Untitled

hugyu

27 1 2020

Contents

- Binary outcomes and proportions
- Comparative parameters of risks and their estimation
- Binomial regression models and comparative parameters
- Adjustment for confounding and evaluation of modification by binomial regression

Main R functions covered:

- `twoby2()` (Epi package)
- `glm()`
- `ci.lin()` (Epi package)

Outcomes in epidemiologic research

Epidemiologic studies address the occurrence of diseases and other health related phenomena:

- (a) cross-sectional: **prevalence** of diseases,
- (b) longitudinal: disease **incidence**, and mortality

Often we want to compare the prevalence or incidence of disease between two groups defined by a binary *risk factor* X

- $X = 1$: `exposed` $X = 0$: `unexposed`

Types of outcome variables

- *Binary* (0/1) variables at individual level
  - disease *status* at a *time point*
  - *change* of status, *event* or *transition* ({*e.g.*} *from healthy to diseased)*
- *Proportions* at group level
  - prevalence
  - incidence proportion or cumulative incidence,
- *Rates* of events
  - incidence or mortality rate (per 1000 y)
  - car accidents (per million km)
- *Time* to event
  - survival time (often censored)

Incidence and prevalence proportions}

- **Incidence proportion** (R) of a binary (0/1) outcome (disease, death etc.) over a fixed risk period is defined

$$R = \frac{D}{N} = \frac{\text{number of new cases during period}}{\text{size of population-at-risk at start}}$$

Also called {**cumulative incidence} (or even "risk").\ NB.**

This formula requires complete follow-up, i.e. no {*censorings}, and absence of {competing risks}.*

- ***Prevalence (proportion)*** *P of disease at time point t*

$$P = \frac{\textit{no. of existing cases at } t}{\textit{total population size at } t}.$$

Two-group comparison

- Binary risk factor X: exposed vs. unexposed.
- Summarizy results from cohort study with fixed risk period and no losses:

Exposure	Cases	Non-cases	Group size
yes	D_1	C_1	N_1
no	D_0	C_0	N_0
total	D_+	C_+	N_+

- Incidence proportions in the two exposure groups

$$R_1 = \frac{D_1}{N_1}, \qquad R_0 = \frac{D_0}{N_0}.$$

- These are crude *estimates* of the true *risks* π_1, and π_0 of outcome in the two exposure categories.

Example: Observational clinical study

Treatment failure in two types of operation for renal calculi (Charig *et al.* 1986. *BMJ* 292: 879-882)

- OS = open surgery (invasive)
- PN = percutaneous nephrolithotomy

Treatment group (j)	Failure (D_j)	Success (C_j)	Patients (N_j)	Failure-% (R_j)
OS ($j = 1$)	77	273	350	*22.0*
PN ($j = 0$)	60	290	350	*17.1*

Crude incidence proportions of treatment failure:

$$R_1 = 77/350 = 22.0\%, \qquad R_0 = 60/350 = 17.1\%$$

Risks and their comparative parameters

The **risk** or **probability** of binary outcome (*e.g.* new case of disease) in the `exposed` π_1 and in the `unexposed` π_0 as to binary risk factor X (values 1 and 0) are typically compared by

- risk difference $\theta = \pi_1 - \pi_0$
- risk ratio $\phi = \pi_1 / \pi_0$
- odds ratio (risk odds ratio)

$$\psi = \frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}$$

The odds ratio is close to the risk ratio when the risks are `small` (less than 0.1 – the `rare-disease assumption`).

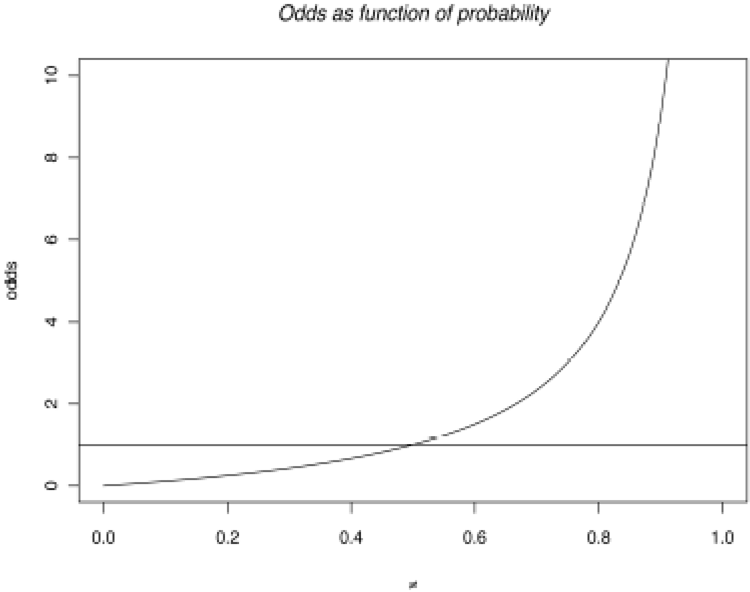
Odds and Odds Ratio (OR)

The **odds (Ω)** is the probability of binary outcome $P(Y = 1) = \pi$ divided by the the probability of binary outcome $P(Y = 0) = 1 - \pi$.

$$\Omega = \frac{\pi}{1 - \pi}$$

- Odds of 2.5 means that the probability of Y=1 (success) is two and half times higher than the probability of Y=0 (failure)
- Odds 0.5 means that success probability of success is 50% of the probability of failure
- Odds of 1 implies that probability of both outcomes 0.5 (equal)

Probability and odds



Risks and comparative parameters estimated

The risks π_1 and π_0 are estimated by empirical incidence proportions $R_1 = D_1/N_1$, and $R_0 = D_0/N_0$.

