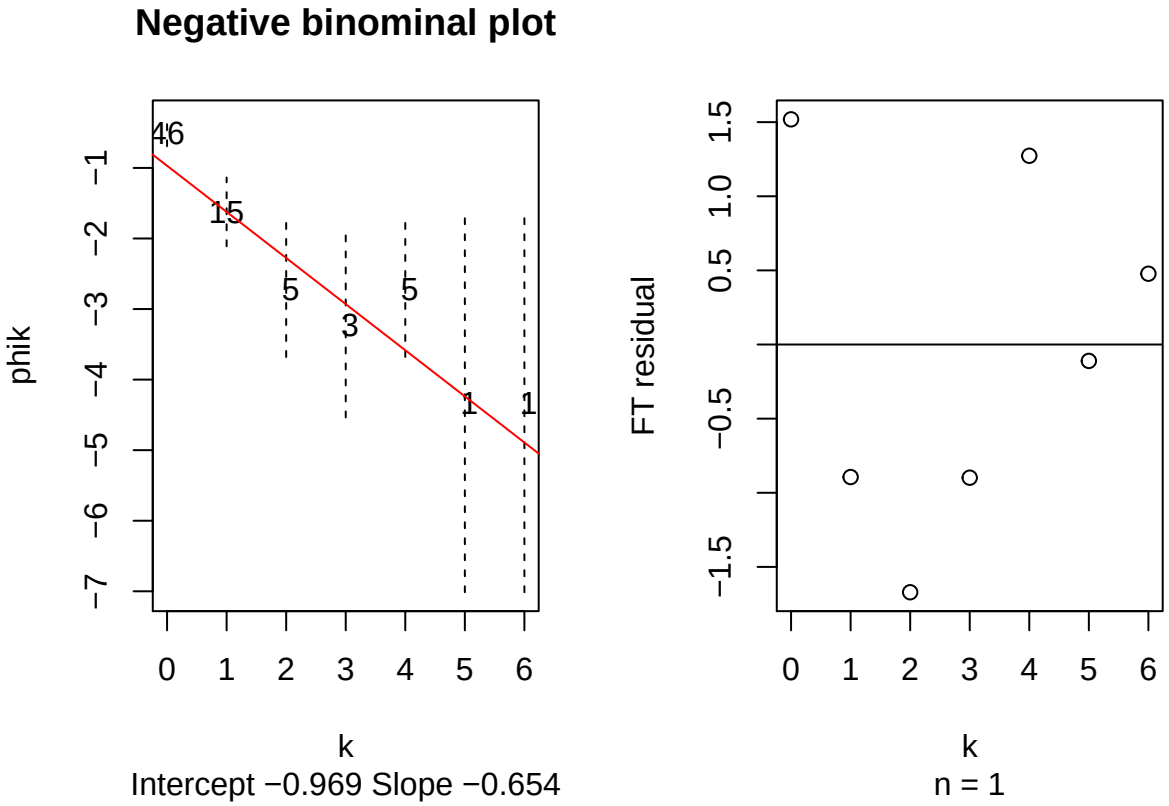


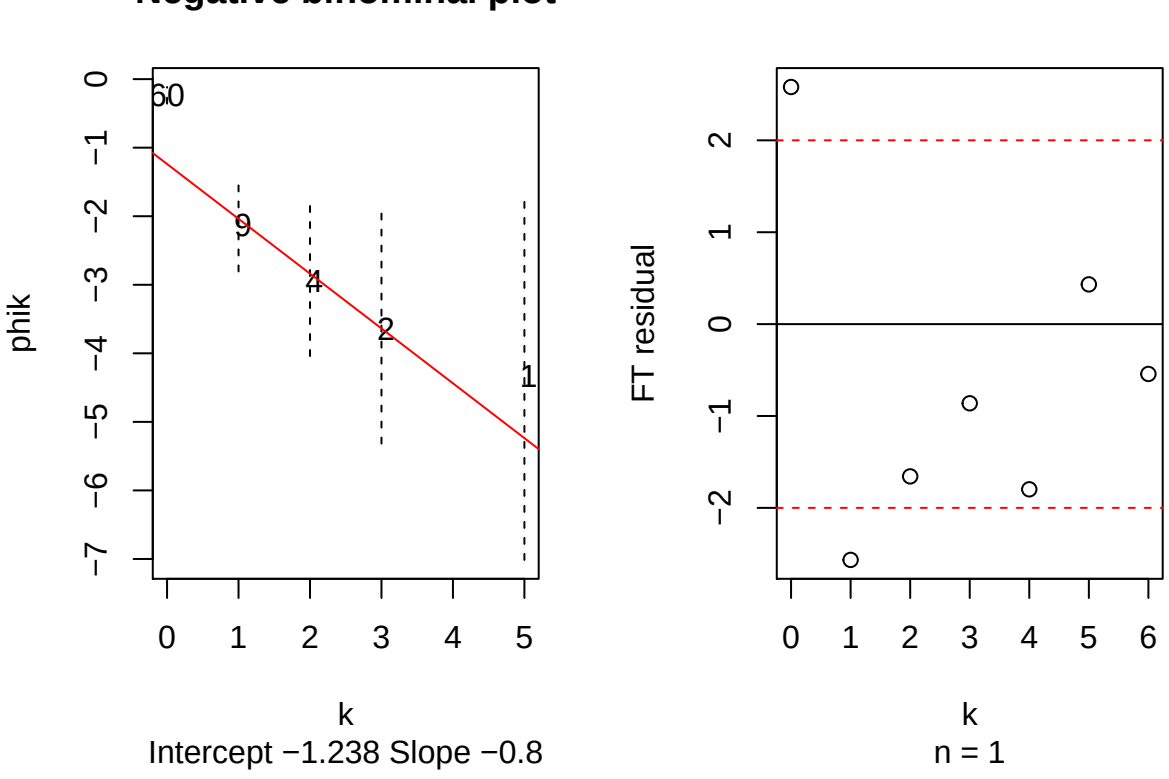
Intercept −1.971 Slope 0.291





Intercept −2.127 Slope −0.047


Negative binominal plot
phik
k
Intercept −0.969 Slope −0.654
FT residual
k
n = 1

