Biostatistics: Exercise 10

Beate Sick, Lisa Herzog

17.11.2020

Exercise 1: Crude and adjusted OR

We are using the data from the case control study on coffee and pancreatic cancer from the lecture. The dataset coffee.csv can be downloaded from the website.

```
# read in the data
coffee = read.csv("data/coffee.csv", sep=";", header=TRUE)
head(coffee)
```

```
##   case sex coffee
## 1    0   0      1
## 2    0   0      1
## 3    0   0      1
## 4    0   0      1
## 5    0   0      1
## 6    0   0      1
```

- Compute the crude OR and the corresponding 95% CI with the function *fisher.test*() measuring the influence of coffee drinking on pancreatic cancer. Test additionally for an association with the χ^2-Test.

```
# consider contingency table
tab = table(coffee$case, coffee$coffee)
tab
##
##       0   1
##   0  88 555
##   1  20 347

# Crude OR and CI with fishers test
fisher.test(tab)
##
##  Fisher's Exact Test for Count Data
##
## data:  tab
## p-value = 3.028e-05
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.641006 4.806538
## sample estimates:
## odds ratio
##   2.748557

# Investigate the influence of coffee on pancreatic cancer incidence
chisq.test(tab)
##
```

```
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tab
## X-squared = 15.746, df = 1, p-value = 7.245e-05
# There is high evidence for an association between coffee drinking and pancreatic cancer
```

- Compute the crude OR and the corresponding 95% CI with a logistic regression model measuring the influence of coffee drinking on pancreatic cancer (adjust for coffee only). Interpret your result.

```
# Crude OR and CI with logistic regression
mod = glm(case~coffee, family=binomial, data=coffee)
summary(mod)
##
## Call:
## glm(formula = case ~ coffee, family = binomial, data = coffee)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9855  -0.9855  -0.9855   1.3822   1.8365
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4816     0.2477  -5.981 2.22e-09 ***
## coffee        1.0120     0.2570   3.938 8.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1323.8  on 1009  degrees of freedom
## Residual deviance: 1305.5  on 1008  degrees of freedom
## AIC: 1309.5
##
## Number of Fisher Scoring iterations: 4
exp(coef(mod)["coffee"]) # OR
##   coffee
## 2.750991

# confidence interal with confint()
exp(confint(mod))
## Waiting for profiling to be done...
##                 2.5 %    97.5 %
## (Intercept) 0.1360523 0.3612225
## coffee      1.6968298 4.6709088

# confidence interval manually
(log_OR = as.vector(coef(mod)["coffee"]))
## [1] 1.011961
(log_OR_se = summary(mod)$coef["coffee","Std. Error"])
## [1] 0.2569966
(CI_log_OR = c(log_OR - 1.96*log_OR_se, log_OR + 1.96*log_OR_se))
## [1] 0.5082478 1.5156746
(CI_OR = exp(CI_log_OR))
## [1] 1.662376 4.552491
```

```
# Interpretation: The risk for pancreatic cancer incidence
# is 2.75 times higher for patients drinking coffee compared
# to patients drinking no coffee
```

- Compute the sex adjusted OR and the corresponding 95% CI. Apply a logistic regression.

```
# Adjusted OR and CI with logistic regression
mod = glm(case~coffee+sex, family=binomial, data=coffee)
summary(mod)
##
## Call:
## glm(formula = case ~ coffee + sex, family = binomial, data = coffee)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0588  -1.0588  -0.9012   1.3009   1.9099
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2423     0.2595  -4.788 1.69e-06 ***
## coffee        0.9566     0.2583   3.704 0.000213 ***
## sex          -0.4055     0.1335  -3.038 0.002378 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1323.8  on 1009  degrees of freedom
## Residual deviance: 1296.2  on 1007  degrees of freedom
## AIC: 1302.2
##
## Number of Fisher Scoring iterations: 4
exp(coef(mod)["coffee"])
##   coffee
## 2.602932

# CI for OR with confint()
exp(confint(mod))
## Waiting for profiling to be done...
##                 2.5 %    97.5 %
## (Intercept) 0.1693146 0.4706545
## coffee      1.6008444 4.4293374
## sex         0.5127257 0.8653701

# Ci for OR manually
(log_OR = as.vector(coef(mod)["coffee"]))
## [1] 0.9566385
(log_OR_se = summary(mod)$coef["coffee","Std. Error"])
## [1] 0.2583015
(CI_log_OR = c(log_OR - 1.96*log_OR_se, log_OR + 1.96*log_OR_se))
## [1] 0.4503676 1.4629094
(CI_OR = exp(CI_log_OR))
## [1] 1.568889 4.318506
```

- Compare the crude and the adjusted OR (and their CIs). Is sex a confounder?

```
# Sex doesn't seem to be a confounder since the confounding effect is not larger than 10%.
```