Crude estimates of comparative parameters

- **incidence proportion difference** $\mathrm{RD} = R_1 - R_0$
- **incidence proportion ratio** $\mathrm{RR} = R_1/R_0$
- **incidence odds ratio**

$$\mathrm{OR} = \frac{R_1/(1 - R_1)}{R_0/(1 - R_0)}$$

NB. To remove *confounding*, the estimated must be adjusted for relevant *confounders*.

Example: OS vs. PN (cont'd)

Crude estimates of true risk difference θ, risk ratio ϕ, and \ odds ratio ψ between OS and PN:

$$
\begin{aligned}
\mathrm{RD} &= \frac{77}{350} - \frac{60}{350} = 0.22 - 0.171 = +\mathbf{0.049}\ (+4.9\%) \\
\mathrm{RR} &= \frac{77/350}{60/350} = \frac{77/60}{350/350} = \frac{0.22}{0.171} = \mathbf{1.283} \\
\mathrm{OR} &= \frac{77/273}{60/290} = \frac{0.22/(1-0.22)}{0.171/(1-0.171)} = \mathbf{1.363}
\end{aligned}
$$

PN appears more successful than OS.

Is this (a) true, (b) due to bias, or (c) due to chance?

Example: OS vs. PN (cont'd)

Standard error of log(RR), 95% error factor (EF) of RR, and 95% CI for true risk ratio ϕ:

$$
\begin{aligned}
\mathrm{SEL} &= \sqrt{\frac{1}{73} + \frac{1}{60} - \frac{1}{350} - \frac{1}{350}} \\
&= \mathbf{0.1547}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{EF} &= \exp\{1.96 \times 0.1547\} \\
&= \mathbf{1.3543}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{CI} &= [1.2833/1.3543, 1.2833 \times 1.3543] \\
&= [\mathbf{0.9476}, \mathbf{1.7380}].
\end{aligned}
$$

Estimating comparative parameters in R

- A multitude of R functions in several packages are readily available for point estimation and CI calculation using either "exact" or/and various approximative methods.
- We shall here demonstrate the use of function `twoby2()` in the `Epi package`. It applies the simple Wald approximations as described above, but for
    - risk difference: the Newcombe method is used, and
    - odds ratio: the `exact` conditional method is also available.
- Hence, similar results are expected as obtained above.

Use of function `twoby2()`

- Loading the `Epi` package:

```
library(Epi)
```

- Reading the counts of the 2 x 2-table into a matrix:

```
counts <- c(77, 273, 60, 290)
tab <- matrix( counts, nrow=2, byrow=T)
```

- Viewing the contents of the matrix/table:

```
tab
```

```
     [,1] [,2]
[1,]   77  273
[2,]   60  290
```

Make 2 by 2 table

- Calling the function with `tab` as its argument:

```
twoby2(tab)
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Col 1
Comparing : Row 1 vs. Row 2

      Col 1 Col 2    P(Col 1) 95% conf. interval
Row 1    77   273      0.2200    0.1797   0.2664
Row 2    60   290      0.1714    0.1355   0.2146

                                   95% conf. interval
             Relative Risk: 1.2833    0.9476   1.7380
         Sample Odds Ratio: 1.3632    0.9362   1.9851
Conditional MLE Odds Ratio: 1.3626    0.9206   2.0237
    Probability difference: 0.0486   -0.0103   0.1071

             Exact P-value: 0.1272
        Asymptotic P-value: 0.1061
------------------------------------------------------
```

Analyses based on binary regression model

Crude estimates and CIs for the comparative parameters can also be obtained by fitting appropriate **binary regression models** for the numbers D_j or proportions R_j.

Special cases of **generalized linear models** (GLM) with

- (i) **random part**: D_j is assumed to obey the binomial distribution or $\{$**family**$\}$ **with index** N_j **and probability** π_j**,**
- (ii) **systematic part**: $\{$**linear predictor**$\}$ $\eta_j = \alpha + \beta X_j$**, in which** $X_j = 0$ **for unexposed and** $X_j = 1$ **for exposed,**
- (iii) **link function**: $g(.)$ that connects the probability π_j and the systematic part η_j by:

$$g(\pi_j) = \eta_j = \alpha + \beta X_j$$