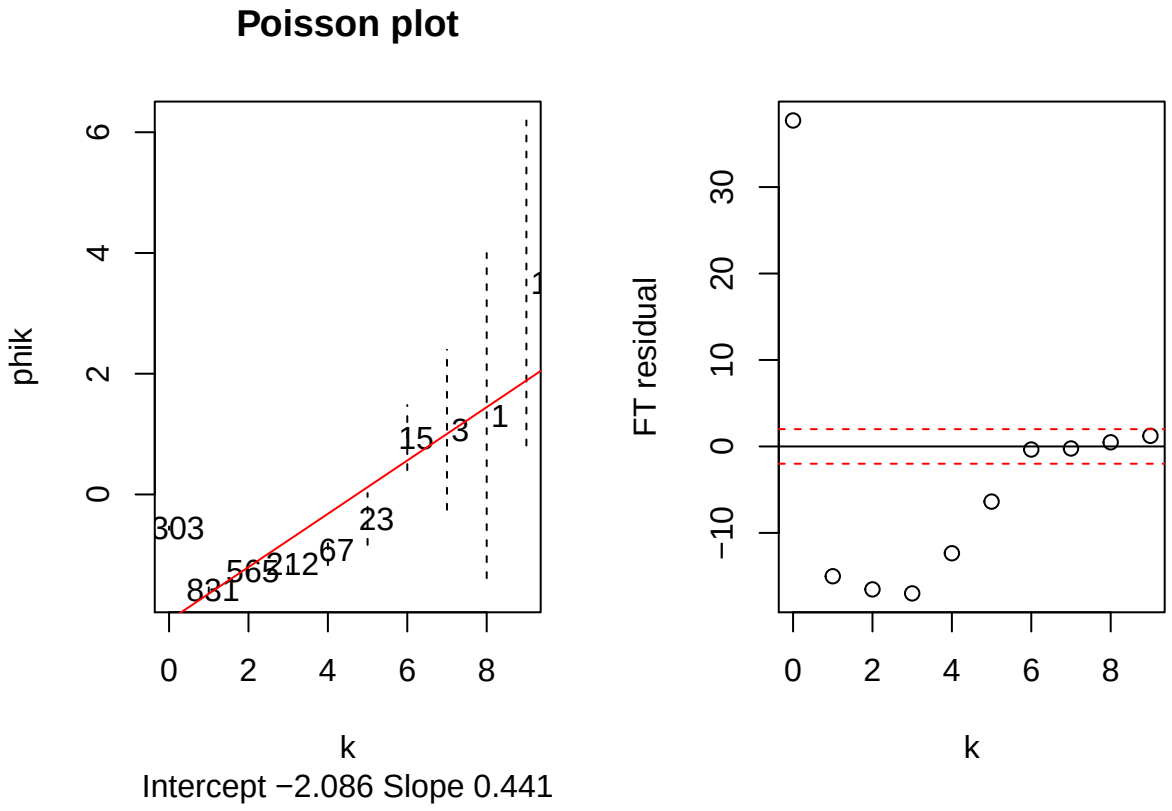

Negative binominal plot
phik
k
Intercept −1.238 Slope −0.8
FT residual
k
n = 1

Solution

The poissonness plots do not seem to be appropriate here. The negative binomial plot is a better fit. Here I choose n = 1.

