Epidemiologic data analysis using R

Practicals 2

Faculty of Social Sciences, University of Tampere

–

Janne Pitkäniemi

26.02.2018

Topics of practical 2

Learning objectives of this practical

- crude and stratified analysis of proportions
- estimation of comparative parameters of risk using function *twoby2()* in package *Epi*
- crude estimation of comparative parameters by binomial regression using function *glm()*
- model-based adjustment for confounding and evaluation of effect modification on comparative parameters of risks.

Read data description:

The UGDP study, conducted in the USA during the 1970's, addressed the effects of tolbutamide (orally administered drug) in treating patients with Type 2 diabetes (T2D). Over 400 patients were randomized into tolbutamide and placebo groups, respectively. The follow-up lasted 5 years for all subjects, and was complete.

The following table displays the summary results concerning total mortality during the follow-up both overall and stratified by age of the patients.

		Tolbutamide	Placebo
Age <55 y	Dead	8	5
	Survived	98	115
	Group size	106	120
Age ≥55 y	Dead	22	16
	Survived	76	69
	Group size	98	85
All patients	Dead	30	21
	Survived	174	184
	Group size	204	205

1. Data for the analysis

Copy the following lines in order to create R dataset for following excercises

```
library(Epi)
library(data.table)

counts<-c( 8, 98, 5, 115,22, 76, 16, 69)
exposed<-c(1,1,0,0,1,1,0,0) #treatment=1, placebo=0
outcome<-c(1,0,1,0,1,0,1,0) # 1=death, 0=Alive
ageg<-c(0,0,0,0,1,1,1,1) # 0= <55 y , 1= >55 y

ugdp<-data.table(ageg, exposed, outcome, counts)
ugdp
```

2. 2by2 Table

Create a 2×2-matrix, e.g. with a name *tab22*, of counts from the '`All patients`'section of the table above, needed as the input for the *twoby2()* function. See thelecture notes, part 2, slide 19, how this was done for the OS *vs.* PN data. Be careful with the contents and order of rows and colums of the matrix.

In order to do this we summarize counts over two age groups and convert the data to matrix for *twoby2()* function.

```
tab22<-ugdp[,
            .(counts=sum(counts)),
            by=list(exposed,outcome)]
tab22
mat<-matrix(tab22$counts, nrow=2, byrow=T)
mat
```

or type numbers directly

```
library(Epi)
counts <- c(30, 174, 21, 184)
tab22 <- matrix(counts, nrow=2, byrow=T)
tab22
```

3. Risk estimates

Crude comparison of risks of death between the treatment groups. Create a 2×2-matrix, e.g. with a name *tab22*, of counts from the All patients section of the table above, needed as the input for the *twoby2()* function. See thelecture notes, part 2, slide 19, how this was done for the OS *vs.* PN data. Be careful with the contents and order of rows and colums of the matrix.

Call function *twoby2()* with the given input. Check, that the mortality proportions are 14.7% in the tolbutamide group and 10.2% in the placebo group. If not, return to previous item, make appropriate corrections to the input data, and try again.

Examine the results. Can you find the point estimates and 95% confidence intervals for the risk difference, risk ratio, and odds ratio parameters?

Use *twoby()* again but now to compute 90% confidence intervals for the parameters. Check how this is done from the help page: *?twoby2*

```
twoby2(tab22)
twoby2(tab22, alpha = 0.1)
```

4. RR, OR and RD

Compute crude estimates and confidence intervals of (a) risk ratio, (b) odds ratio and (c) risk difference, respectively, by regression modelling. Use function *glm()* to fit the corresponding generalized linear models on the mortality proportions by treatment group with appropriate specifications of *family* and *link*. Save the fitting results into model objects named *RRmod*, *ORmod*, and *RDmod*, respectively.Display the results of each model fitting using *ci.lin()* function. See lecture notes, slides 25-28, how this was done for the OS *vs.* PN data. % For the identity link you must use the trick shown on slide 28.

NB. If the 2×2 matrix of counts (named *e.g. tab22*) is correctly created in exercise 2., then the vectors containing the number of cases D, group sizes N, and mortality proportions P can be extracted from the columns of this matrix:

(a) Relative Risk

```
RRmod <- glm( outcome ~ exposed, family=binomial("log"), w = counts, data=ugdp)
summary(RRmod)
round(ci.lin( RRmod, Exp=T)[ , -(3:4)], 4)
round(exp( confint(RRmod)), 4)
```

(b) Odds Ratio

```
ORmod <- glm( outcome ~ exposed, family=binomial("logit"), w = counts, data=ugdp)
round(ci.lin( ORmod, Exp=T)[ , -(3:4)], 4)
round(exp( confint(ORmod)), 4)
```

(b) Risk Difference

```
RDmod <- glm( outcome ~ exposed, family=binomial(link="identity"), w = counts,data=ugdp)
summary(RDmod)$coef
round(ci.lin( RDmod )[ , -(3:4)], 4)
confint(RDmod)
```

5. Stratified analysis

Compute the point estimates and 95% confidence intervals for the values of the three comparative parameters within both age strata separately using function *twoby2()*, as you did in exercise 2. for all patients. What observations do you make on the heterogeneity of stratum-specific results, and what may be conluded about the homogeneity of the comparative parameters?

```
lt55<-matrix(ugdp[1:4,]$counts, nr=2, byrow=T)
twoby2(lt55)

ge55<-matrix(ugdp[5:8,]$counts, nr=2, byrow=T)
twoby2(ge55)
```

6. Effect modification - Risk Difference

Perform a model-based adjustment for confounding and evaluation of effect-modification (interaction) by age group on the risk difference between tolbutamide and placebo. See slides 37-43.

```
RDmod2 <- glm(PP ~ exposed+ ageg, family=binomial(link="identity"), w=NN,data=ugdp2)
round(ci.lin( RDmod2 )[ , -(3:4)], 4)
confint(RDmod2) # profile-likelihood based confidence intervals
RDmod3 <- glm(PP ~ exposed*ageg, family=binomial(link="identity"), w=NN,data=ugdp2)
round(ci.lin( RDmod3 )[ , -(3:4)], 4)
```

7. Effect modification RR and OR

Perform a model-based adjustment for confounding and evaluation of effect-modification (interaction) by age group on (a) the risk ratio, and (b) the odds ratio, respectively, between tolbutamide and placebo.

(a) RR

```
RRmod2 <- glm(PP ~ exposed+ ageg, fam=binomial("log"), w=NN,data=ugdp2)
round(ci.lin( RRmod2, Exp=T )[ , -(3:4)], 4)
exp(cbind(coef(RRmod2), confint(RRmod2)))
RRmod3 <- glm(PP ~ exposed*ageg, fam=binomial("log"), w=NN,data=ugdp2)
round(ci.lin( RRmod3, Exp=T )[ , -(3:4)], 4)
```

(b) OR

```
ORmod2 <- glm(PP ~ exposed + ageg, fam=binomial("logit"), w=NN,data=ugdp2)
round(ci.lin( ORmod2, Exp=T )[ , -(3:4)], 4)
exp(cbind(coef(ORmod2), confint(ORmod2)))
ORmod3 <- glm(PP ~ exposed*ageg, fam=binomial("logit"), w=NN,data=ugdp2)
round(ci.lin( ORmod3, Exp=T )[ , -(3:4)], 4)
```