Fatal Dog Attacks and the Poisson distribution

"Three violent incidents involving pit bulls in one day in Salinas is a statistical anomaly, said animal control officers, but dog attack cases involving the pit breeds and other large dogs are common." A quote from the San Jose Mercury article:

Pit bull attacks common, animal control officers say

Three incidents in one day in Salinas

By KEVIN HOWE

Herald Staff Writer

Posted: 02/28/2012 04:55:47 PM PST

Updated: 02/29/2012 04:48:20 PM PST

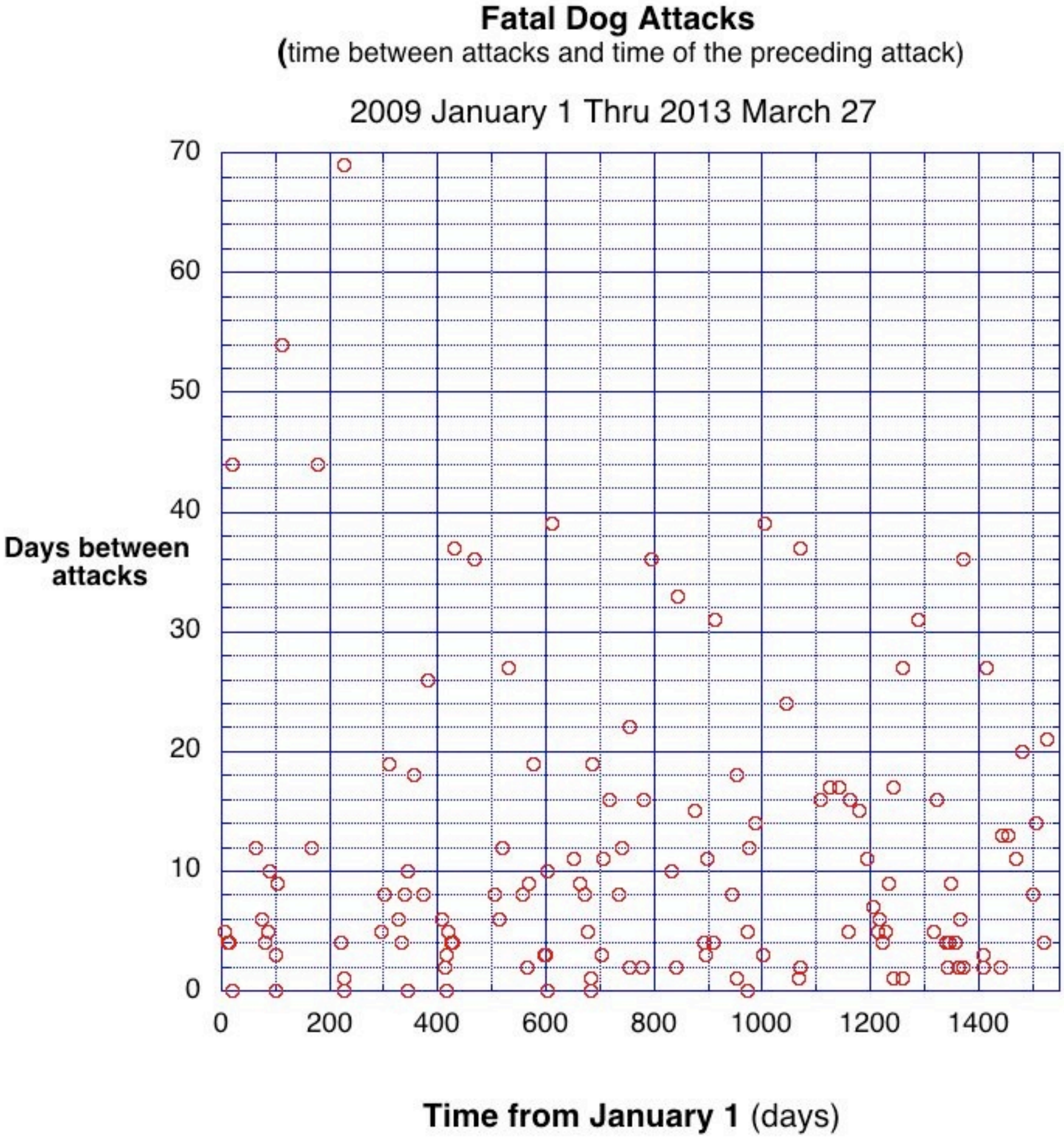
The above article caused me to think, really? Are not attacks independent, the possibility high, but the average small. *I.e.* Dog animal interaction often, but the number of attacks small. These are the requirements for the Poisson interval distribution. Therefore, short time intervals between attacks should have a higher probability than longer intervals. -- Mathematically, exponentially increasing. Also the space between is Poisson governed. Unfortunately, I didn't find this data to ascertain if also a Poisson distribution.

