IEHC0046 BASIC STATISTICS FOR MEDICAL SCIENCES

Confounding and Effect Modification: Practical

09 November, 2020

In this practical session we will learn how to use R to deal with confounding and effect modification in analyses. We will use the ELSA dataset that you have been using throughout the course so far. We will look at associations between two variables in this dataset and whether there is suggestion of confounding or effect modification by other variables. We’ll also be using the tidyverse, summarytools, and mStats packages. You may need to install the mStats package.

# install.packages("mStats") # Uncomment if need to install.  
library(tidyverse)  
library(summarytools)  
library(mStats)  
load("elsa.Rdata")

This practical is split into three sections. For each section there is a video that shows you how to use the Stata commands. You might choose to watch the video first and then work through the practical exercise by yourself. Or you might instead choose to work through the practical alongside watching the video.

1. Dealing with confounding in associations with a binary outcome
2. Dealing with effect modification in associations with a binary outcome
3. Dealing with multiple confounders or effect modifiers in associations with a binary outcome

This practical deviates slightly from the Stata practical. There we look at the Mantel-Haenszel test for trend. Here we use the simple Mantel-Haenszel test.

## 1. Dealing with confounding in associations with a binary outcome variable

In this first section we are going to look at the association between physical activity (the exposure) and heart attack (the outcome). Physical activity is represented by variable physact and incident heart attack by heart\_attack.

### Examining the variables

We should first examine these two key variables. Let’s use the freq() function from summarytools.

**Q. How is physical activity distributed in our sample?**

freq(elsa$physact)

## Frequencies   
## elsa$physact   
## Label: 3 categories of physical activity (1=low, 3=high)   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ---------------------- ------ --------- -------------- --------- --------------  
## Group 1 -low 1321 42.23 42.23 42.22 42.22  
## Group 2 - medium 1051 33.60 75.83 33.59 75.81  
## Group 3 - high 756 24.17 100.00 24.16 99.97  
## <NA> 1 0.03 100.00  
## Total 3129 100.00 100.00 100.00 100.00

There are 3128 people with information on physical activity in our sample, 1321 (42.2%) reported low physical activity, 1051 (33.6%) report medium activity, and 756 (24.2%) report high physical activity.

**Q. How many people in our sample had a heart attack?**

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

There are 3127 people with complete information on heart attack. 224 people in our sample had a heart attack (7.2%).

## Examining the association between physical activity and heart attack

We can examine the association between physical activity and heart attack using a chi-squared test as we have two categorical variables. Remember, to do this we cross-tabulate these two variables with ctable() (from summarytools), setting the argument chisq = TRUE to report the chi-squared statistics. We’ll also need to set the argument useNA = "no" to ignore missing values.

**Q. Is physical activity associated with heart attack?**

ctable(elsa$physact, elsa$heart\_attack, useNA = "no", chisq = TRUE)

## Cross-Tabulation, Row Proportions   
## physact \* heart\_attack   
## Data Frame: elsa   
##   
##   
## ------------------ -------------- --------------- ------------ ---------------  
## heart\_attack Not mentioned Mentioned Total  
## physact   
## Group 1 -low 1196 (90.5%) 125 (9.5%) 1321 (100.0%)  
## Group 2 - medium 990 (94.4%) 59 (5.6%) 1049 (100.0%)  
## Group 3 - high 716 (94.7%) 40 (5.3%) 756 (100.0%)  
## Total 2902 (92.8%) 224 (7.2%) 3126 (100.0%)  
## ------------------ -------------- --------------- ------------ ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 18.22 2 1e-04   
## ----------------------------

Yes, there appears to be an association between physical activity and having a heart attack. 9.5% of those reporting low physical activity had a heart attack compared to 5.6% of those reporting medium and 5.3% of those reporting high physical activity. The null hypothesis is that there is no association between physical activity and heart attack. The chi-squared statistic (fairly high) and association p value (very low, <0.001) suggest that there is little compatibility with being selected from a population where those with the three levels of physical activity had the same proportion of people with heart attacks. We can therefore conclude that there is an association between physical activity and heart attack in this sample.

**Q. Describe the association between physical activity and heart attack as an odds ratio and 95% confidence intervals.**

We can also express the association between physical activity and heart attack in terms of an odds ratio using the mhor() function from mStats. This function only reports the odds ratios, so let’s also use the tabOdds() function (from mStats) to report the odds of heart attach in this group.