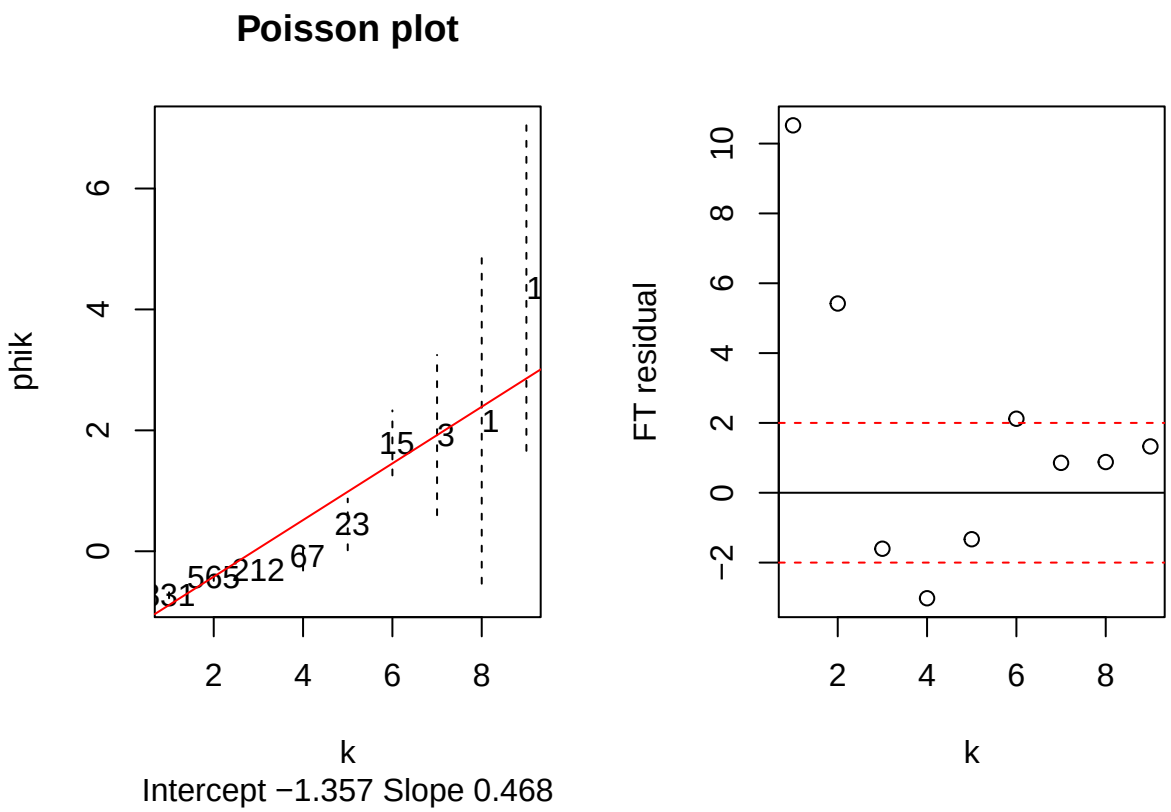
Ex2

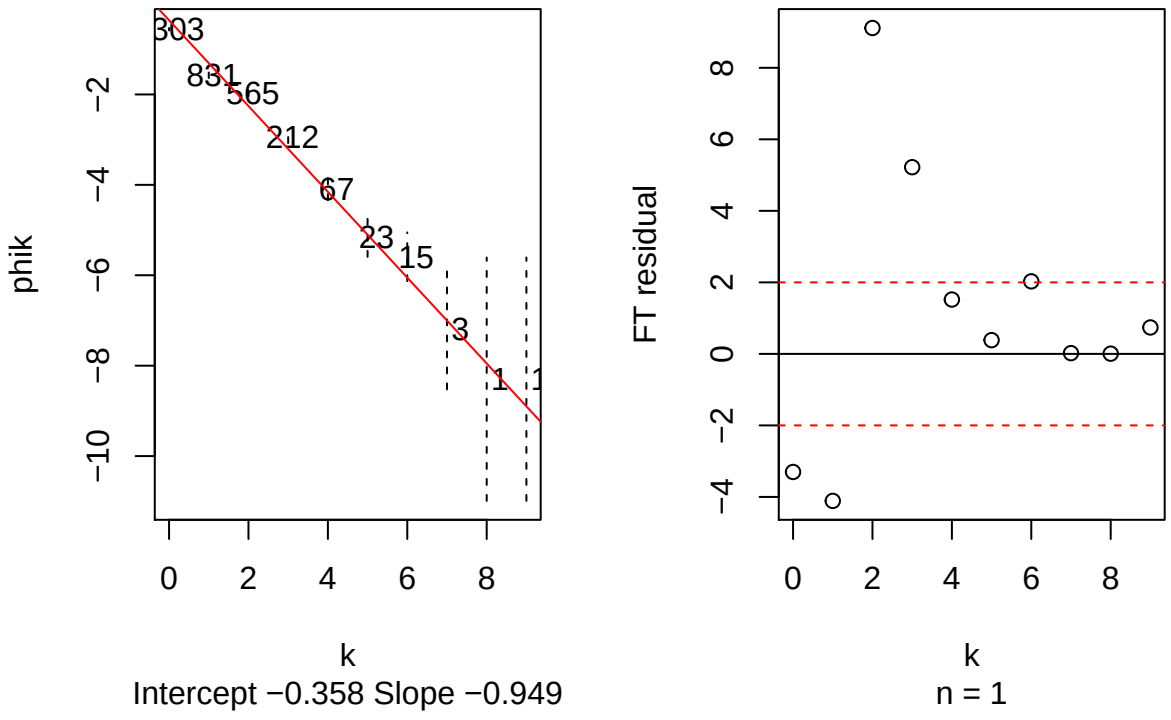
A sample of 4021 British households in 1950 yielded the following numbers of households (nk) with exactly k children (under 14 years of age). Use a Poissonness plot to judge the appropriateness of a Poisson plot. About how many households with no children would you have expected? How well does the model fit for k >= 1? Does a negative binomial distribution fit better?