Well .... , Three attacks on the same day in the same small city -- intuitively very rare, but what about fatal dog attacks in the United States? These should fit the Poisson requirements. And the data, tho probably incomplete, is available. If the "incompleteness" is random (normally distributed) my intuition thinks the data is still Poisson distributed.

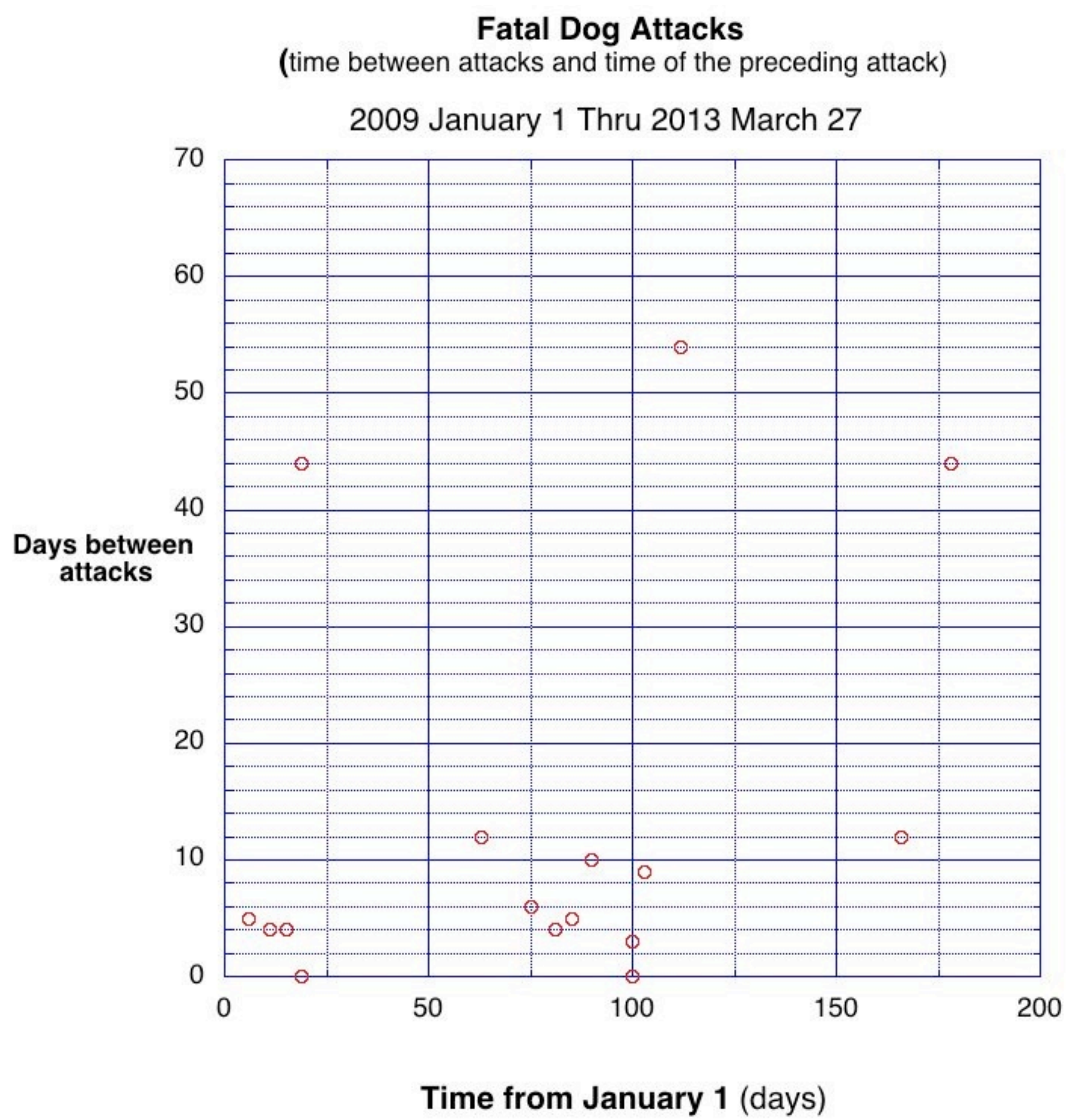
As reported by the Wikipedia page "List of fatal dog attacks in the United States", (1) there are 30 to 35 annual fatalities for the years 2009 thru 2012. Before that the rate is somewhat less. Assuming the recent data is rather complete record of the fatalities, (or as above) an analysis follows with the assumption the deaths are a Poisson distributed.

