British Standards Institution Study Day

Detecting a single event

Martin Bland
Prof. of Health Statistics
University of York
http://martinbland.co.uk

Detecting a single event

The problem is this. If we have a series of cases where no event has taken place, what is the estimated event rate? Our best estimate of the proportion of cases which have an event is zero, but there will be uncertainty in this estimate. Just because we have not seen an event yet does not mean we will never see one. We need a confidence interval for this estimate.

Because the number of events observed is zero, we cannot use the usual standard error estimate for the confidence interval. Instead, we use a small sample confidence interval for the estimate, based on the exact probabilities of the Binomial distribution. The Binomial distribution has two parameters, p, the proportion of observations which are an event, and n, the number of observations. We find p which would give a probability 0.025 of having zero events. This is the upper limit of the 95% confidence interval. The lower limit would be the value for p which would have a probability of 0.975 of zero events, which is, of course, $p = 0.0$.

For example, suppose we observe $n = 40$ cases with no events. What is the estimated proportion in the population who would experience the event? For this, the 95% confidence interval is 0 to 0.088. The upper limit for the estimates proportion having events would be 8.8%. If the proportion were more than 8.8%, samples of 40 observations with no events would quite unusual, only 2.5% of possible samples doing this.

Suppose we observe a larger sample, say $n = 100$ cases with no events. Then the 95% confidence interval = 0 to 0.036, the upper limit for the proportion with events would be 3.6%. Suppose we observe $n = 1000$ cases with no events. The 95% confidence interval would be 0 to 0.0037, upper limit = 0.37%.

It is easy to tabulate this upper limit for selected sample sizes (Table 1). These proportions may be greater than intuition would suggest. Thus, for example, if we want to be fairly sure that the rate is less than 1% (0.01), we need to observe no events in 350 cases.

There is a program available, biconf.exe, which can be used to calculate the exact 95% confidence interval for a Binomial proportion. It is now a bit difficult to use as Microsoft have dropped MS-DOS from Windows 7. You can download this from:

> http://www-users.york.ac.uk/~mb55/soft/soft.htm

or go to

> http://martinbland.co.uk/

and follow the link to "Simple statistical software" on the menu. If you have a machine running Windows 7, you can use a free download called DOSBox. This was written for running old games, but works for any software. Biconf.exe can be used when there are events, too.

Table 1. Upper 95% confidence limits for the Binomial proportion when no event has been observed

Number	Proportion	Number	Proportion	Number	Proportion
10	0.31	41	0.086	140	0.026
11	0.28	42	0.084	150	0.024
12	0.26	43	0.082	160	0.023
13	0.25	44	0.080	170	0.021
14	0.23	45	0.079	180	0.020
15	0.22	46	0.077	190	0.019
16	0.21	47	0.075	200	0.018
17	0.20	48	0.074	250	0.015
18	0.19	49	0.073	300	0.012
19	0.18	50	0.071	350	0.010
20	0.17	51	0.070	400	0.0092
21	0.16	52	0.068	450	0.0082
22	0.15	53	0.067	500	0.0074
23	0.15	54	0.066	600	0.0061
24	0.14	55	0.065	700	0.0053
25	0.14	56	0.064	800	0.0046
26	0.13	57	0.063	900	0.0041
27	0.13	58	0.062	1000	0.0037
28	0.12	59	0.061	1100	0.0034
29	0.12	60	0.060	1200	0.0031
30	0.12	65	0.055	1300	0.0028
31	0.11	70	0.051	1400	0.0026
32	0.11	75	0.048	1500	0.0025
33	0.11	80	0.045	1600	0.0023
34	0.10	85	0.042	1700	0.0022
35	0.10	90	0.040	1800	0.0021
36	0.097	95	0.038	1900	0.0020
37	0.095	100	0.036	2000	0.0018
38	0.093	110	0.033		
39	0.090	120	0.030		
40	0.088	130	0.028		

How does the method work? It is easy to write down the formula for probabilities in the Binomial distribution. The probability of seeing r events out of a possible n is:

$$\Pr(r) = \frac{n!}{r!\,(n-r)!} p^r (1-p)^{n-r}$$

(but we only do that to make ourselves look clever). Then to find the probability of being less or equal to than some number of events, we sum all the probabilities for r up to that number. For the confidence interval, we find p so that this is 0.025. This is all built into the program.

