Practical: Effect Estimates Binary Data

Alice Kamau

18 Feb 2016

We will use various R packages to compute these measures such as epicalc, Epi, epiR, and epitools to obtain effet estimates for binary data

Part 1: Load the dataset

We will use the lung capacity dataset to compute effect estimates for binary data (Risk Ratio and Odds Ratio)

Load the *LungCapData.csv* dataset into memory.

```
LungCapData<-read.csv("/Users/akamau/Desktop/Stats forum/Stat training/LungCapData.csv", header=TRUE)
```

Part 2: Explore the dataset

```
attach(LungCapData)
names(LungCapData)
class(Gender)
levels(Gender)
class(Smoke)
levels(Smoke)
```

Explore the relationship between various variables

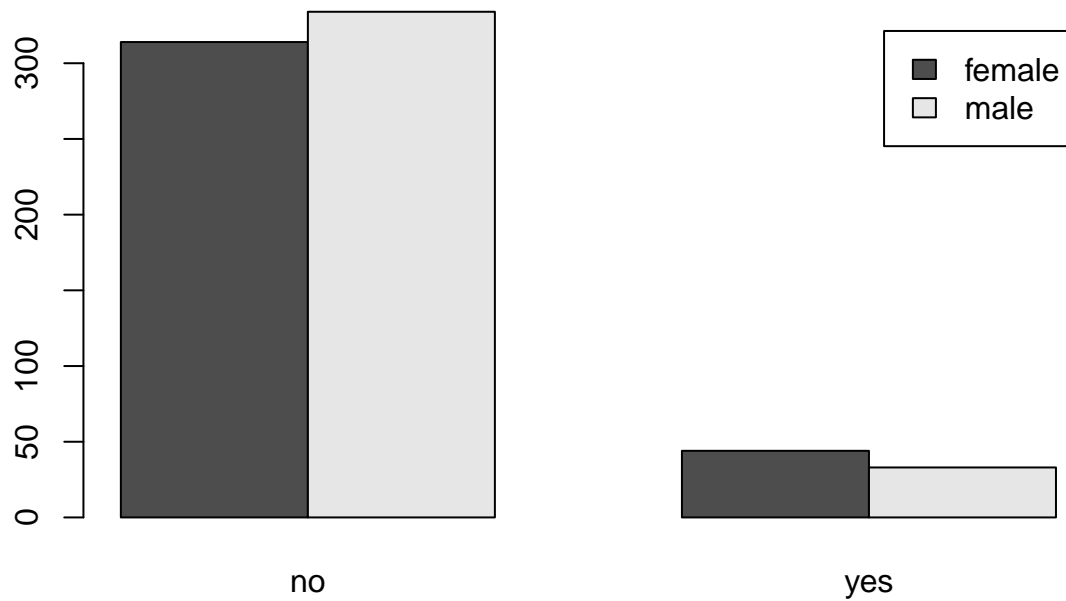Load the R packages 'ggplot2', 'lme4', 'descr'.
You can use the help function to determine what each package does e.g. ?ggplot

```
library(ggplot2)
library(lme4)
library(descr)
```

```
tab<-table(Gender, Smoke)
tab
```

```
##         Smoke
## Gender    no yes
##   female 314  44
##   male   334  33
```

```
barplot(tab, beside=T, legend=T)
```

```r
chisq.test(tab, correct = F)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab
## X-squared = 2.0773, df = 1, p-value = 0.1495
```

```r
fisher.test(tab, conf.int=T, conf.level=0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  tab
## p-value = 0.1845
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.4233701 1.1659426
## sample estimates:
## odds ratio
##  0.7054345
```

Part 3: Load the R package 'epitools', 'epiR', 'abd' to calculate the RR and OR

```r
library(epitools)
library(epiR)
library(abd)
library(epiDisplay)
```