The following graph is a time series. Each point is the time of the attack and the time between it and the succeeding attack.

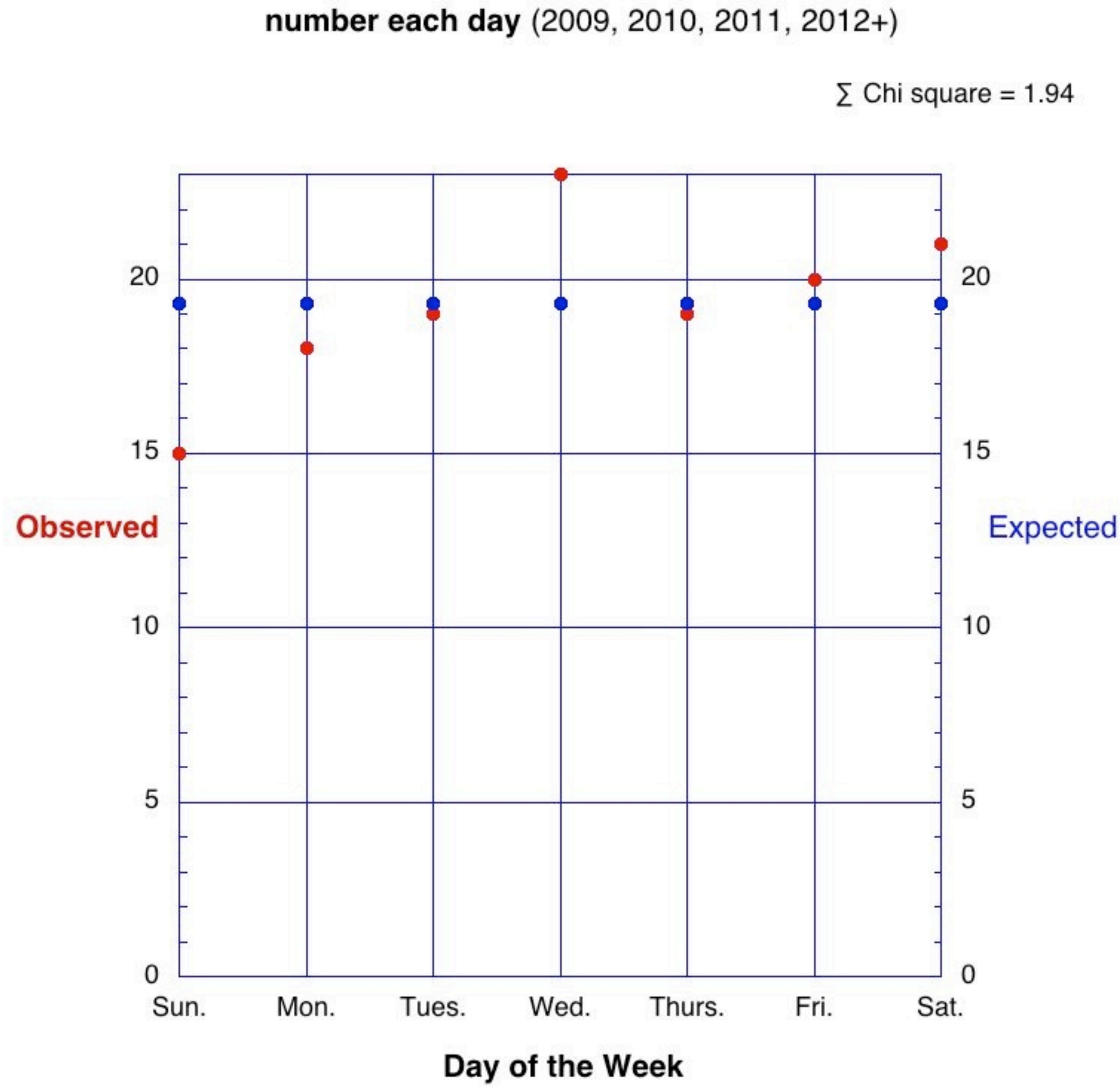
(1) List of fatal dog attacks in the United States - Wikipedia, the free encyclopedia
http://en.wikipedia.org/wiki/List_of_fatal_dog_attacks_in_the_United_States