Solution

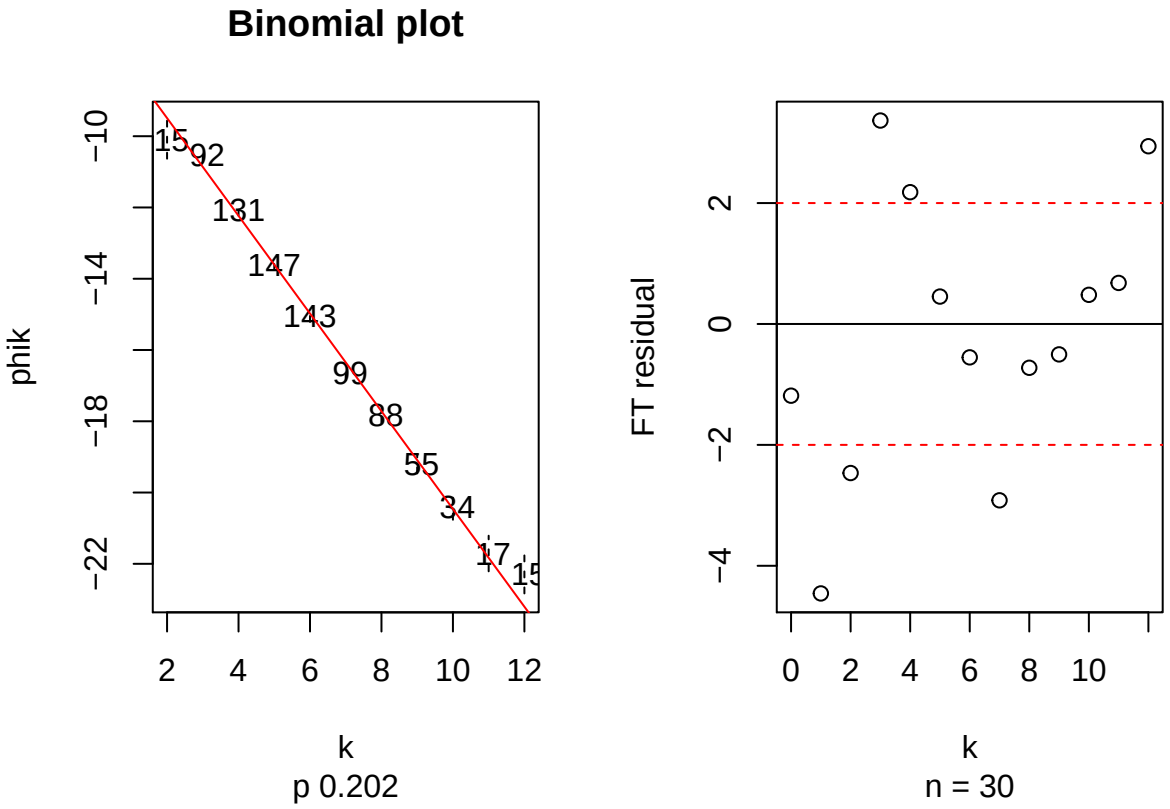
Based on the poisson model, I will expect 850 households with no child.


Poisson plot
phik
k
Intercept −1.357 Slope 0.468
FT residual
k


Negative binominal plot
phik
k
Intercept −0.358 Slope −0.949
FT residual
k
n = 1

Ex3

Sylvia Ann Howland data: number of coincidences between corresponding downstrokes in comparisons of all possible pairs from 42 uncontested signature of Sylvia Ann Howland.