A power calculation

How big a sample do we need to have a 90% chance of finding an event? We postulate an event probability for the population, p. We then ask: what is the probability of no events in a sample of size n? This is the same asking: what proportion of possible samples of size n would have no event?

The probability that a given observation has no adverse event $=1-p$. The probability of no observations in n observations $=(1-p)^n$. We set this equal to $1 - \text{power} = 1 - 0.90 = 0.1$.

For example, suppose adverse events happen once in 100 trials. What sample to we need to have 90% chance of seeing an event?

$$(1-0.01)^n = 0.1$$

$$n\log(0.99)=\log(0.1)$$

$$n=\log(0.1)/\log(0.99)=229.1$$

We need 229 observations to have a 90% chance of an adverse event.

We can do this for any value of p. Figure 1 shows a graph of the sample size required for detecting adverse event proportions as small as 0.001 or 0.1%. For more practical sample sizes, Figure 2 shows the portion of this graph for proportions as small as 0.01 or 1%.

Figure 1. Sample size required to give power 90% to detect adverse event proportions.

3
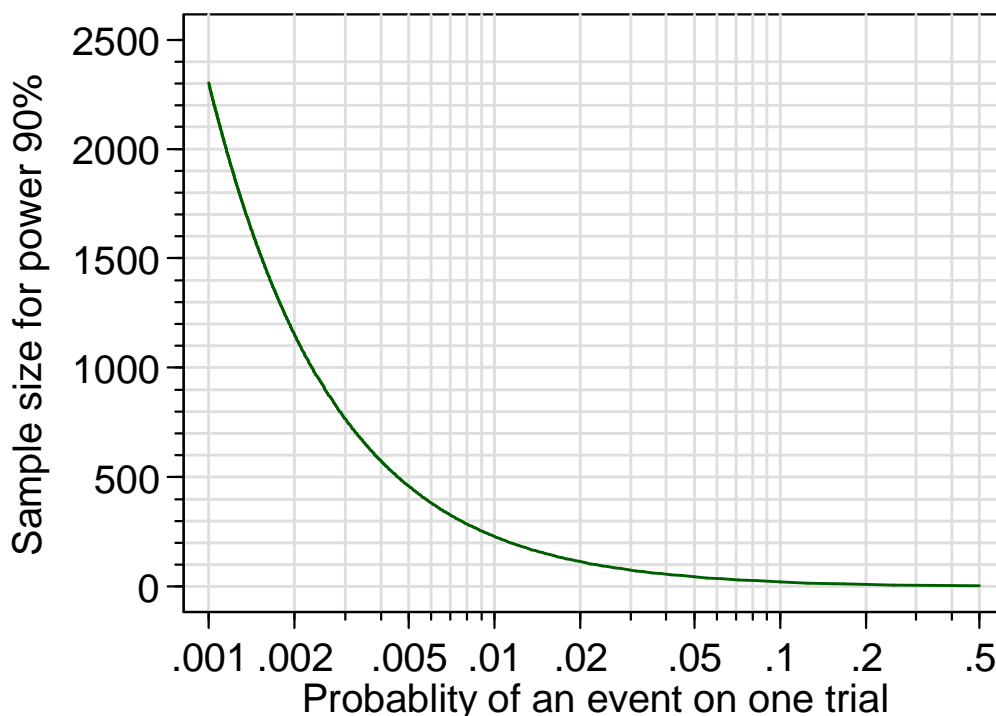
Figure 2. Sample size required to give power 90% to detect adverse event proportions for proportions greater than 1%.