Fatal Dog Attacks
(time between attacks and time of the preceding attack)
2009 January 1 Thru 2013 March 27
Days between attacks
Time from January 1 (days)
0
10
20
30
40
50
60
70
0
200
400
600
800
1000
1200
1400

The most obvious characteristic of the Poisson distribution is its asymmetry. This is especially obvious with small mean values, e.g. <10. For this reason I chose intervals of 25 days. Below an extract of the preceding graph showing the method, *i.e.* Four events in the first 25 day interval; none in the next, and one the next, etc.



For my first analysis, I used fifty day intervals and my stats. Guru replied:
“Why 50 day intervals? A week or multiples thereof might make more sense as weekends may be crucial.”
So I found the rates for each day of the week, as graphed below:

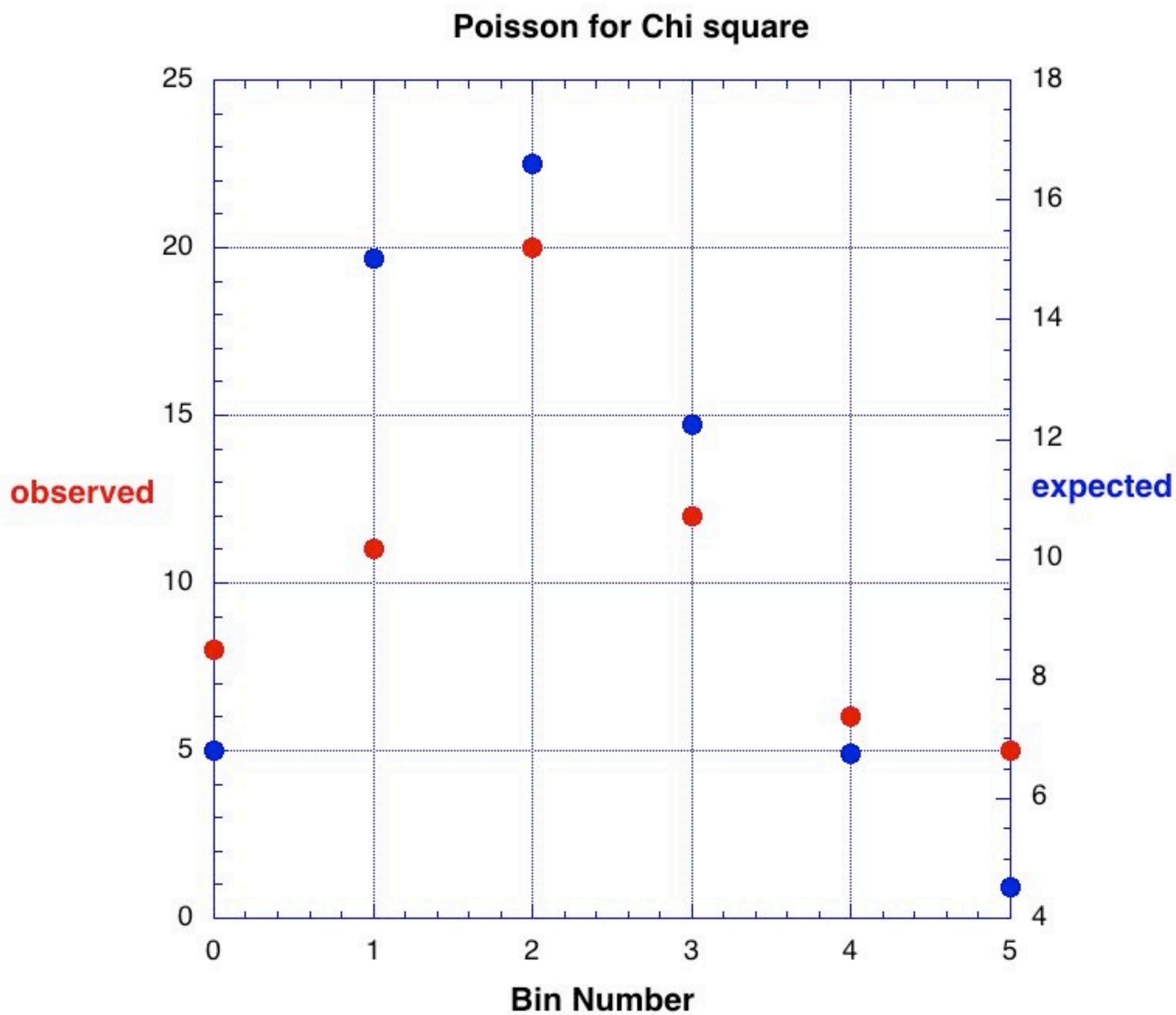


Tho the graph may arouse suspicion, the reduced Chi square (Chi square sum / five degrees of freedom) is somewhat less than one, and, therefore, one will conclude the deaths by day are indeed random.

So, whats the data distribution, and does it agree with the expected Poisson distribution?

$$P_X = \frac{m^X}{X!} e^{-M}$$

Here's the graph of the observed and expected, Poisson distribution.