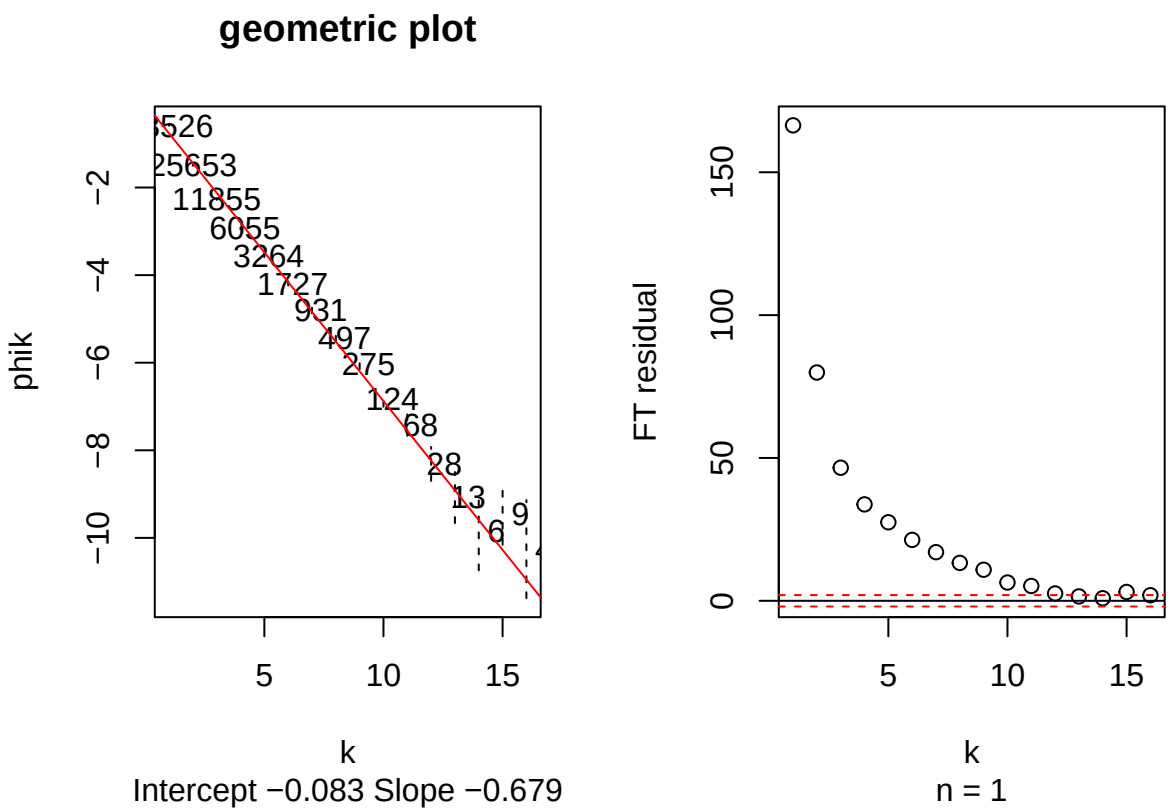


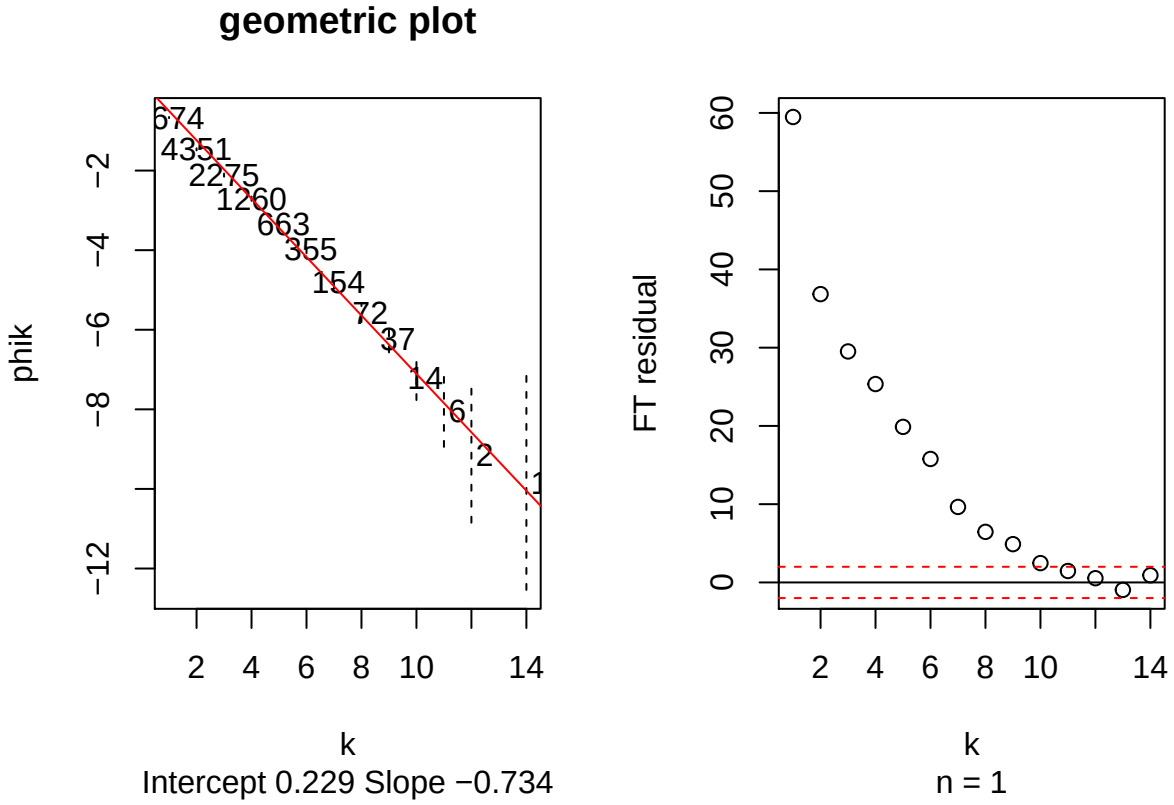
Solution

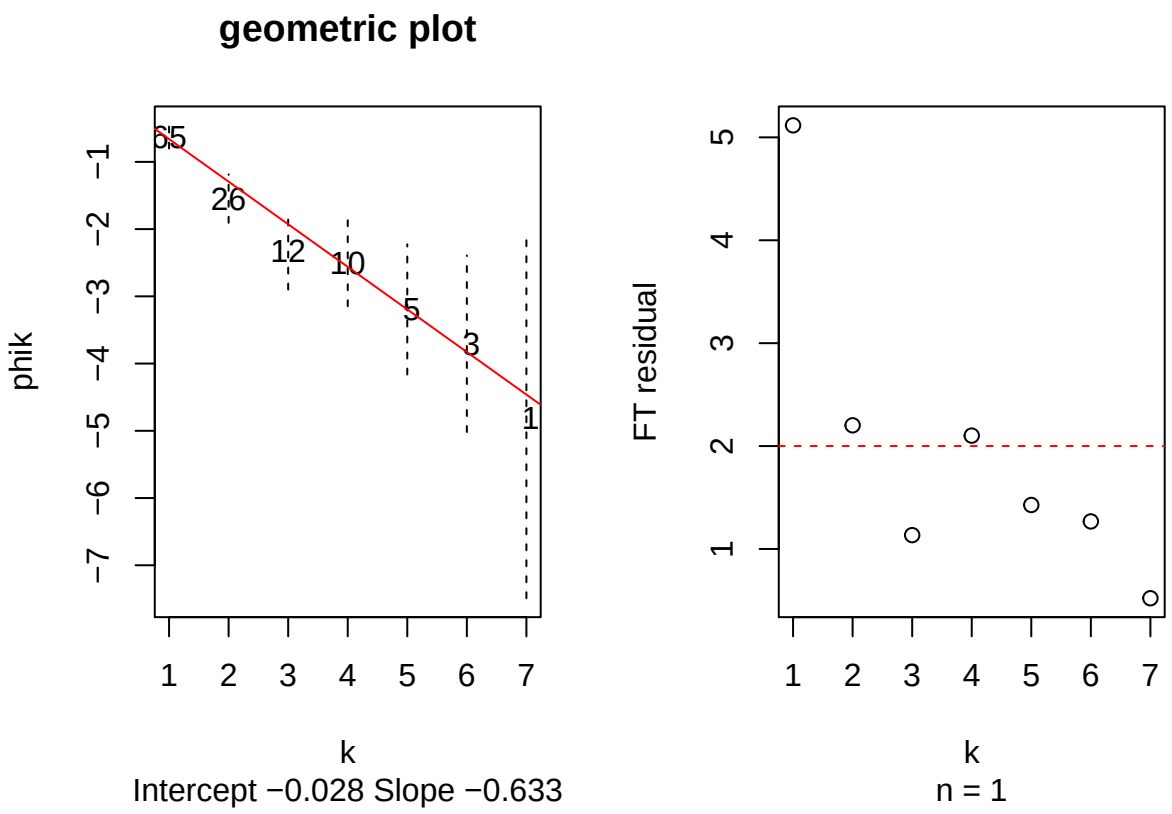
Based on the shape of n_k, I fit a binomial distribution with n=30. It work pretty well. p in the binomial distribution is 0.202.

