IEHC0046 BASIC STATISTICS FOR MEDICAL SCIENCES

Analysis of Categorical Data II: Practical

08 November, 2020

In this practical we will use R to calculate chi-squared test and odds ratios using the elsa dataset.

Remember to use a script to save your code and to change your working directory so you can load the ELSA dataset easily.

load("elsa.Rdata")

We’ll be using the tidyverse and summarytools packages, as well as a new package, mStats. Load the packages - and install, if necessary.

# install.packages("mStats") # uncomment if you need to install!  
library(tidyverse)  
library(summarytools)  
library(mStats)

Most of the social and demographic variables and health behaviours were measured at the baseline of the ELSA study, while incident events were recorded at the end of longitudinal follow-up. In this session, we are interested in the association between some social and demographic variables and health behaviours and incident heart attacks.

**Q. Look at the variable heart\_attack. How many events are there?**

Use the freq() function from summarytools (or table() from Base R).

freq(elsa$heart\_attack)

## Frequencies   
## elsa$heart\_attack   
## Label: incident heart attack   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ------------------- ------ --------- -------------- --------- --------------  
## Not mentioned 2903 92.84 92.84 92.78 92.78  
## Mentioned 224 7.16 100.00 7.16 99.94  
## <NA> 2 0.06 100.00  
## Total 3129 100.00 100.00 100.00 100.00

**Q. Look at the frequency distribution of heart\_attack by sex, age groups, smoking(smok\_bin), and occupational status (manual). How do you interpret the results?**

Let’s create a variable with ten year age groups. We can do this using the cut\_width() function from ggplot2 which is part of the tidyverse.

elsa <- elsa %>%  
 mutate(age\_group = cut\_width(age, width = 10, centre = 5))  
by(elsa$age, elsa$age\_group, summary) # To check worked correctly

## elsa$age\_group: [45,55]  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 45.00 48.00 51.00 50.56 53.00 55.00   
## ------------------------------------------------------------   
## elsa$age\_group: (55,65]  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 56.00 58.00 61.00 60.44 63.00 65.00   
## ------------------------------------------------------------   
## elsa$age\_group: (65,75]  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 66.00 67.00 70.00 69.94 72.00 75.00   
## ------------------------------------------------------------   
## elsa$age\_group: (75,85]  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 76.00 77.00 78.00 78.47 80.00 85.00   
## ------------------------------------------------------------   
## elsa$age\_group: (85,95]  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 86.00 86.25 87.50 87.80 89.00 90.00

table(elsa$age\_group)

##   
## [45,55] (55,65] (65,75] (75,85] (85,95]   
## 1216 986 679 238 10

There are only 10 individuals in the upper-age group categoriy (85-95). Let’s collapse the 75-85 and 85-95 age groups into a single category using the fct\_recode() function from forcats (part of the tidyverse).

elsa <- elsa %>%  
 mutate(age\_group = fct\_recode(age\_group, "(75, 95]" = "(75,85]", "(75, 95]" = "(85,95]"))

Now, to cross-tabulate the variables, use the ctable() function from summarytools.

ctable(elsa$sex, elsa$heart\_attack)

## Cross-Tabulation, Row Proportions   
## sex \* heart\_attack   
## Data Frame: elsa   
##   
## -------- -------------- --------------- ------------- ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## sex   
## male 1215 (88.8%) 152 (11.1%) 1 (0.1%) 1368 (100.0%)  
## female 1688 (95.9%) 72 ( 4.1%) 1 (0.1%) 1761 (100.0%)  
## Total 2903 (92.8%) 224 ( 7.2%) 2 (0.1%) 3129 (100.0%)  
## -------- -------------- --------------- ------------- ---------- ---------------

ctable(elsa$age\_group, elsa$heart\_attack)

## Cross-Tabulation, Row Proportions   
## age\_group \* heart\_attack   
## Data Frame: elsa   
##   
## ----------- -------------- --------------- ------------- ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## age\_group   
## [45,55] 1176 (96.7%) 40 ( 3.3%) 0 (0.0%) 1216 (100.0%)  
## (55,65] 905 (91.8%) 81 ( 8.2%) 0 (0.0%) 986 (100.0%)  
## (65,75] 602 (88.7%) 75 (11.0%) 2 (0.3%) 679 (100.0%)  
## (75, 95] 220 (88.7%) 28 (11.3%) 0 (0.0%) 248 (100.0%)  
## Total 2903 (92.8%) 224 ( 7.2%) 2 (0.1%) 3129 (100.0%)  
## ----------- -------------- --------------- ------------- ---------- ---------------

ctable(elsa$smok\_bin, elsa$heart\_attack)

