Epidemiologic data analysis using R

Practicals 2

Faculty of Social Sciences, University of Tampere

–

Janne Pitkäniemi

14.03.2018

# 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.

## 1. Data for the analysis

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

library(Epi)

Attaching package: 'Epi'

The following object is masked from 'package:base':  
  
 merge.data.frame

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

ageg exposed outcome counts  
1: 0 1 1 8  
2: 0 1 0 98  
3: 0 0 1 5  
4: 0 0 0 115  
5: 1 1 1 22  
6: 1 1 0 76  
7: 1 0 1 16  
8: 1 0 0 69

## 2. 2by2 Table

Create a -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

exposed outcome counts  
1: 1 1 30  
2: 1 0 174  
3: 0 1 21  
4: 0 0 184

mat<-matrix(tab22$counts, nrow=2, byrow=T)  
mat

[,1] [,2]  
[1,] 30 174  
[2,] 21 184

or type numbers directly

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

[,1] [,2]  
[1,] 30 174  
[2,] 21 184

## 3. Risk estimates

Crude comparison of risks of death between the treatment groups. Create a -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)

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 30 174 0.1471 0.1048 0.2026  
Row 2 21 184 0.1024 0.0677 0.1520  
  
 95% conf. interval  
 Relative Risk: 1.4356 0.8510 2.4216  
 Sample Odds Ratio: 1.5107 0.8333 2.7387  
Conditional MLE Odds Ratio: 1.5091 0.8014 2.8873  
 Probability difference: 0.0446 -0.0200 0.1096  
  
 Exact P-value: 0.1813   
 Asymptotic P-value: 0.1741   
------------------------------------------------------

twoby2(tab22, alpha = 0.1)

2 by 2 table analysis:   
------------------------------------------------------   
Outcome : Col 1   
Comparing : Row 1 vs. Row 2   
  
 Col 1 Col 2 P(Col 1) 90% conf. interval  
Row 1 30 174 0.1471 0.1108 0.1927  
Row 2 21 184 0.1024 0.0725 0.1429  
  
 90% conf. interval  
 Relative Risk: 1.4356 0.9257 2.2264  
 Sample Odds Ratio: 1.5107 0.9169 2.4889  
Conditional MLE Odds Ratio: 1.5091 0.8802 2.6145  
 Probability difference: 0.0446 -0.0094 0.0989  
  
 Exact P-value: 0.1813   
 Asymptotic P-value: 0.1741   
------------------------------------------------------

## 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 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:

1. Relative Risk

RRmod <- glm( outcome ~ exposed, family=binomial("log"), w = counts, data=ugdp)  
summary(RRmod)

Call:  
glm(formula = outcome ~ exposed, family = binomial("log"), data = ugdp,   
 weights = counts)  
  
Deviance Residuals:   
 Min 1Q Median 3Q Max   
-5.5836 -4.9342 0.4557 6.2883 9.1839   
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -2.2785 0.2067 -11.022 <2e-16 \*\*\*  
exposed 0.3616 0.2668 1.355 0.175   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 307.71 on 7 degrees of freedom  
Residual deviance: 305.84 on 6 degrees of freedom  
AIC: 309.84  
  
Number of Fisher Scoring iterations: 7

round(ci.lin( RRmod, Exp=T)[ , -(3:4)], 4)

Estimate StdErr exp(Est.) 2.5% 97.5%  
(Intercept) -2.2785 0.2067 0.1024 0.0683 0.1536  
exposed 0.3616 0.2668 1.4356 0.8510 2.4216

round(exp( confint(RRmod)), 4)

Waiting for profiling to be done...

2.5 % 97.5 %  
(Intercept) 0.0659 0.1488  
exposed 0.8562 2.4578

1. Odds Ratio

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

Estimate StdErr exp(Est.) 2.5% 97.5%  
(Intercept) -2.1704 0.2303 0.1141 0.0727 0.1793  
exposed 0.4126 0.3035 1.5107 0.8333 2.7387

round(exp( confint(ORmod)), 4)

Waiting for profiling to be done...