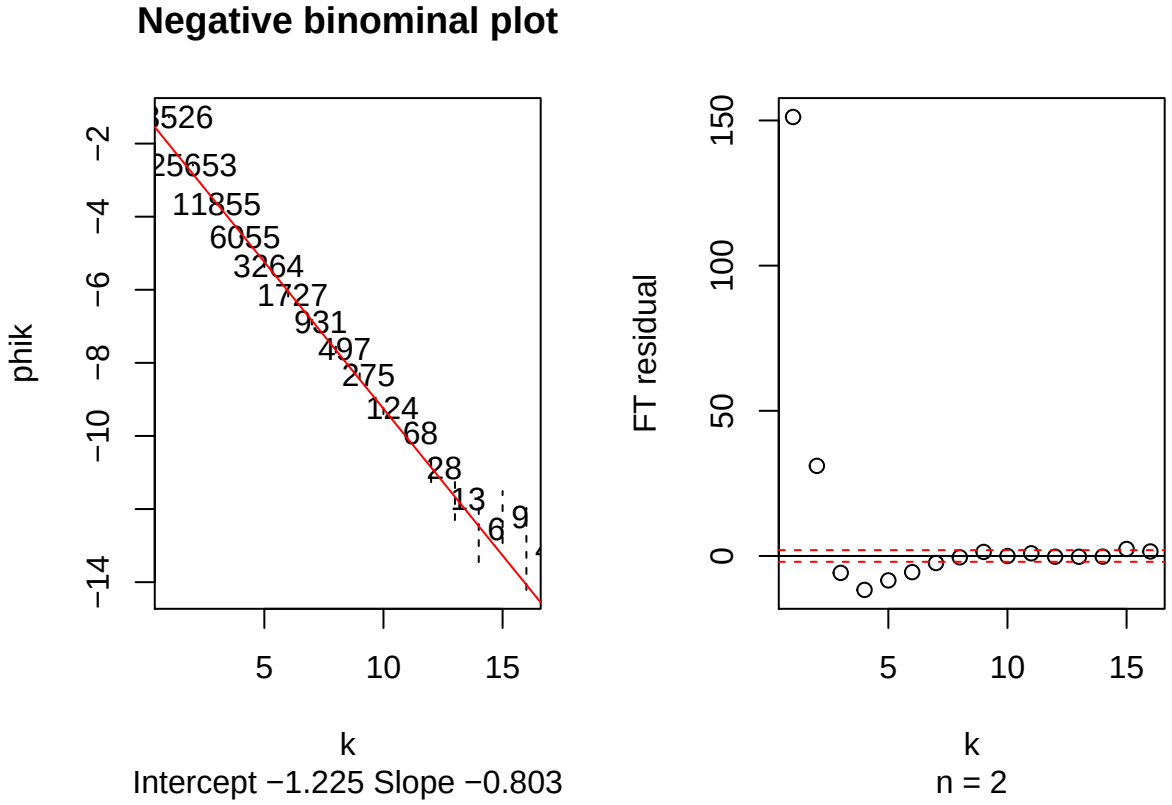
Ex4

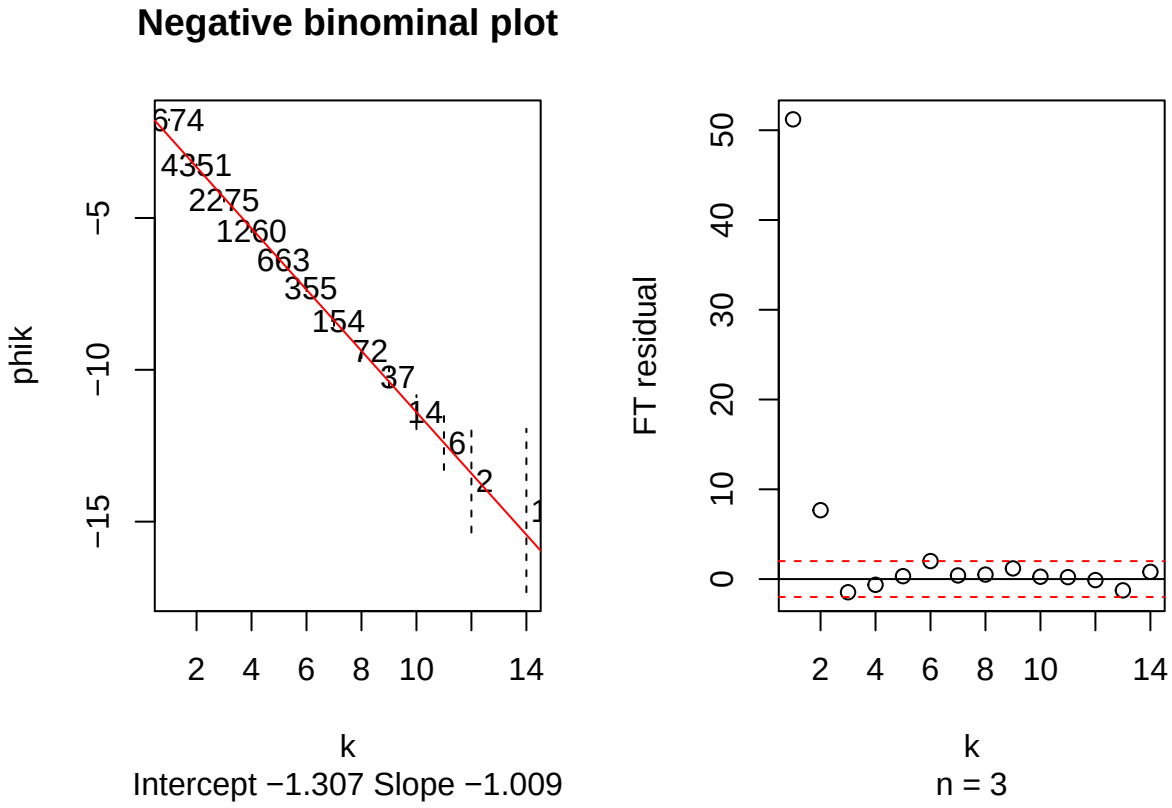
Below are the frequency distributions of library circulation data of the Wishart Library, University of Cambridge, and the "long-loan" collection of the Sussex University Library. Using plots for the geometric distribution and the negative binomial distribution, compare the behavior of these data sets to that of the data from the Hillman library.