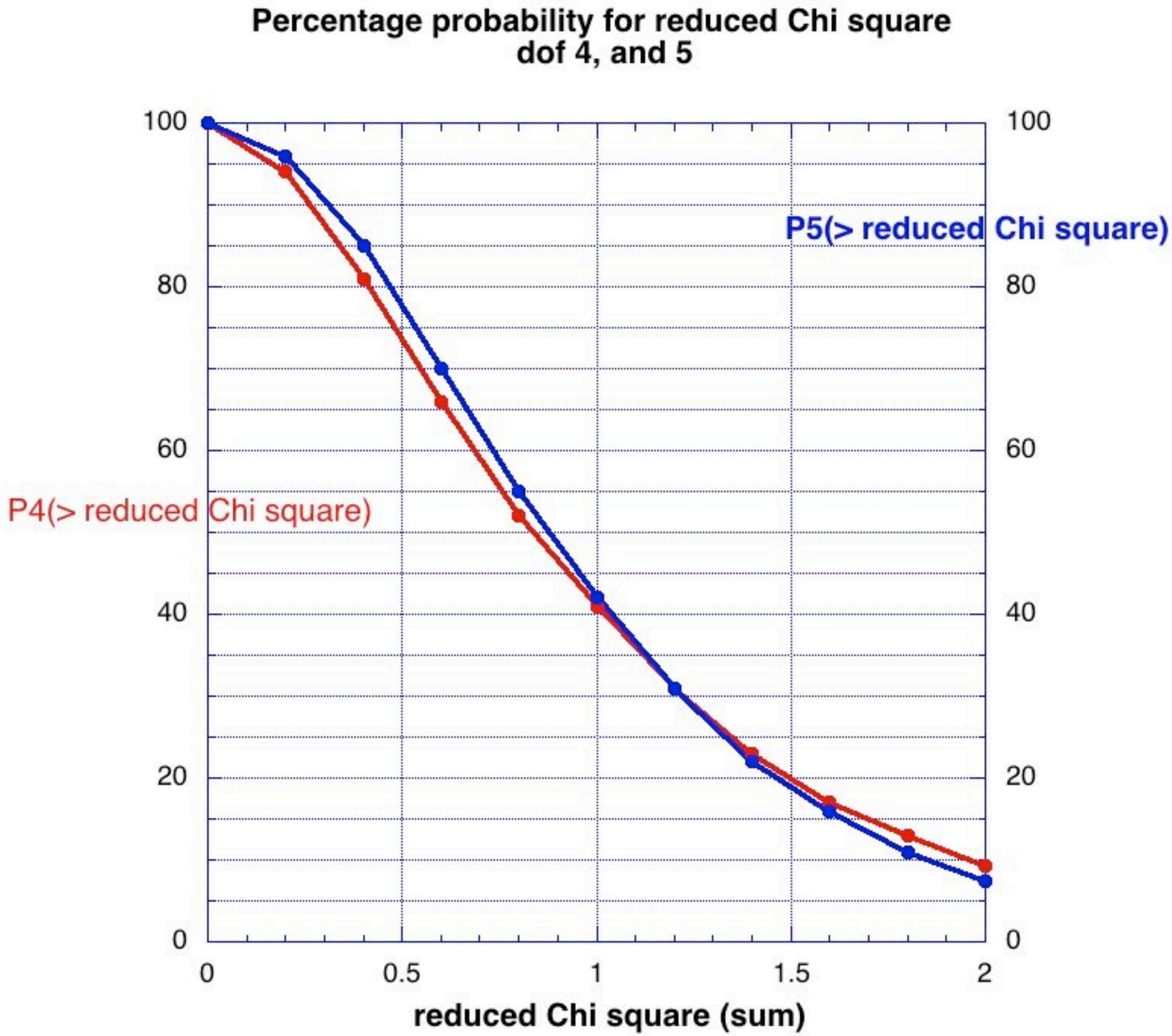


And here’s the data analysis used to create the graph:

25 day # C11	# in each 25 days C12	new (137) C13	counts(X)in25 days C14	Initial events in X C15	P(X) C16	n*P(X) n=62 C17	Chi square C18	Bin # (r C19
1	5	0	0	8	0.109736	6.803632	0.2104	none
2	0	0	1	11	0.242481	15.033821	1.0823	one
3	2	0	2	20	0.267903	16.609985	0.6919	two
4	5	0	3	12	0.1973263	12.234231	0.0045	three
5	2	0	4	6	0.109007	6.758434	0.0851	four
6	0	0	5	4	0.048174	2.986788	0.3437	> four
7	1	0	6	1	0.017742	1.100004	0.0091	
8	1	0	7	0	0.0056005	0.347231	0.3472	
9	1	1	8	0	0.001547	0.095914	0.0959	
10	4	1						
11	0	1		137		61.970001		
12	1	1		62				
13	2	1		2.20968				
14	5	1						
15	2	1						

in X	P(X) C16	n*P(X) n=62 C17	Chi square C18	Bin # (regrouped) C19	observed C20	expected C21	Chi square C22
8	0.109736	6.803632	0.2104	none	8	6.803632	0.2104
11	0.242481	15.033821	1.0823	one	11	15.033821	1.0823
20	0.267903	16.609985	0.6919	two	20	16.609985	0.6919
12	0.1973263	12.234231	0.0045	three	12	12.234231	0.0045
6	0.109007	6.758434	0.0851	four	6	6.758434	0.0851
4	0.048174	2.986788	0.3437	> four	5	4.56	0.0425
1	0.017742	1.100004	0.0091				
0	0.0056005	0.347231	0.3472				sum = 2.1167
0	0.001547	0.095914	0.0959				sum/4 (DOF)=0.529
137		61.970001					
62							
0968							

And here's a graph of an extract of the reduced Chi square probability table from AN INTRODUCTION TO ERROR ANALYSIS. (Taylor) So the probability of a repeat will have a greater reduced Chi square sum is ~70%.

