IEHC0046 BASIC STATISTICS FOR MEDICAL SCIENCES

Analysis of Categorical Data III: Practical

08 November, 2020

In this practical we will again use R to calculate chi-squared test, chi-squared test for trend, odds ratios and risk 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, summarytools and mStats package. Load the packages - and install, if necessary.

library(tidyverse)  
library(summarytools)  
library(mStats)

Today, we will be interested in obesity. We have two relevant variables in our data: bmi and bmi4. Let’s use one of these to and create new binary variable obesity for whether BMI is over 30kg/m2. This will be our dependent variable (outcome) of interest.

elsa <- elsa %>%  
 mutate(obese = ifelse(bmi > 30, "Obese", "Not Obese"),  
 obese = factor(obese, levels = c("Obese", "Not Obese")))

**Q. How many cases of obesity do we have**

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

freq(elsa$obese)

## Frequencies   
## elsa$obese   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## --------------- ------ --------- -------------- --------- --------------  
## Obese 717 24.47 24.47 22.91 22.91  
## Not Obese 2213 75.53 100.00 70.73 93.64  
## <NA> 199 6.36 100.00  
## Total 3129 100.00 100.00 100.00 100.00

**Q. Whatthe distribution of wealth5 (quintiles of household wealth, 1=low, 5=high)**

str(elsa$wealth5)

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

freq(elsa$wealth5)

## Frequencies   
## elsa$wealth5   
## Label: quintiles of household wealth   
## Type: Numeric   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 1 541 17.58 17.58 17.29 17.29  
## 2 551 17.91 35.49 17.61 34.90  
## 3 576 18.72 54.21 18.41 53.31  
## 4 650 21.12 75.33 20.77 74.08  
## 5 759 24.67 100.00 24.26 98.34  
## <NA> 52 1.66 100.00  
## Total 3129 100.00 100.00 100.00 100.00

**Q. Look at the association between obesity and wealth. Calculate chi square test and draw your conclusions about the association.**

We can use the ctable() function from the last session using the argument chisq = TRUE. We’ll also use the argument useNA = "no" to remove missing values.

ctable(elsa$wealth5, elsa$obese, chisq = TRUE, useNA = "no")

## Cross-Tabulation, Row Proportions   
## wealth5 \* obese   
## Data Frame: elsa   
##   
##   
## --------- ------- ------------- -------------- ---------------  
## obese Obese Not Obese Total  
## wealth5   
## 1 154 (30.9%) 344 (69.1%) 498 (100.0%)  
## 2 156 (30.4%) 357 (69.6%) 513 (100.0%)  
## 3 136 (25.1%) 406 (74.9%) 542 (100.0%)  
## 4 126 (20.7%) 482 (79.3%) 608 (100.0%)  
## 5 134 (18.6%) 585 (81.4%) 719 (100.0%)  
## Total 706 (24.5%) 2174 (75.5%) 2880 (100.0%)  
## --------- ------- ------------- -------------- ---------------  
##   
## ----------------------------  
## Chi.squared df p.value   
## ------------- ---- ---------  
## 38.93 4 0   
## ----------------------------

Ordinary chi square test suggests the associationbetween wealth and obesity – chi-square statistic is large and p<0.001. Strong evidence against null hypothesis of no association.

**Q. Wealth classified in 5 quintiles is an example of ordinal variable. We can look at chi square test for trend.**

We can use the function prop.trend.test() from Base R. The function takes two inputs, x, the number of “events” in each groups (in this case people with obesity) , and n the number of “trials” in each group (in this case the number of people).

events <- table(elsa$wealth5[elsa$obese == "Obese"])  
trials <- table(elsa$wealth5)  
prop.trend.test(x = events, n = trials)

##   
## Chi-squared Test for Trend in Proportions  
##   
## data: events out of trials ,  
## using scores: 1 2 3 4 5  
## X-squared = 33.314, df = 1, p-value = 7.841e-09

The chi-square test for trend suggest strong dose response relationship (null hypothesis of no association can be rejected with p< 0.001). The results from the previous question suggest there is successively lower proportions of obese people in higher wealth groups.