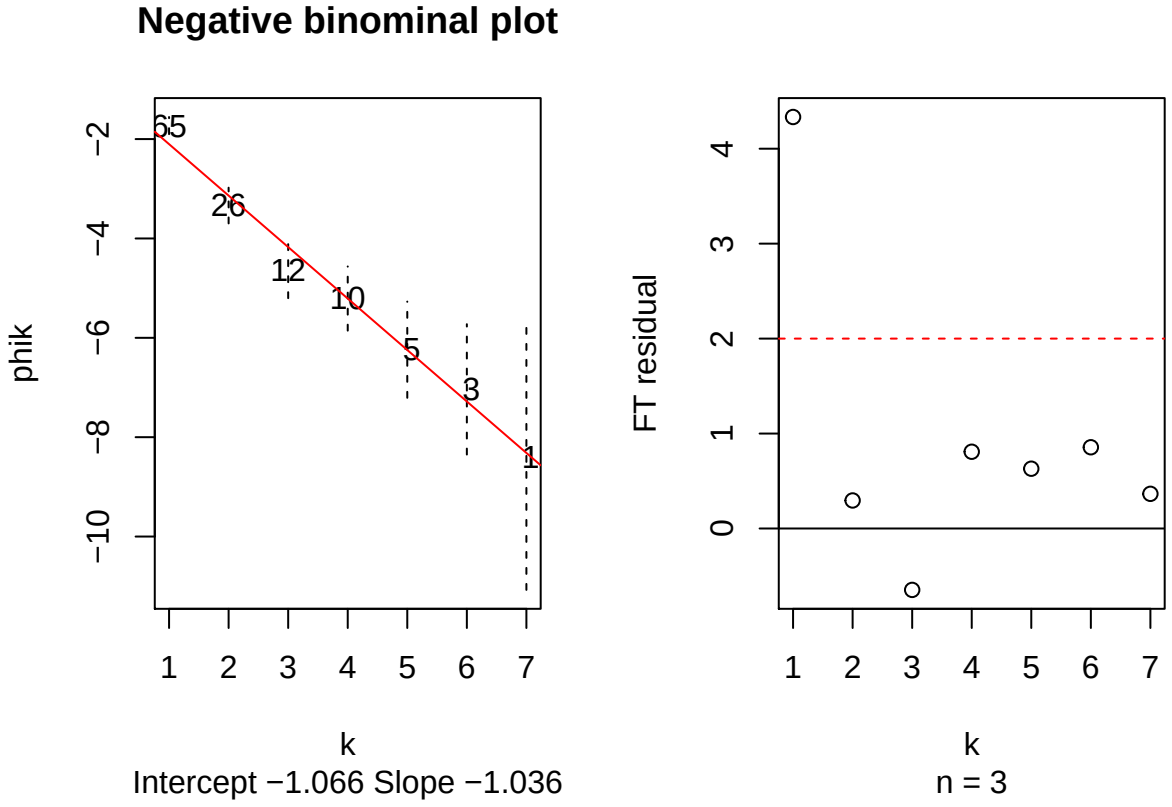

geometric plot
phik
k
Intercept −0.083 Slope −0.679
FT residual
k
n = 1


geometric plot
phik
k
Intercept 0.229 Slope −0.734
FT residual
k
n = 1


geometric plot
phik
k
Intercept −0.028 Slope −0.633
FT residual
k
n = 1


Negative binominal plot
phik
k
Intercept −1.225 Slope −0.803
FT residual
k
n = 2

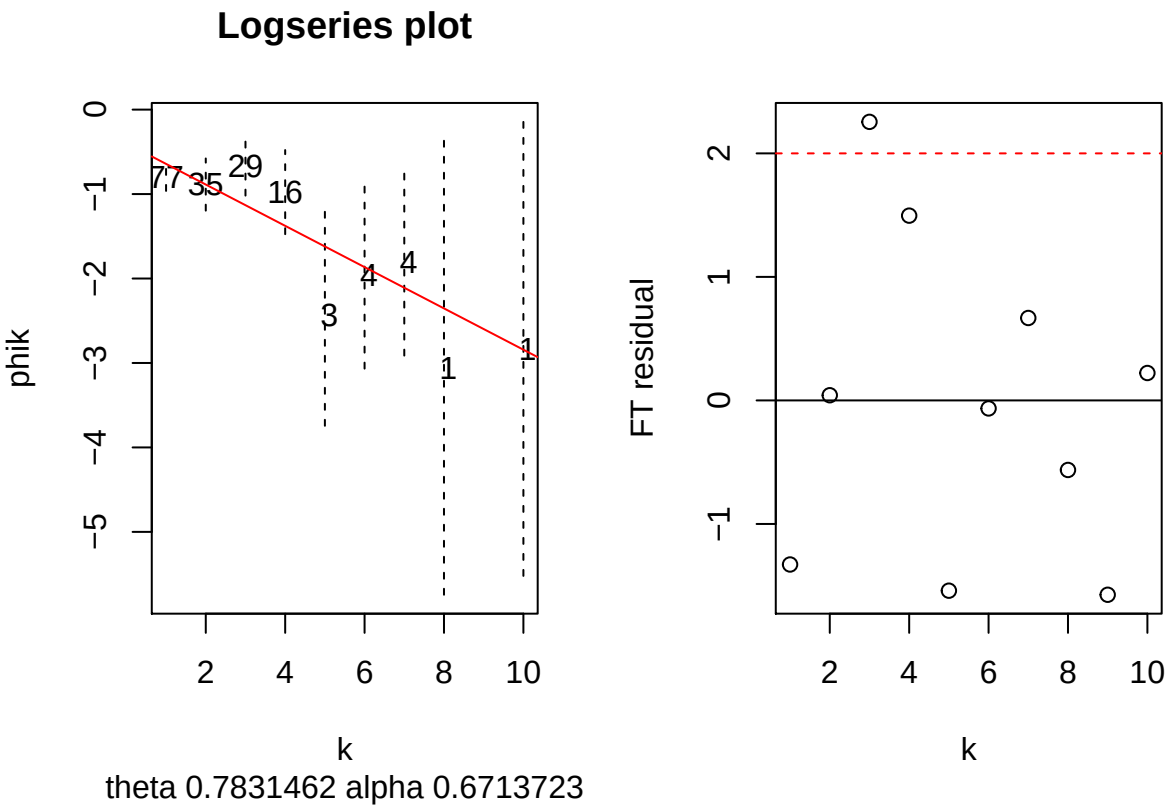

Negative binominal plot
674
4351
2275
1260
663
355
154
72
37
14
6
2
1
phik
−5
−10
−15
2
4
6
8
10
14
k
Intercept −1.307 Slope −1.009
FT residual
50
40
30
20
10
0