Part 4: Calculate the Risk Ratio and Odds Ratios

Recall the table 'tab'. We need to set it in the standard abcd format.The standard format dictates that the exposure be presented in rows while the outcome be presented in columns

```
tab2<-cbind(tab[,2],tab[,1])
colnames(tab2)<-c("yes","no")
tab2
```

```
##        yes  no
## female  44 314
## male    33 334
```

```
fisher.test(tab2, conf.int=T, conf.level=0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  tab2
## p-value = 0.1845
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8576752 2.3619999
## sample estimates:
## odds ratio
##   1.417566
```

Using epitools package

```
epitab(tab2, method="oddsratio")
```

```
## $tab
##        yes        p0  no        p1 oddsratio     lower    upper   p.value
## female  44 0.5714286 314 0.4845679  1.000000        NA       NA        NA
## male    33 0.4285714 334 0.5154321  1.418259 0.8803057 2.284955 0.1845426
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```

Using epiR package

```r
epi.2by2(tab2, method="cohort.count",conf.level=0.95)
```

```
##              Outcome +   Outcome -     Total     Inc risk *
## Exposed +          44         314       358          12.29
## Exposed -          33         334       367           8.99
## Total              77         648       725          10.62
##                 Odds
## Exposed +      0.1401
## Exposed -      0.0988
## Total          0.1188
##
## Point estimates and 95 % CIs:
## -------------------------------------------------------------
## Inc risk ratio (W)                        1.37 (0.89, 2.10)
## Odds ratio (W)                            1.42 (0.88, 2.28)
## Attrib risk (W) *                         3.30 (-1.19, 7.79)
## Attrib risk in population (W) *           1.63 (-2.06, 5.32)
## Attrib fraction in exposed (%)            26.84 (-12.15, 52.28)
## Attrib fraction in population (%)         15.34 (-8.10, 33.69)
## -------------------------------------------------------------
##  X2 test statistic: 2.077 p-value: 0.15
##  W: Wald confidence limits
##  * Cases per 100 population units
```

Part 5: Interpretation

The odds of smoking among female is 1.42 times higher than the odds of smoking among male

Refer to the example on the relationship between lung cancer and smoking **

**Using the example used in the slides to compute effect estimates using R **

```r
smoking <- matrix(c(709, 154, 142, 308), nrow = 2)
rownames(smoking) <- c("smokers", "non-smokers")
colnames(smoking) <- c("lungcancer", "no-lung cancer")


smoking
```

```
##             lungcancer no-lung cancer
## smokers            709            142
## non-smokers        154            308
```

```r
epi.2by2(smoking, method="cohort.count",conf.level=0.95)
```

```
##               Outcome +      Outcome -       Total        Inc risk *
## Exposed +          709            142          851              83.3
## Exposed -          154            308          462              33.3
## Total              863            450         1313              65.7
##                   Odds
## Exposed +         4.99
## Exposed -         0.50
## Total             1.92
##
## Point estimates and 95 % CIs:
## ---------------------------------------------------------
## Inc risk ratio (W)                           2.50 (2.19, 2.85)
## Odds ratio (W)                               9.99 (7.67, 13.01)
## Attrib risk (W) *                            49.98 (45.01, 54.96)
## Attrib risk in population (W) *              32.39 (27.39, 37.40)
## Attrib fraction in exposed (%)               59.99 (54.33, 64.95)
## Attrib fraction in population (%)            49.29 (43.36, 54.59)
## ---------------------------------------------------------
##  X2 test statistic: 332.057 p-value: < 0.001
##  W: Wald confidence limits
##  * Cases per 100 population units
```

```r
epitab(smoking, method="oddsratio", rev="both")
```

```
## $tab
##              no-lung cancer        p0 lungcancer          p1 oddsratio
## non-smokers             308 0.6844444       154 0.1784473  1.000000
## smokers                 142 0.3155556       709 0.8215527  9.985915
##              lower    upper       p.value
## non-smokers     NA       NA            NA
## smokers    7.66614 13.00766 4.826448e-74
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```

Part 6: Assignment

1. Evaluate the birthweight2 data

2. Look at the association between LBW and gestational weeks. Divide gestwks into quartiles and analyse as groups.

3. Check the Odds ratio for the association between LBW and ethinicity

4. Finally look at LBW and sex, maternal age, height.

5. Make Conclusion