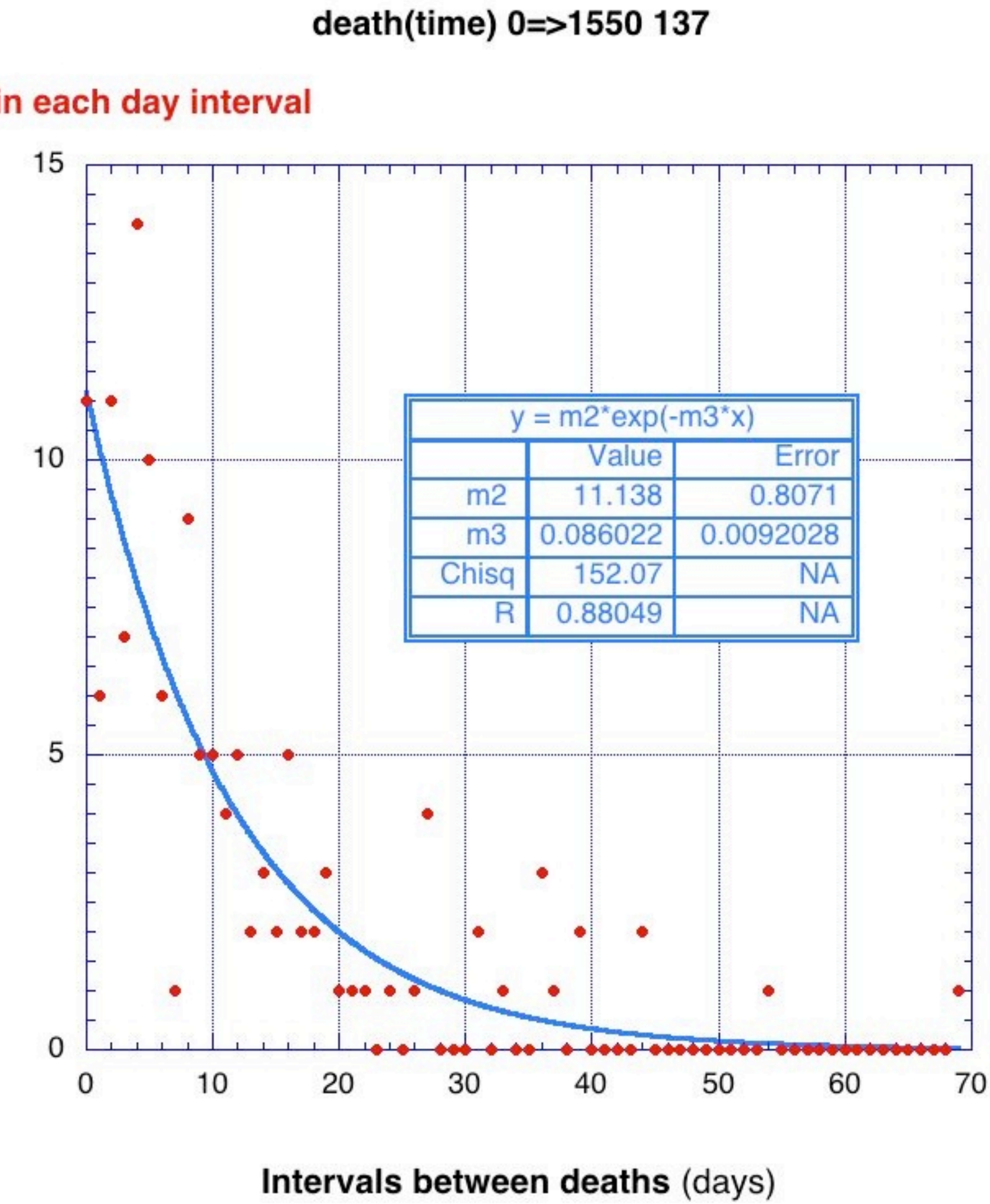


So .... How about the Poisson interval? I claimed the Poisson interval distribution will favour shorter times between dog attack deaths; more accurately the distribution is exponential. Below is an extract from the table of the time of deaths and their respective intervals:

Date (wikipedia)	Time (days)	diff(Time); S-1	S-1 ascending sort
2009-Jan-1			0
2009-Jan-6	6	5	0
2009-Jan-11	11	4	0
2009-Jan-15	15	4	0
2009-Jan-19	19	0	0
2009-Jan-19	19	44	0
2009-Mar-4	63	12	0
2009-Mar-16	75	6	0
2009-Mar-22	81	4	0
2009-Mar-26	85	5	0
2009-Mar-31	90	10	0
2009-Apr-10	100	0	1
2009-Apr-10	100	3	1
2009-Apr-13	103	9	1
2009-Apr-22	112	54	1
2009-Jun-15	166	12	1
2009-Jun-27	178	44	1
2009-Aug-10	222	4	2
2009-Aug-14	226	0	2
2009-Aug-14	226	1	2

From the last column (an ascending sort of the previous one) one obtains eleven deaths on the same day, six spaced by one day, *etc.*

Below is a plot of the intervals (same day to 69 -- no deaths with a greater spacing than 69 days) and the number in each. ***N.B.*** the exponential fit. There were 137 deaths in 1550 days.