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

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Estimates of Risks of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not.mentioned Odds [95% Conf. Interval]  
## ---------------- + --------- ------------- ----- ---------- ---------  
## Group 1 -low | 125 1196 0.105 0.095 0.115  
## Group 2 - medium | 59 990 0.060 0.054 0.066  
## Group 3 - high | 40 716 0.056 0.051 0.062  
## ---------------- + --------- ------------- ----- ---------- ---------  
## Total | 224 2902 0.077 0.070 0.085  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)

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

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 125 1196 1.754 1.272 2.417  
## Group 2 - medium | 59 990   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 125 1196 1.871 1.295 2.703  
## Group 3 - high | 40 716   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.001  
##   
## -----  
## 0.001  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)

As the frequency of physical activity increases the likelihood of having a heart attack decreases. People in the lowest category of physical activity have 75.4% higher odds (95% CI = 27.2%, 141.7%) of heart attack than individuals in the medium category and 87.1% higher odds (95% CI = 29.5%, 170.3%) than individuals in the highest category. The 95% CIs do not cross 1 and the p-values = 0.001, strong evidence against the null hypothesis.

## Assessing the role of self-rated health in the association between physical activity and heart attack

In this section we will assess whether self-rated health is a confounder of the association between physical activity and heart attack.

First self-rated health needs to be recoded into a new binary variable as we would like to have two categories comparing those with ‘very good’ and ‘good’ health, with people who have ‘fair’, ‘bad’ or ‘very bad’ health. We use mutate() from the tidyverse to do this.

**Q. Generate a new self-rated health variable with two such categories. Ensure that the categories are labelled.**

levels(elsa$srh)

## [1] "...very good," "good," "fair," "bad,"   
## [5] "or, very bad?"

elsa <- elsa %>%  
 mutate(srh\_2 = ifelse(srh %in% levels(srh)[1:2], # True if very good or good, false otherwise  
 "Very good or good", "Fair or bad"),  
 srh\_2 = ifelse(is.na(srh), NA, srh\_2), # Need to recode as missing is srh is missing  
 srh\_2 = factor(srh\_2)) # Convert to factor  
table(elsa$srh, elsa$srh\_2, useNA = "ifany")

##   
## Fair or bad Very good or good <NA>  
## ...very good, 0 722 0  
## good, 0 1279 0  
## fair, 862 0 0  
## bad, 169 0 0  
## or, very bad? 45 0 0  
## <NA> 0 0 52

Remember, to be a confounder the variable should be associated with both the exposure and outcome. Also it should not lie on the causal pathway. To meet the conditions of a confounder we will firstly assess whether the new binary self-rated health variable is associated with both physical activity and heart attack. Again, we can do this using a chi-squared test.

**Q. Run two Chi-squared tests to check whether self-rated health is associated with both the exposure (physical activity) and outcome (heart attack). What do you conclude?**

ctable(elsa$srh\_2, elsa$heart\_attack, useNA = "no", chisq = TRUE)

## Cross-Tabulation, Row Proportions   
## srh\_2 \* heart\_attack   
## Data Frame: elsa   
##   
##   
## ------------------- -------------- --------------- ------------- ---------------  
## heart\_attack Not mentioned Mentioned Total  
## srh\_2   
## Fair or bad 927 (86.3%) 147 (13.7%) 1074 (100.0%)  
## Very good or good 1928 (96.4%) 73 ( 3.6%) 2001 (100.0%)  
## Total 2855 (92.8%) 220 ( 7.2%) 3075 (100.0%)  
## ------------------- -------------- --------------- ------------- ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 104.5 1 0   
## ----------------------------

Here we can see that there is an association between self-rated health and heart attack. 3.6% of those with very good or good health had a heart attack, however 13.7% of those with fair or bad health had a heart attack. The Chi-squared statistic is high (104.5) with a low p value (<0.001). Therefore we can say that self-rated health and heart attack are associated. It appears that those with poorer health are more likely to have had a heart attack.

ctable(elsa$physact, elsa$srh\_2, useNA = "no", chisq = TRUE, prop = "c")

