Меры Ассоциации. Домашняя работа

Victoria Zaitceva

2024-10-06

# 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$Диабет, diabetes$ИМТ) %>%  
 addmargins() %>%  
 as.data.frame.matrix() %>%  
 rownames\_to\_column(var = "Диабет") %>%  
 as\_tibble()   
# add prefix to column names  
colnames(table\_diabetes) <- c("Диабет", "Высокий ИМТ", "Нормальный ИМТ", "Всего")  
  
table\_diabetes\_flex <- table\_diabetes %>%  
 flextable()  
table\_diabetes\_flex

| Диабет | Высокий ИМТ | Нормальный ИМТ | Всего |
| --- | --- | --- | --- |
| Есть | 67 | 23 | 90 |
| Нет | 46 | 64 | 110 |
| Sum | 113 | 87 | 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 = 1.8"

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 1.8, абсолютная разница в рисках (risk difference) равна 0.3. Это означает, что у людей с высоким ИМТ риск заболеть диабетом 2-го типа в 1.8 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

Здесь мы хотим сравнить две независимые группы (пациенты с пневмонией и контрольная группа) по тому, какие места они посещали. Для этого мы можем использовать критерий хи-квадрат для проверки независимости двух категориальных переменных.

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

colnames(pneumonia) <- c("ID", "Группа", "Торговый\_центр", "Ресторан", "Общественные\_мероприятия")  
   
  
   
pneumonia\_table <- pneumonia %>%  
 group\_by(Группа) %>%  
 summarise(  
 ТЦ\_Да = sum(Торговый\_центр == "Да"),  
 ТЦ\_Нет = sum(Торговый\_центр == "Нет"),  
 Ресторан\_Да = sum(Ресторан == "Да"),  
 Ресторан\_Нет = sum(Ресторан == "Нет"),  
 Мероприятия\_Да = sum(Общественные\_мероприятия == "Да"),  
 Мероприятия\_Нет = sum(Общественные\_мероприятия == "Нет")  
 ) %>%  
 flextable()  
  
  
pneumonia\_table

| Группа | ТЦ\_Да | ТЦ\_Нет | Ресторан\_Да | Ресторан\_Нет | Мероприятия\_Да | Мероприятия\_Нет |
| --- | --- | --- | --- | --- | --- | --- |
| Контроль | 344 | 406 | 356 | 394 | 381 | 369 |
| Пневмония | 142 | 108 | 125 | 125 | 126 | 124 |

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

##   
## Анализ для: Торговый\_центр   
##   
## Контроль Пневмония  
## Да 344 142  
## Нет 406 108  
## Chi-Squared Statistic: 8.540029   
## p-value: 0.003474202   
##   
## Анализ для: Ресторан   
##   
## Контроль Пневмония  
## Да 356 125  
## Нет 394 125  
## Chi-Squared Statistic: 0.3858906   
## p-value: 0.5344674   
##   
## Анализ для: Общественные\_мероприятия   
##   
## Контроль Пневмония  
## Да 381 126  
## Нет 369 124  
## Chi-Squared Statistic: 0.001333595   
## p-value: 0.970869

По результатам анализа, место посещения, которое наиболее ассоциировано с возникновением пневмонии, - это торговый центр, так как он имеет p-value < 0.05 и наибольший хи-квадрат (8.54).