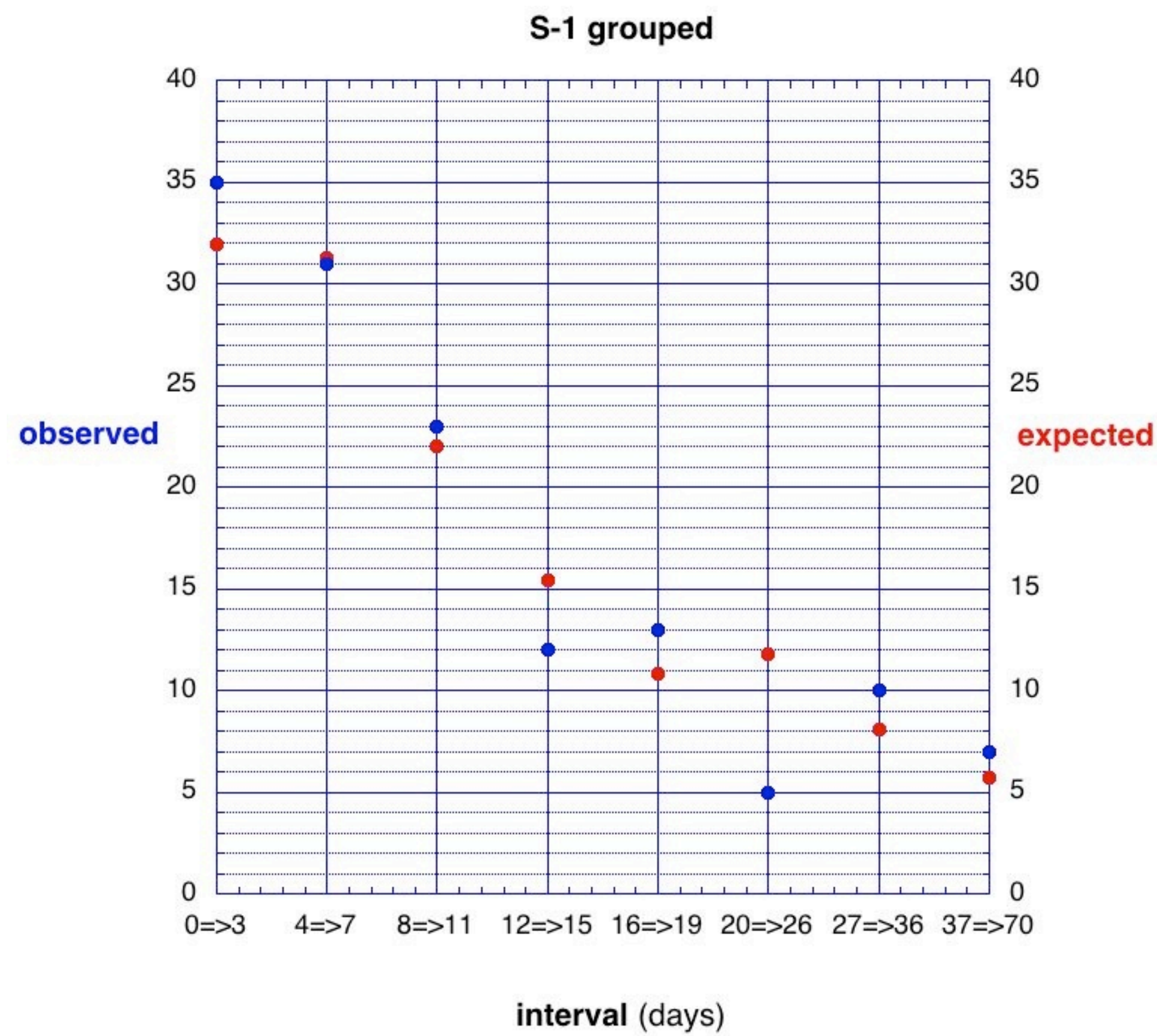
Not very good agreement? Reminds me of Press' statement: "... This approach is known as chi-by-eye. Luckily, its practitioners get what they deserve." Here's the data grouped so each category has at least five events and five categories with the calculated expected number from the Poisson interval distribution.

interval (days)	number observed	number expected	Chi square	labels
0=>3	35	31.91	0.2992	
4=>7	31	31.3	0.0029	
8=>11	23	22	0.0455	
12=>15	12	15.43	0.7625	
16=>19	13	10.8	0.4481	
20=>26	5	11.79	3.9104	
27=>36	10	8.08	0.4562	
37=>70	7	5.69	0.3016	
			6.2264	Sum
			1.04	Sum / DOF (6)
			35	P(6) > percent

N.B. The intervals need not be equal. I'll post the derivation of the Poisson interval used above on my site: **cleyet.org**

The probability of another 1550 day data set exceeding this one is well above the 5% "cut off". The next slide is a graph of the first three columns.

o



So, like the famous horse kick deaths in the Prussian Army, dog attack deaths are also governed by the Poisson distribution.