Домашнее задание по мерам заболеваемости/случаемости

BioStat 2024

Ustin Zolotikov

2024-09-22

Оглавление

# Датасет

Описывает аренду автомобилей водителями и попадание ДТП в течение года.

* id - идентификатор,
* experience - стаж вождения (0 - нет стажа, 1 - есть стаж),
* accident - ДТП (0 - возврат из аренды целого автомобиля, 1 - возврат из аренды автомобиля после ДТП),
* start - день начала аренды,
* stop - день прекращения аренды.

summary(data %>% select(!id))

## experience start stop accident   
## Min. :0.00 Min. : 1.00 Min. : 12.0 Min. :0.00   
## 1st Qu.:0.00 1st Qu.: 68.75 1st Qu.:146.2 1st Qu.:0.00   
## Median :0.00 Median :158.00 Median :250.5 Median :0.00   
## Mean :0.32 Mean :166.74 Mean :239.0 Mean :0.22   
## 3rd Qu.:1.00 3rd Qu.:258.75 3rd Qu.:356.5 3rd Qu.:0.00   
## Max. :1.00 Max. :360.00 Max. :365.0 Max. :1.00

head(data)

## # A tibble: 6 × 5  
## id experience start stop accident  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 351 365 0  
## 2 2 1 128 149 0  
## 3 3 1 40 41 0  
## 4 4 0 79 147 0  
## 5 5 0 53 103 0  
## 6 6 0 61 93 1

# Находим распространенность “experience”

prevalance <- sum(data$experience) / length(data$experience)

## [1] "Prevalance = 0.32"

Распространенность показывает, что 32% водителей, участвовавших в исследовании, имели стаж.

# Находим риск “accident” в общей группе, в группе “experience” и в группе без “experience”

exp <- data %>% filter(experience == 1)  
noexp <- data %>% filter(experience == 0)  
incidence\_all <- sum(data$accident) / length(data$accident)  
incidence\_exp <- sum(exp$accident) / length(exp$accident)  
incidence\_noexp <- sum(noexp$accident) / length(noexp$accident)

## [1] "Incidence (1 year) = 0.22"

## [1] "Incidence in experienced group (1 year) = 0.09375"

## [1] "Incidence in unexperienced group (1 year) = 0.279411764705882"

Из годового наблюдения можно вывести риск водителей попасть в ДТП = 22%, при этом риск опытных водителей = 9.375%, а неопытных = 2.794%. В целом, ожидаемо, что риск попасть в аварию выше у водителей без опыта.

Риски рассчитывались отношением количества новых случаев дтп (в группе) к количеству новых случаев взятия в аренду автомобиля (в группе) в течение периода наблюдения (1 год).

# Находим плотность событий (incidence rate) в общей группе, в группе “experience” и в группе без “experience”

patient\_time\_all <- sum(data$stop - data$start)  
patient\_time\_exp <- sum(exp$stop - exp$start)  
patient\_time\_noexp <- sum(noexp$stop - noexp$start)  
  
incidence\_rate\_all <- sum(data$accident) / patient\_time\_all  
incidence\_rate\_exp <- sum(exp$accident) / patient\_time\_exp  
incidence\_rate\_noexp <- sum(noexp$accident) / patient\_time\_noexp

## [1] "Incidence rate (1 day) = 0.00304624757684852"

## [1] "Incidence rate in experienced group (1 day) = 0.00317796610169492"

## [1] "Incidence rate in unexperienced group (1 day) = 0.00302644154189232"

Плотность событий за день показывает, что в день в среднем происходит 0.3% ДТП среди всех водителей, 0.04% среди опытных водителей и 0.26%(выше в 6.5 раза!) среди неопытных водителей. Эти данные подкрепляют вывод о том, что опыт вождения сильно влияет на вероятность ДТП.