## Cross-Tabulation, Column Proportions   
## physact \* srh\_2   
## Data Frame: elsa   
##   
##   
## ------------------ ------- --------------- ------------------- ---------------  
## srh\_2 Fair or bad Very good or good Total  
## physact   
## Group 1 -low 595 ( 55.3%) 693 ( 34.6%) 1288 ( 41.9%)  
## Group 2 - medium 287 ( 26.7%) 751 ( 37.5%) 1038 ( 33.7%)  
## Group 3 - high 193 ( 18.0%) 557 ( 27.8%) 750 ( 24.4%)  
## Total 1075 (100.0%) 2001 (100.0%) 3076 (100.0%)  
## ------------------ ------- --------------- ------------------- ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 124 2 0   
## ----------------------------

We added the argument prop = "c" to get the column proportions rather than the row proportions. Again there appears to be an association between self-rated health and physical activity. 27.8% of those with very good or good health reported high physical activity. 18% of those with fair or bad health reported high physical activity. Similarly, a greater percentage of people with fair or bad health reported low physical activity. The Chi-squared statistic is again high (124) and the p value is low (p<0.001) suggesting low compatibility with people with good or bad health having reporting the same levels of physical activity.

Therefore we can conclude that self-rated health is associated with both our exposure (physical activity) and outcome (heart attack). It is also unlikely to be on the causal pathway between these two variables and we therefore consider it to be a confounder of this association.

The next step is to use the mantelhaen.test() function to see what effect the inclusion of self-rated health has in our analysis of the association between physical activity and heart attack.

**Q. Assess the role of self-rated health in the association between physical activity and heart attack using the mantelhaen.test() function. Interpret your findings.**

We showed earlier that physical activity was associated with heart attack. Here we will assess the role of self-rated health in this association using the mantelhaen.test() function. This function only reports the results of the test, so let’s also use the mhor() function to see how the odds ratios differ by self-rated health.

mhor(elsa[elsa$srh\_2 == "Very good or good", ],  
 physact, by = heart\_attack,  
 case\_value = "Mentioned", na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 33 660 1.583 0.920 2.723  
## Group 2 - medium | 23 728   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 33 660 1.588 0.875 2.883  
## Group 3 - high | 17 540   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.095  
##   
## -----  
## 0.125  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)

mhor(elsa[elsa$srh\_2 == "Fair or bad", ],  
 physact, by = heart\_attack,  
 case\_value = "Mentioned", na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 90 505 1.273 0.837 1.935  
## Group 2 - medium | 35 250   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 90 505 1.385 0.842 2.278  
## Group 3 - high | 22 171   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.258  
##   
## -----  
## 0.198  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)

mantelhaen.test(elsa$physact, elsa$heart\_attack, elsa$srh\_2)

##   
## Cochran-Mantel-Haenszel test  
##   
## data: elsa$physact and elsa$heart\_attack and elsa$srh\_2  
## Cochran-Mantel-Haenszel M^2 = 5.5996, df = 2, p-value = 0.06082

We can see that the strength of the association between physical activity and heart attack is weaker among those with fair or bad self-rated health than among those with good self-rated health. For instance, low physical activity is related to 27.3% (95% CI = -16.3%, 93.5%) higher odds of heart attack compared with medium physical activity among the fair/bad self-rated health group, but 58.3% (95% CI = -8%, 172.3%) higher odds of heart attack among the good/very good self-rated health group. The p-value of the Mantel-Haenszel test is 0.06082 so the test is not significant at the 95% level. Notice also that the size of the odds-ratios is smaller than when we did not stratify by self-rated health. This suggests that part of the association between physical activity and heart attack is confounded by self-rated health.

## 2. Dealing with effect modification in associations with a binary outcome

### Assessing the role of sex in the association between physical activity and heart attack

In this second practical we are going to apply the methods for testing for effect modification in associations involving a binary outcome as shown in part B of the lecture. Please watch part B of the lecture before attempting this practical exercise.

We will return to the association between physical activity and heart attack that we tested in part 1 of this practical. Here we will assess whether sex plays a role in this association as an effect modifier.