2.5 % 97.5 %  
(Intercept) 0.0706 0.1749  
exposed 0.8375 2.7696

1. Risk Difference

RDmod <- glm( outcome ~ exposed, family=binomial(link="identity"), w = counts,data=ugdp)  
summary(RDmod)$coef

Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.1024390 0.02117814 4.837018 1.318018e-06  
exposed 0.0446198 0.03260949 1.368307 1.712159e-01

round(ci.lin( RDmod )[ , -(3:4)], 4)

Estimate StdErr 2.5% 97.5%  
(Intercept) 0.1024 0.0212 0.0609 0.1439  
exposed 0.0446 0.0326 -0.0193 0.1085

confint(RDmod)

Waiting for profiling to be done...

2.5 % 97.5 %  
(Intercept) 0.06590969 0.1488359  
exposed -0.01945402 0.1095387

## 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)

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 8 98 0.0755 0.0382 0.1437  
Row 2 5 115 0.0417 0.0174 0.0962  
  
 95% conf. interval  
 Relative Risk: 1.8113 0.6112 5.3679  
 Sample Odds Ratio: 1.8776 0.5949 5.9260  
Conditional MLE Odds Ratio: 1.8723 0.5204 7.5212  
 Probability difference: 0.0338 -0.0300 0.1043  
  
 Exact P-value: 0.3919   
 Asymptotic P-value: 0.2827   
------------------------------------------------------

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

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 22 76 0.2245 0.1526 0.3175  
Row 2 16 69 0.1882 0.1186 0.2854  
  
 95% conf. interval  
 Relative Risk: 1.1926 0.6713 2.1188  
 Sample Odds Ratio: 1.2484 0.6066 2.5692  
Conditional MLE Odds Ratio: 1.2468 0.5723 2.7647  
 Probability difference: 0.0363 -0.0832 0.1513  
  
 Exact P-value: 0.5874   
 Asymptotic P-value: 0.5469   
------------------------------------------------------

## 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(cbind(outcome,counts) ~ exposed+ ageg, family=binomial(link="identity"), data=ugdp)  
round(ci.lin(RDmod2,Exp=F)[,c(1,5:6)],4) # wald-likelihood based confidence intervals

Estimate 2.5% 97.5%  
(Intercept) 0.0087 -0.0062 0.0237  
exposed 0.0000 -0.0188 0.0189  
ageg 0.0020 -0.0172 0.0213

RDmod3 <- glm(cbind(outcome,counts) ~ exposed\*ageg, family=binomial(link="identity"), data=ugdp)  
round(ci.lin(RDmod3,Exp=F)[,c(1,5:6)],4) # wald-likelihood based

Estimate 2.5% 97.5%  
(Intercept) 0.0083 -0.0079 0.0244  
exposed 0.0011 -0.0233 0.0254  
ageg 0.0034 -0.0244 0.0312  
exposed:ageg -0.0026 -0.0413 0.0360

## 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.

1. RR

RDmod3 <- glm(cbind(outcome,counts) ~ exposed\*ageg, family=binomial(link="log"), data=ugdp)  
round(ci.lin(RDmod3,Exp=F)[,c(1,5:6)],4) # wald-likelihood based

Estimate 2.5% 97.5%  
(Intercept) -4.7958 -6.7476 -2.8439  
exposed 0.1230 -2.6366 2.8825  
ageg 0.3414 -2.4165 3.0994  
exposed:ageg -0.2637 -4.1643 3.6369

1. OR

RDmod3 <- glm(cbind(outcome,counts) ~ exposed\*ageg, family=binomial(link="logit"), data=ugdp)  
round(ci.lin(RDmod3,Exp=F)[,c(1,5:6)],4) # wald-likelihood based

Estimate 2.5% 97.5%  
(Intercept) -4.7875 -6.7553 -2.8197  
exposed 0.1241 -2.6598 2.9079  
ageg 0.3448 -2.4407 3.1303  
exposed:ageg -0.2664 -4.2056 3.6728