## Cross-Tabulation, Row Proportions   
## smok\_bin \* heart\_attack   
## Data Frame: elsa   
##   
## ---------------- -------------- --------------- ------------ ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## smok\_bin   
## never/ex occ 1316 (95.7%) 59 (4.3%) 0 (0.0%) 1375 (100.0%)  
## ex/current reg 1587 (90.5%) 165 (9.4%) 2 (0.1%) 1754 (100.0%)  
## Total 2903 (92.8%) 224 (7.2%) 2 (0.1%) 3129 (100.0%)  
## ---------------- -------------- --------------- ------------ ---------- ---------------

ctable(elsa$manual, elsa$heart\_attack)

## Cross-Tabulation, Row Proportions   
## manual \* heart\_attack   
## Data Frame: elsa   
##   
## --------- -------------- --------------- ------------- ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## manual   
## non man 1734 (94.0%) 109 ( 5.9%) 1 (0.1%) 1844 (100.0%)  
## manual 1107 (91.0%) 108 ( 8.9%) 1 (0.1%) 1216 (100.0%)  
## <NA> 62 (89.9%) 7 (10.1%) 0 (0.0%) 69 (100.0%)  
## Total 2903 (92.8%) 224 ( 7.2%) 2 (0.1%) 3129 (100.0%)  
## --------- -------------- --------------- ------------- ---------- ---------------

**Q. Next, look at the association between sexand heart attacksin more details. Calculate chi square test and draw your conclusions about the association.**

To get the chi-square test, we can change the chisq argument in the ctable() function to TRUE.

ctable(elsa$sex, elsa$heart\_attack, chisq = TRUE)

## Cross-Tabulation, Row Proportions   
## sex \* heart\_attack   
## Data Frame: elsa   
##   
##   
## -------- -------------- --------------- ------------- ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## sex   
## male 1215 (88.8%) 152 (11.1%) 1 (0.1%) 1368 (100.0%)  
## female 1688 (95.9%) 72 ( 4.1%) 1 (0.1%) 1761 (100.0%)  
## Total 2903 (92.8%) 224 ( 7.2%) 2 (0.1%) 3129 (100.0%)  
## -------- -------------- --------------- ------------- ---------- ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 56.1 1 0   
## ----------------------------

The chi-square test statistic is large and p < 0.01. This is very strong evidence against the null hypothesis of no association between sex and incident heart attack. Heart attacks are more common among males than females.

**Q. Evaluate the association between smoking, occupational status and heart attacksusing chi-squared test.**

We can swap sex for manual and smok\_bin in the above code.

ctable(elsa$manual, elsa$heart\_attack, chisq = TRUE)

## Cross-Tabulation, Row Proportions   
## manual \* heart\_attack   
## Data Frame: elsa   
##   
##   
## --------- -------------- --------------- ------------- ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## manual   
## non man 1734 (94.0%) 109 ( 5.9%) 1 (0.1%) 1844 (100.0%)  
## manual 1107 (91.0%) 108 ( 8.9%) 1 (0.1%) 1216 (100.0%)  
## <NA> 62 (89.9%) 7 (10.1%) 0 (0.0%) 69 (100.0%)  
## Total 2903 (92.8%) 224 ( 7.2%) 2 (0.1%) 3129 (100.0%)  
## --------- -------------- --------------- ------------- ---------- ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 9.382 1 0.0022   
## ----------------------------

ctable(elsa$smok\_bin, elsa$heart\_attack, chisq = TRUE)

## Cross-Tabulation, Row Proportions   
## smok\_bin \* heart\_attack   
## Data Frame: elsa   
##   
##   
## ---------------- -------------- --------------- ------------ ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## smok\_bin   
## never/ex occ 1316 (95.7%) 59 (4.3%) 0 (0.0%) 1375 (100.0%)  
## ex/current reg 1587 (90.5%) 165 (9.4%) 2 (0.1%) 1754 (100.0%)  
## Total 2903 (92.8%) 224 (7.2%) 2 (0.1%) 3129 (100.0%)  
## ---------------- -------------- --------------- ------------ ---------- ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 29.68 1 0   
## ----------------------------

The p-values in both chi-square tests were low. There is evidence that heart attacks are more common among smokers and among individuals in manual jobs.

