6

Evaluating risks

In prospective studies the risk or incidence of contracting a disease is often represented by a probability p, the probability that someone drawn from a cohort contracts a disease during a certain period of time. Or p could represent the prevalence of a disease within a certain population at the present time. In either case a random sample of individuals is examined at a fixed time and the number within the sample with the disease is noted. The question is then how to use this information to estimate p, or to test hypotheses regarding p.

6.1 Methodology

Data and model

- The data are a set of n dichotomous observations; that is, each taking on one of two possibilities, say D for diseased, $\bar{D}$ for not diseased, or labeled numerically by 1 and 0.
- The binomial model assumes that there are n independent, identically distributed variables, say $I_1, \ldots, I_n$, with $P(I_j = 1) = p$, $P(I_j = 0) = 1 - p$, where $0 < p < 1$ is unknown.

Questions

- What is the evidence that the risk p exceeds a certain fixed level p_0?
- What is a confidence interval for the risk p?

Meta Analysis: A Guide to Calibrating and Combining Statistical Evidence Elena Kulinskaya, Stephan Morgenthaler, and Robert G. Staudte

Test statistic and distribution

- The test statistic is the number of 1's amongst the n outcomes, $S = \sum_{j=1}^{n} I_j$; it has the binomial distribution with parameters n, p.

Transformation to evidence and distributional properties

- Let $\tilde{p} = (S + 3/8)/(n + 3/4)$. Then the evidence for the alternative $p > p_0$ to the null $p = p_0$ is given by the classic transformation

$$T = 2\sqrt{n}\left\{\arcsin(\sqrt{\tilde{p}}\,) - \arcsin(\sqrt{p_0}\,)\right\}.$$

- This T is approximately normal for $np(1 - p) \geq 5$.
- The expected evidence $E[T] \doteq \sqrt{n}\,\mathcal{K}(p)$, where the Key Inferential Function is defined by

$$\mathcal{K}(p) = 2\left\{\arcsin(\sqrt{p}\,) - \arcsin(\sqrt{p_0}\,)\right\}.$$

- The evidence T has standard deviation lying between 0.95 and 1.0 for $0.2 < p < 0.8$ for sample size $n = 9$, and this range expands to $0.07 < p < 0.93$ for $n = 30$. For any n, as p approaches 0 or 1, the standard deviation of T approaches 0; but this does not mean that T is not a good estimator of its expected value. For more information, see Figure 18.1 and accompanying text.

Interpretation

- Positive values of T are evidence for the alternative $p > p_0$, while the magnitude $|T|$ of a negative value of T is positive evidence for the alternative $p < p_0$. Evidence $T^{\pm}$ for the two-sided alternative $p \neq p_0$ can be obtained from $|T|$ via the transformation (2.3).

Choosing the sample size

- For testing $p = p_0$ against $p > p_0$ one may choose n_1 so that the expected evidence for a fixed p_1 of interest is at least τ_1. This requires n_1 to satisfy $\tau_1 \leq \sqrt{n}_1\,\mathcal{K}(p_1)$, or $n_1 \geq \{\tau_1/\mathcal{K}(p_1)\}^2$.
- For example, if the null is $p = 0.5$ to achieve 'strong' expected evidence $\tau_1 = 5$ against $p = 0.9$ one requires $n_1 \approx 29$. Some other values are also shown in Table 6.1.

Confidence intervals

- Letting $\tilde{p} = (S + 3/8)/(n + 3/4)$ and $T = 2\sqrt{n}\,\arcsin\left(\sqrt{\tilde{p}}\,\right)$, a 95 % confidence interval for p is given by

$$\left[\left\{\sin\left(\frac{T - z_{0.975}}{2\sqrt{n}}\right)\right\}^2, \left\{\sin\left(\frac{T + z_{0.975}}{2\sqrt{n}}\right)\right\}^2\right].$$

  It is understood that if the sine values are less than 0 or greater than 1, they are replaced, respectively, by 0 and 1, before squaring.

Table 6.1 Approximate sample sizes required to achieve weak, moderate or strong expected evidence for alternatives p_1 to the null $p_0 = 0.5$.

p_1	$\arcsin(\sqrt{p_1})$	$\tau_1 = 1.645$	$\tau_1 = 3.3$	$\tau_1 = 5.0$
0.5	0.78540	—	—	—
0.6	0.88608	67	267	617
0.7	0.99116	16	64	148
0.8	1.10715	7	26	60
0.9	1.24905	4	13	29

- These intervals are much more accurate than traditional large sample intervals of the form $\hat{p} \pm z_{0.975}\sqrt{\hat{p}(1-\hat{p})/n}$, where $\hat{p} = X/n$ (see Section 18.2).
- When p is near 0, confidence intervals for p are often derived after a log-transformation of $\hat{p} = S/n$. Such intervals are comparable in performance to those based on the formula displayed above (see Section 18.4). A rule of thumb suggested based on simulations reported in Section 18.4 is that the when conditions $np(1-p) \geq 5$ and $n \geq 25$ are satisfied, then the arcsine intervals will have empirical coverage between 93 and 97 %; and for $np(1-p) \geq 11$ and $n \geq 100$, the coverages will lie between 94 and 96 %.

6.2 Examples

These methods have already been illustrated for the case of $p_0 = 0.5$ in matched pair experiments in Section 1.2.

6.2.1 Ultrasound and left-handedness

A study by Salvesen *et al.* (1993) found a slight positive association between *in utero* routine ultrasonography and subsequent left-handedness of 8- and 9-year-old children. Similar reports for only boys in a different study were reported by Kieler *et al.* (1998). If the proportion of left-handers in the general population is $p_0 = 0.1$, how large a sample is required to obtain strong evidence that *in utero* routine ultrasonography leads to a proportion p of left-handers which exceeds the general population proportion 0.1 by 10 %? That is, what is the minimum sample size required to obtain expected evidence 5 for an alternative $p = p_1 = 0.11$?

We require $n_1 \geq \{\tau_1/\mathcal{K}(p_1)\}^2 = (5/0.03263)^2 = 23\,481.3$, or 23 482. For only moderate evidence 3.3 of a 10 % increase, one needs a minimum sample size of $n_1 = 10\,229$.

6.2.2 Treatment of recurrent urinary tract infections

If untreated, recurrent urinary tract infections continue in 65 % of observed patients (see Section 19.5). Let p represent the risk of continued infection following treatment by antibiotics. In study 2 of Table 19.1 eight of 21 patients treated by antibiotics had further infections during the study period. How much evidence is there for the alternative $p < 0.65$ to the null $p = 0.65$ based on these data?

In the notation of this chapter $n = 21$ and $S = 8$. An estimate of p is $\tilde{p} = (S + 3/8)/(n + 3/4) = 0.3895$. Hence the evidence for the alternative $p < 0.65$ is $T = 2\sqrt{n}\,\{\arcsin(\sqrt{p_0}) - \arcsin(\sqrt{\tilde{p}})\} = 2\sqrt{21}\,\{\arcsin(\sqrt{0.65}) - \arcsin(\sqrt{0.3895})\} = 2.4$, which is between weak and moderate. An analysis based on comparing treatment patients to similar controls is given in Section 7.2.1.