Меры Ассоциации. Домашняя работа

Victoria Zaitceva

2024-10-13

# Задание 1.

Врачи решили исследовать, как индекс массы тела (ИМТ) ассоциирован с риском развития диабета 2-го типа. Файл diabetes.csv содержит данные о случайной выборке из 200 жителей населённого пункта N. Для каждого респондента известен ИМТ (высокий или нормальный) и статус по диабету (наличие/отсутствие диабета 2-го типа). Определите, как высокий ИМТ ассоциирован c развитием диабета 2-го типа, укажите относительный риск (relative risk) и абсолютную разницу в рисках (risk difference). Как вы проинтерпретируете полученные результаты?

## Решение

Мера ассоциации для двух категориальных (дихотомических) переменных - это относительный риск (relative risk). Он определяется как отношение вероятности заболеть в группе с высоким ИМТ к вероятности заболеть в группе с нормальным ИМТ.

Построим таблицу сопряженности для переменных ИМТ и диабета:

diabetes <- read.csv("diabetes.csv")

table\_diabetes <- table(diabetes$ИМТ, diabetes$Диабет) %>%  
 addmargins() %>%  
 as.data.frame.matrix() %>%  
 rownames\_to\_column(var = "ИМТ") %>%  
 as\_tibble()   
  
  
colnames(table\_diabetes) <- c("ИМТ", "Диабет", "Нет диабета", "Всего")  
  
# для красоты  
table\_diabetes\_flex <- table\_diabetes %>%  
 flextable()  
table\_diabetes\_flex

| ИМТ | Диабет | Нет диабета | Всего |
| --- | --- | --- | --- |
| Высокий | 67 | 46 | 113 |
| Нормальный | 23 | 64 | 87 |
| Sum | 90 | 110 | 200 |

# relative risk  
Risk\_diabetes <- table\_diabetes[1, 2] / table\_diabetes[1, 4]  
Risk\_no\_diabetes <- table\_diabetes[2, 2] / table\_diabetes[2, 4]  
  
RR <- round(as.numeric(Risk\_diabetes / Risk\_no\_diabetes),1)  
print(paste("Risk Ratio = ", RR))

## [1] "Risk Ratio = 2.2"

RD <- round(as.numeric(Risk\_diabetes - Risk\_no\_diabetes),1)  
print(paste("Risk Difference = ", RD))

## [1] "Risk Difference = 0.3"

Относительный риск (relative risk) равен 2.2, абсолютная разница в рисках (risk difference) равна 0.3. Это означает, что у людей с высоким ИМТ риск заболеть диабетом 2-го типа в 2.2 раза выше, чем у людей с нормальным ИМТ. Абсолютная разница в рисках равна 0.3, что означает, что у людей с высоким ИМТ риск заболеть диабетом 2-го типа на 30% выше, чем у людей с нормальным ИМТ.

# Задание 2.

В городе N зафиксирована вспышка пневмонии. Пострадало 250 человек, проживающих в разных домах. Все они на протяжении последних двух недель посещали различные места: торговые центры, рестораны и общественные мероприятия. Для контроля взяли 750 человек, которые не заболели пневмонией. Был проведен опрос о том, какие места каждый человек посещал (pneumonia.csv). Используя подходящую меру ассоциации, определите, какое место посещения с наибольшей вероятностью связано с возникновением пневмонии.

## Решение

Данное исследование типа случай-контроль. Участники были отобраны по наличию или отсутствию пневмонии, и для каждого участника были ретроспективно собраны данные о том, какие места он посещал. Так как мы начинаем наше исследование с исхода, наша выборка оприори нерепрезентативна, и мы не можем на основании этого анализа узнать распространенность заболевания (как можно было бы, если бы мы отобрали всех здоровых изначально и смотрели, сколько заболеет в течение некоторого промежутка времени).

Лучшей мерой ассоциации для исследования типа случай-контроль является отношение шансов (odds ratio). Оно определяется как отношение шансов заболеть в группе пациентов (с пневмонией) к шансам заболеть в контрольной группе (здоровым).

В дополнение к основно мере ассоциации, мы можем использовать критерий хи-квадрат для проверки независимости двух категориальных переменных (группа (случай/контроль) и факт посещения места (да/нет)).

pneumonia <- read.csv("pneumonia.csv")  
colnames(pneumonia) <- c("ID", "Группа", "Торговый\_центр", "Ресторан", "Общественные\_мероприятия") # чтобы пробелы не смущали

places <- c("Торговый\_центр", "Ресторан", "Общественные\_мероприятия")  
chi\_squared\_OR\_results <- list()  
  
for (place in places) {  
 # таблица сопряженности для каждого места  
 contingency\_table <- table(pneumonia[[place]], pneumonia$Группа)  
 contingency\_table <- contingency\_table[, c(2, 1)] # меняем местами столбцы  
   
 # chi-squared test  
 chi\_test <- chisq.test(contingency\_table)  
   
 # odds ratio  
 odds\_ratio <- (contingency\_table[1, 1]/contingency\_table[1, 2]) / (contingency\_table[2, 1]/contingency\_table[2, 2])  
   
 # Store results  
 chi\_squared\_OR\_results[[place]] <- list(  
 "Contingency Table" = contingency\_table,  
 "Chi-Squared Statistic" = chi\_test$statistic,  
 "p-value" = chi\_test$p.value,  
 "OR" = odds\_ratio  
 )  
}  
  
  
for (place in places) {  
 cat("\nАнализ для:", place, "\n")  
 print(chi\_squared\_OR\_results[[place]]$`Contingency Table`)  
 cat("Chi-Squared Statistic:", round(chi\_squared\_OR\_results[[place]]$`Chi-Squared Statistic`,2), "\n")  
 cat("p-value:", round(chi\_squared\_OR\_results[[place]]$`p-value`,2), "\n")  
 cat("Odds Ratio:", round(chi\_squared\_OR\_results[[place]]$`OR`,2), "\n")  
}

##   
## Анализ для: Торговый\_центр   
##   
## Пневмония Контроль  
## Да 142 344  
## Нет 108 406  
## Chi-Squared Statistic: 8.54   
## p-value: 0   
## Odds Ratio: 1.55   
##   
## Анализ для: Ресторан   
##   
## Пневмония Контроль  
## Да 125 356  
## Нет 125 394  
## Chi-Squared Statistic: 0.39   
## p-value: 0.53   
## Odds Ratio: 1.11   
##   
## Анализ для: Общественные\_мероприятия   
##   
## Пневмония Контроль  
## Да 126 381  
## Нет 124 369  
## Chi-Squared Statistic: 0   
## p-value: 0.97   
## Odds Ratio: 0.98

По результатам анализа, посещение торгового центра наиболее ассоциировано с возникновением пневмонии: посещение ТЦ повышает шансы заболеть пневмонией на 55% (OR = 1.55). Критерий хи-квадрат подтверждает, что между посещением ТЦ и заболеванием пневмонией есть статистически значимая ассоциации (хи-квадрат 8.54, p-value < 0.05).

Посещение общественных мероприятий и ресторанов не ассоциировано с возникновением пневмонии, что подтверждается OR близким к 1 и статистически незначимым хи-квадратом.