**Q. Using the mhor() and mantelhaen.test() functions, assess whether sex plays a role in the association between physical activity and heart attack. What do you conclude?**

mhor(elsa[elsa$sex == "male", ],  
 physact, by = heart\_attack,  
 case\_value = "Mentioned", na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 76 472 1.504 1.005 2.251  
## Group 2 - medium | 41 383   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 76 472 1.656 1.085 2.529  
## Group 3 - high | 35 360   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.046  
##   
## -----  
## 0.019  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)

mhor(elsa[elsa$sex == "female", ],  
 physact, by = heart\_attack,  
 case\_value = "Mentioned", na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 49 724 2.282 1.316 3.959  
## Group 2 - medium | 18 607   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 49 724 4.819 1.903 12.200  
## Group 3 - high | 5 356   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.003  
##   
## -----  
## 0.000  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)

mantelhaen.test(elsa$physact, elsa$heart\_attack, elsa$sex)

##   
## Cochran-Mantel-Haenszel test  
##   
## data: elsa$physact and elsa$heart\_attack and elsa$sex  
## Cochran-Mantel-Haenszel M^2 = 21.891, df = 2, p-value = 1.764e-05

Focussing firstly on the output of the mantelhaen.test() function, we can see that there is evidence that sex is an effect modifier (p < 0.001). Looking at the ouput of the mhor() functions we see that for both men and women, there is an association between physical activity and heart attack, but the association is larger among women. For instance, comparing low and medium physical activity groups among women, the odds of heart attack is 128.2% higher in the former group (95% CI = 31.6%, 295.9%). Among men, the corresponding figure is 50.4% (95% CI = 0.05%, 125.1%). We can therefore conclude that sex is an effect modifier of the relationship between physical activity and heart attack, and that the association is stronger for women.

### Assessing the role of wealth in the association between physical activity and heart attack

**Q. Assess whether wealth plays a role in the association between physical activity and heart attack following a similar approach to the previous section. Use variable wealth5 in your analysis.**

Rather than write the mhor function five times, we will use the map() function from purrr which is part of the tidyverse.

str(elsa$wealth5)

## num [1:3129] 2 4 1 2 3 2 1 5 2 4 ...  
## - attr(\*, "label")= chr "quintiles of household wealth"

map(1:5,  
 ~ mhor(elsa[elsa$wealth5 == .x, ],  
 physact, by = heart\_attack,  
 case\_value = "Mentioned", na.rm = TRUE))

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 34 244 1.033 0.555 1.921  
## Group 2 - medium | 17 126   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 34 244 1.368 0.668 2.801  
## Group 3 - high | 11 108   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.919  
##   
## -----  
## 0.390  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 27 241 1.180 0.609 2.288  
## Group 2 - medium | 15 158   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 27 241 3.959 1.175 13.335  
## Group 3 - high | 3 106   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.624  
##   
## -----  
## 0.017  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 22 210 1.928 0.890 4.177  
## Group 2 - medium | 10 184   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 22 210 1.846 0.800 4.264  
## Group 3 - high | 8 141   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.091  
##   
## -----  
## 0.146  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 23 240 3.674 1.469 9.186  
## Group 2 - medium | 6 230   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 23 240 1.713 0.746 3.931  
## Group 3 - high | 8 143   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.003  
##   
## -----  
## 0.200  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'heart\_attack'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 19 244 1.961 0.915 4.202  
## Group 2 - medium | 11 277   
## ---------------- + --------- ------------- ---------- ---------- ---------  
## Group 1 -low | 19 244 1.722 0.762 3.889  
## Group 3 - high | 9 199   
## -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## -----  
## 0.079  
##   
## -----  
## 0.187  
##   
## -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (physact: 3 categories of physical activity (1=low, 3=high))  
## (heart\_attack: incident heart attack)