2
4
6
8
10
14
k
n = 3


Negative binominal plot
65
26
12
10
5
3
1
phik
−2
−4
−6
−8
−10
1
2
3
4
5
6
7
k
Intercept −1.066 Slope −1.036
FT residual
4
3
2
1
0
1
2
3
4
5
6
7
k
n = 3

Solution

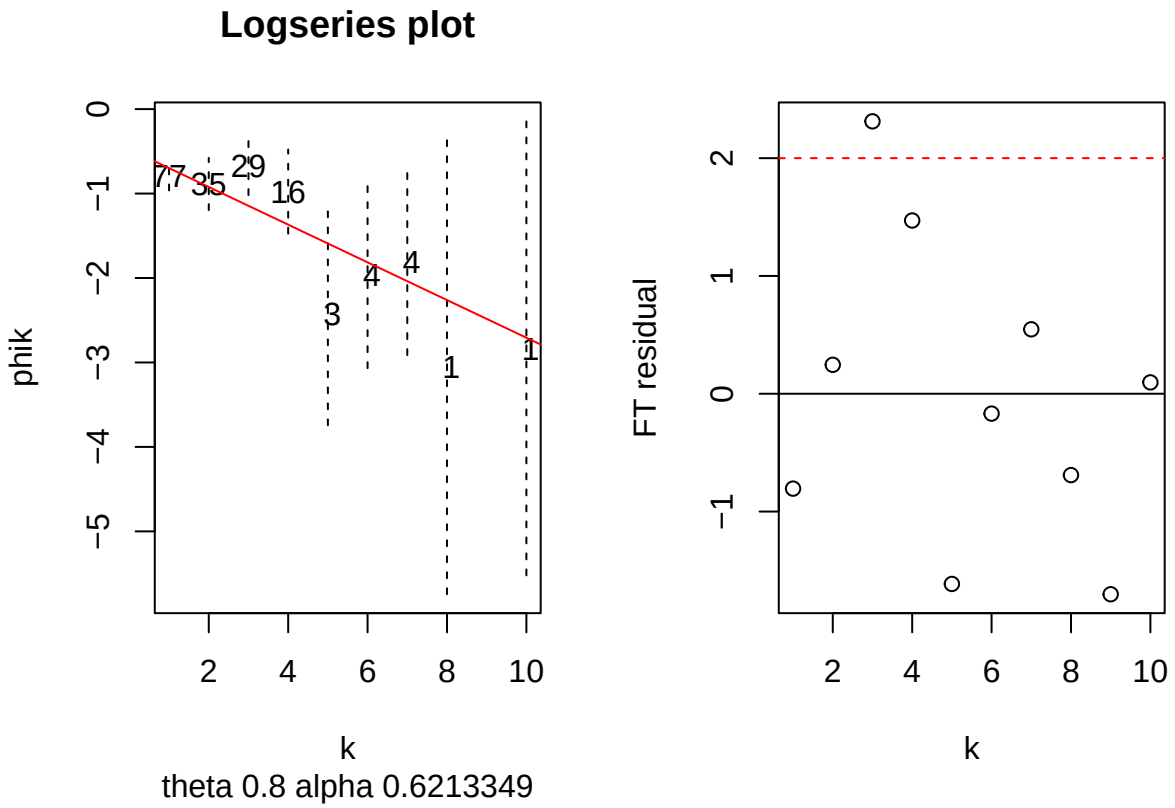
No matter the value of n in negative binomial plot, FT residual at k=1 is always an outlier. Also k=1 has greatest number of observation, which makes concern whether negative binomial is an appropriate fit.

Ex5



Solution

Here I used $\theta = 0.7831462$, $\alpha = 0.6713723$.



$\theta = 0.8$ works equally well.