**Q. Calculate odds ratios to assess the magnitude of the wealth effect on obesity.**

We’ll use the tabOdds() and mhor() functions from the previous session.

tabOdds(elsa, wealth5, by = obese, plot = FALSE, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Estimates of Risks of 'obese'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## wealth5 | Obese Not.Obese Odds [95% Conf. Interval]  
## ------- + ----- --------- ----- ---------- ---------  
## 1 | 154 344 0.448 0.422 0.475  
## 2 | 156 357 0.437 0.412 0.464  
## 3 | 136 406 0.335 0.315 0.356  
## 4 | 126 482 0.261 0.246 0.278  
## 5 | 134 585 0.229 0.216 0.243  
## ------- + ----- --------- ----- ---------- ---------  
## Total | 706 2174 0.325 0.306 0.345  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (wealth5: quintiles of household wealth)

mhor(elsa, wealth5, by = obese, na.rm = TRUE)

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## Odds Ratio Estimates of 'obese'   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
##   
## wealth5 | Obese Not Obese Odds Ratio [95% Conf. Interval] P>|z|  
## ------- + ----- --------- ---------- ---------- --------- -----  
## 1 | 154 344 1.024 0.784 1.339 0.859  
## 2 | 156 357   
## ------- + ----- --------- ---------- ---------- --------- -----  
## 1 | 154 344 1.336 1.018 1.754 0.036  
## 3 | 136 406   
## ------- + ----- --------- ---------- ---------- --------- -----  
## 1 | 154 344 1.713 1.303 2.250 0.000  
## 4 | 126 482   
## ------- + ----- --------- ---------- ---------- --------- -----  
## 1 | 154 344 1.954 1.496 2.553 0.000  
## 5 | 134 585   
## ------- + ----- --------- ---------- ---------- --------- -----  
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
## (Labels)  
## (wealth5: quintiles of household wealth)

With increasing categories of wealth, odds ratio gets smaller and smaller – individuals in higher quintiles of wealth are less likely to bein obesity category. Those in wealth category 1 are 1.95-times more likely to have BMI > 30kg/m2 than those in category 5. 95% CIs typically do not cross 1, suggesting that there is higher levels of obesity in lower wealth groups.

**Q. Finally, let’s look at risk ratios. We will use the binary variable sex.**

We can use the ctable() function again but setting argument RR = TRUE.

ctable(elsa$sex, elsa$obese, OR = TRUE, RR = TRUE, useNA = "no")

## Cross-Tabulation, Row Proportions   
## sex \* obese   
## Data Frame: elsa   
##   
##   
## -------- ------- ------------- -------------- ---------------  
## obese Obese Not Obese Total  
## sex   
## male 271 (21.0%) 1021 (79.0%) 1292 (100.0%)  
## female 446 (27.2%) 1192 (72.8%) 1638 (100.0%)  
## Total 717 (24.5%) 2213 (75.5%) 2930 (100.0%)  
## -------- ------- ------------- -------------- ---------------  
##   
## ----------------------------------  
## Odds Ratio Lo - 95% Hi - 95%   
## ------------ ---------- ----------  
## 0.71 0.60 0.84   
## ----------------------------------  
##   
## ----------------------------------  
## Risk Ratio Lo - 95% Hi - 95%   
## ------------ ---------- ----------  
## 0.77 0.67 0.88   
## ----------------------------------

Odds ratio more extreme than relative risk; both suggesting effect in the same direction (men less likely to have BMI>30 than women). Both 95% confidence intervals haveh lower and upper limit below 1 - strong evidence of the association between sex and obesity.

**Q. Use binary smoking variable (smok\_bin) and calculate risk and odds ratios. Interpret and give conclusions.**

ctable(elsa$smok\_bin, elsa$obese, OR = TRUE, RR = TRUE, useNA = "no")

## Cross-Tabulation, Row Proportions   
## smok\_bin \* obese   
## Data Frame: elsa   
##   
##   
## ---------------- ------- ------------- -------------- ---------------  
## obese Obese Not Obese Total  
## smok\_bin   
## never/ex occ 312 (24.4%) 969 (75.6%) 1281 (100.0%)  
## ex/current reg 405 (24.6%) 1244 (75.4%) 1649 (100.0%)  
## Total 717 (24.5%) 2213 (75.5%) 2930 (100.0%)  
## ---------------- ------- ------------- -------------- ---------------  
##   
## ----------------------------------  
## Odds Ratio Lo - 95% Hi - 95%   
## ------------ ---------- ----------  
## 0.99 0.83 1.17   
## ----------------------------------  
##   
## ----------------------------------  
## Risk Ratio Lo - 95% Hi - 95%   
## ------------ ---------- ----------  
## 0.99 0.87 1.13   
## ----------------------------------

No evidence for association between smoking and obesity. RR 1.01, OR 1.01 – no difference in risk of obesity or odds of obesity between smokers and non-smokers. 95% confidence interval almost symmetrical around 1.00.