## [[1]]  
## [[1]][[1]]  
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## 1 ---------------- + --------- ------------- ---------- ---------- ---------  
## 2 Group 1 -low | 34 244 1.033 0.555 1.921  
## 3 Group 2 - medium | 17 126   
## 4 ---------------- + --------- ------------- ---------- ---------- ---------  
## 5 Group 1 -low | 34 244 1.368 0.668 2.801  
## 6 Group 3 - high | 11 108   
## 7 -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## 1 -----  
## 2 0.919  
## 3   
## 4 -----  
## 5 0.390  
## 6   
## 7 -----  
##   
##   
## [[2]]  
## [[2]][[1]]  
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## 1 ---------------- + --------- ------------- ---------- ---------- ---------  
## 2 Group 1 -low | 27 241 1.180 0.609 2.288  
## 3 Group 2 - medium | 15 158   
## 4 ---------------- + --------- ------------- ---------- ---------- ---------  
## 5 Group 1 -low | 27 241 3.959 1.175 13.335  
## 6 Group 3 - high | 3 106   
## 7 -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## 1 -----  
## 2 0.624  
## 3   
## 4 -----  
## 5 0.017  
## 6   
## 7 -----  
##   
##   
## [[3]]  
## [[3]][[1]]  
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## 1 ---------------- + --------- ------------- ---------- ---------- ---------  
## 2 Group 1 -low | 22 210 1.928 0.890 4.177  
## 3 Group 2 - medium | 10 184   
## 4 ---------------- + --------- ------------- ---------- ---------- ---------  
## 5 Group 1 -low | 22 210 1.846 0.800 4.264  
## 6 Group 3 - high | 8 141   
## 7 -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## 1 -----  
## 2 0.091  
## 3   
## 4 -----  
## 5 0.146  
## 6   
## 7 -----  
##   
##   
## [[4]]  
## [[4]][[1]]  
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## 1 ---------------- + --------- ------------- ---------- ---------- ---------  
## 2 Group 1 -low | 23 240 3.674 1.469 9.186  
## 3 Group 2 - medium | 6 230   
## 4 ---------------- + --------- ------------- ---------- ---------- ---------  
## 5 Group 1 -low | 23 240 1.713 0.746 3.931  
## 6 Group 3 - high | 8 143   
## 7 -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## 1 -----  
## 2 0.003  
## 3   
## 4 -----  
## 5 0.200  
## 6   
## 7 -----  
##   
##   
## [[5]]  
## [[5]][[1]]  
## physact | Mentioned Not mentioned Odds Ratio [95% Conf. Interval]  
## 1 ---------------- + --------- ------------- ---------- ---------- ---------  
## 2 Group 1 -low | 19 244 1.961 0.915 4.202  
## 3 Group 2 - medium | 11 277   
## 4 ---------------- + --------- ------------- ---------- ---------- ---------  
## 5 Group 1 -low | 19 244 1.722 0.762 3.889  
## 6 Group 3 - high | 9 199   
## 7 -------------- + --------- ------------- ---------- ---------- ---------  
## P>|z|  
## 1 -----  
## 2 0.079  
## 3   
## 4 -----  
## 5 0.187  
## 6   
## 7 -----

mantelhaen.test(elsa$physact, elsa$heart\_attack, elsa$wealth5)

##   
## Cochran-Mantel-Haenszel test  
##   
## data: elsa$physact and elsa$heart\_attack and elsa$wealth5  
## Cochran-Mantel-Haenszel M^2 = 14.274, df = 2, p-value = 0.0007952

# # 4. Confounding and effect modification ----  
# rm(list = ls())  
# load("elsa.Rdata")  
#   
# freq(elsa$physact)  
# freq(elsa$heart\_attack)  
# ctable(elsa$physact, elsa$heart\_attack, useNA = "no", chisq = TRUE)  
#   
# library(mStats)  
# tabOdds(elsa, physact, by = heart\_attack, plot = FALSE, na.rm = TRUE)  
# mhor(elsa, physact, by = heart\_attack, na.rm = TRUE)  
#   
# events <-table(elsa$physact[elsa$heart\_attack == "Mentioned"])  
# trials <-table(elsa$physact)  
# prop.trend.test(x = events, n = trials)  
#   
# str(elsa$srh)  
# freq(elsa$srh)  
# elsa <- elsa %>%  
# mutate(srh\_2 = ifelse(srh %in% levels(srh)[1:2],  
# "Very good or good", "Fair or bad"),  
# srh\_2 = ifelse(is.na(srh), NA, srh\_2),  
# srh\_2 = factor(srh\_2))  
# table(elsa$srh, elsa$srh\_2, useNA = "ifany")  
#   
# ctable(elsa$srh\_2, elsa$heart\_attack, useNA = "no", chisq = TRUE)  
# ctable(elsa$physact, elsa$srh\_2, useNA = "no", chisq = TRUE)  
#   
# tab\_3 <- table(elsa$heart\_attack, elsa$physact, elsa$srh\_2)  
# mantelhaen.test(tab\_3)  
# mantelhaen.test(elsa$physact, elsa$heart\_attack, elsa$srh\_2)  
# library(magrittr)  
# elsa %>%  
# dplyr::select(physact, heart\_attack, srh\_2) %>%  
# drop\_na() %$%  
# epiDisplay::mhor(physact, heart\_attack, srh\_2, graph = F)