**Q. Evaluate the association between age using 10-years age groups (age10)and heart attacks. Is there any difference in the interpretation of your findings when you have exposure with 4 categories rather than just two?**

ctable(elsa$age\_group, elsa$heart\_attack, chisq = TRUE)

## Cross-Tabulation, Row Proportions   
## age\_group \* heart\_attack   
## Data Frame: elsa   
##   
##   
## ----------- -------------- --------------- ------------- ---------- ---------------  
## heart\_attack Not mentioned Mentioned <NA> Total  
## age\_group   
## [45,55] 1176 (96.7%) 40 ( 3.3%) 0 (0.0%) 1216 (100.0%)  
## (55,65] 905 (91.8%) 81 ( 8.2%) 0 (0.0%) 986 (100.0%)  
## (65,75] 602 (88.7%) 75 (11.0%) 2 (0.3%) 679 (100.0%)  
## (75, 95] 220 (88.7%) 28 (11.3%) 0 (0.0%) 248 (100.0%)  
## Total 2903 (92.8%) 224 ( 7.2%) 2 (0.1%) 3129 (100.0%)  
## ----------- -------------- --------------- ------------- ---------- ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 51.03 3 0   
## ----------------------------

We have 4 categories of age but this does not change how we should interpret results of chi-square test. Again, we have a chi-square statistic and p<0.001. This is strong evidence against null hypothesis. In other words, there is strong evidence of association between age and incidence of heart attacks.

**Q. Let’s look again at the association between sex and heart attacks. Calculate OR (and 95% CI) to see the magnitude of the effect of sex on heart attacks.**

We’ll use two functions from the mStats package: tabOdds() and mhor(). We’ll use tabOdds() to get the odds (+ 95% CIs) of heart attack in each sex. We’ll use mhor() to get the odds ratios.

We’ll need to set the argument na.rm to TRUE as there is some missing data in the sex and heart\_attack variables.

tabOdds(elsa, sex, by = heart\_attack, plot = FALSE, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Estimates of Risks of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## sex | Not.mentioned Mentioned Odds [95% Conf. Interval]  
## ------ + ------------- --------- ------ ---------- ---------  
## male | 1215 152 7.993 7.261 8.800  
## female | 1688 72 23.444 21.296 25.809  
## ------ + ------------- --------- ------ ---------- ---------  
## Total | 2903 224 12.960 11.772 14.267  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (sex: Sex)  
## (heart\_attack: incident heart attack)

mhor(elsa, sex, by = heart\_attack, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## sex | Not mentioned Mentioned Odds Ratio [95% Conf. Interval] P>|z|  
## ------ + ------------- --------- ---------- ---------- --------- -----  
## male | 1215 152 0.341 0.255 0.456 0.000  
## female | 1688 72   
## ------ + ------------- --------- ---------- ---------- --------- -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (sex: Sex)  
## (heart\_attack: incident heart attack)

Note, because of how the heart\_attack factor is ordered, the functions have given us the odds of **not** having a heart attack in each sex. If we want to change this, we can use the case\_value argument in the tabOdds() and mhor() functions.

tabOdds(elsa, sex, by = heart\_attack, case\_value = 'Mentioned',  
 plot = FALSE, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Estimates of Risks of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## sex | Mentioned Not.mentioned Odds [95% Conf. Interval]  
## ------ + --------- ------------- ----- ---------- ---------  
## male | 152 1215 0.125 0.114 0.138  
## female | 72 1688 0.043 0.039 0.047  
## ------ + --------- ------------- ----- ---------- ---------  
## Total | 224 2903 0.077 0.070 0.085  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (sex: Sex)  
## (heart\_attack: incident heart attack)

mhor(elsa, sex, by = heart\_attack, case\_value = 'Mentioned', na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## sex | Mentioned Not mentioned Odds Ratio [95% Conf. Interval] P>|z|  
## ------ + --------- ------------- ---------- ---------- --------- -----  
## male | 152 1215 2.933 2.195 3.919 0.000  
## female | 72 1688   
## ------ + --------- ------------- ---------- ---------- --------- -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (sex: Sex)  
## (heart\_attack: incident heart attack)

The odds of heart attack are very different by sex, and 95% CIs do not overlap. Men are 2.93 times more likely to have incident heart attack than women. The 95% confidence interval is between 2.19 and 3.93, with a p-value < 0.001.

**Q. Now, do the same with smok\_bin and manual.**

tabOdds(elsa, smok\_bin, by = heart\_attack, case\_value = 'Mentioned', plot = FALSE, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Estimates of Risks of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## smok\_bin | Mentioned Not.mentioned Odds [95% Conf. Interval]  
## -------------- + --------- ------------- ----- ---------- ---------  
## never/ex occ | 59 1316 0.045 0.041 0.049  
## ex/current reg | 165 1587 0.104 0.094 0.114  
## -------------- + --------- ------------- ----- ---------- ---------  
## Total | 224 2903 0.077 0.070 0.085  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (smok\_bin: Binary smoking status)  
## (heart\_attack: incident heart attack)

mhor(elsa, smok\_bin, by = heart\_attack, case\_value = 'Mentioned', na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## smok\_bin | Mentioned Not mentioned Odds Ratio [95% Conf. Interval] P>|z|  
## -------------- + --------- ------------- ---------- ---------- --------- -----  
## never/ex occ | 59 1316 0.431 0.317 0.586 0.000  
## ex/current reg | 165 1587   
## -------------- + --------- ------------- ---------- ---------- --------- -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (smok\_bin: Binary smoking status)  
## (heart\_attack: incident heart attack)

Similar interpretation as before – odds of heart attack are 0.045 among non-smokers and 0.104 among smokers, OR 2.32 for smokers compared to non-smokers: smokers are 2.32 times more likely to have incident heart attack than non smokers. Confidence interval for OR 1.70-3.15, farfrom 1.00, p<0.001, strong evidence against null hypothesis, supporting evidence of association between smoking and incident heart attacks.

tabOdds(elsa, manual, by = heart\_attack, case\_value = 'Mentioned',  
 plot = FALSE, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Estimates of Risks of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## manual | Mentioned Not.mentioned Odds [95% Conf. Interval]  
## ------- + --------- ------------- ----- ---------- ---------  
## non man | 109 1734 0.063 0.057 0.069  
## manual | 108 1107 0.098 0.088 0.108  
## ------- + --------- ------------- ----- ---------- ---------  
## Total | 217 2841 0.076 0.069 0.084  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (manual: Manual vs non-manual sclass)  
## (heart\_attack: incident heart attack)

mhor(elsa, manual, by = heart\_attack, case\_value = 'Mentioned', na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## manual | Mentioned Not mentioned Odds Ratio [95% Conf. Interval] P>|z|  
## ------- + --------- ------------- ---------- ---------- --------- -----  
## non man | 109 1734 0.644 0.489 0.850 0.002  
## manual | 108 1107   
## ------- + --------- ------------- ---------- ---------- --------- -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (manual: Manual vs non-manual sclass)  
## (heart\_attack: incident heart attack)

Similar interpretation as before – odds of heart attack are 0.063 among non-manual workers and 0.098 among manual workers, OR 1.55 for manual vs non-manual class: non-manual workersare 0.64 times as likely to have incident heart attack than manual workers. Confidence interval for OR 0.49-0.85, relatively far from 1.00 (but closer than in previous examples). The p-value = 0.02, strong evidence against null hypothesis, supporting evidence of association between occupational class and incident heart attacks.

**Q. Finally, let’s look at age10 variable.**

tabOdds(elsa, age\_group, by = heart\_attack, case\_value = 'Mentioned',  
 plot = FALSE, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Estimates of Risks of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## age\_group | Mentioned Not.mentioned Odds [95% Conf. Interval]  
## --------- + --------- ------------- ----- ---------- ---------  
## [45,55] | 40 1176 0.034 0.031 0.037  
## (55,65] | 81 905 0.090 0.081 0.099  
## (65,75] | 75 602 0.125 0.113 0.137  
## (75, 95] | 28 220 0.127 0.116 0.140  
## --------- + --------- ------------- ----- ---------- ---------  
## Total | 224 2903 0.077 0.070 0.085  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (heart\_attack: incident heart attack)

mhor(elsa, age\_group, by = heart\_attack, case\_value = 'Mentioned', na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## age\_group | Mentioned Not mentioned Odds Ratio [95% Conf. Interval] P>|z|  
## --------- + --------- ------------- ---------- ---------- --------- -----  
## [45,55] | 40 1176 0.380 0.258 0.560 0.000  
## (55,65] | 81 905   
## --------- + --------- ------------- ---------- ---------- --------- -----  
## [45,55] | 40 1176 0.273 0.184 0.406 0.000  
## (65,75] | 75 602   
## --------- + --------- ------------- ---------- ---------- --------- -----  
## [45,55] | 40 1176 0.267 0.161 0.442 0.000  
## (75, 95] | 28 220   
## --------- + --------- ------------- ---------- ---------- --------- -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (heart\_attack: incident heart attack)

By default, the 45-55 has been used as the reference category in the mhor() function. We can change this using the exp\_value argument. For instance:

mhor(elsa, age\_group, by = heart\_attack, exp\_value = '(55,65]',  
 case\_value = 'Mentioned', na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## age\_group | Mentioned Not mentioned Odds Ratio [95% Conf. Interval] P>|z|  
## --------- + --------- ------------- ---------- ---------- --------- -----  
## (55,65] | 81 905 2.631 1.784 3.881 0.000  
## [45,55] | 40 1176   
## --------- + --------- ------------- ---------- ---------- --------- -----  
## (55,65] | 81 905 0.718 0.516 1.000 0.049  
## (65,75] | 75 602   
## --------- + --------- ------------- ---------- ---------- --------- -----  
## (55,65] | 81 905 0.703 0.447 1.108 0.127  
## (75, 95] | 28 220   
## --------- + --------- ------------- ---------- ---------- --------- -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (heart\_attack: incident heart attack)

**Q. Please try to make conclusions about your today’s analysis. Which variables seem to be risk factors for incident heart attack –on what basis do you make these conclusions?**