Home\_work\_lecture\_1

Zarva\_I

2024-09-22

#Чтение файла  
df <- read.csv("data/raw/carrental\_id.csv", header = TRUE)  
head(df, 5)

## X id experience start stop accident id\_rent  
## 1 1 1 0 351 365 0 a3f7d4a39b65b170005aafbbeed05106  
## 2 2 2 1 128 149 0 5391009fc72643b4f98a7307baf4a4c9  
## 3 3 3 1 40 41 0 7b00a8676bfbbf3ca1278d66289505e8  
## 4 4 4 0 79 147 0 80027a04a698b863ba848ed65ae365bf  
## 5 5 5 0 53 103 0 fb5375101937f15f22b2c5a5a4bb77f7  
## id\_driver  
## 1 1a8258e7d2adc48ac67a89a1854278a2  
## 2 c8f7037c66966f7979df22c34fc853f0  
## 3 97eefbb028cc3f0ad70dba5a3b8f0299  
## 4 651e9b286029d0c2e79fa13b3866b90a  
## 5 ce247f71bbfd47b642866dd803de4d8f

#DF без 1 столбца  
df <- df %>% select(-X)  
head(df, 5)

## id experience start stop accident id\_rent  
## 1 1 0 351 365 0 a3f7d4a39b65b170005aafbbeed05106  
## 2 2 1 128 149 0 5391009fc72643b4f98a7307baf4a4c9  
## 3 3 1 40 41 0 7b00a8676bfbbf3ca1278d66289505e8  
## 4 4 0 79 147 0 80027a04a698b863ba848ed65ae365bf  
## 5 5 0 53 103 0 fb5375101937f15f22b2c5a5a4bb77f7  
## id\_driver  
## 1 1a8258e7d2adc48ac67a89a1854278a2  
## 2 c8f7037c66966f7979df22c34fc853f0  
## 3 97eefbb028cc3f0ad70dba5a3b8f0299  
## 4 651e9b286029d0c2e79fa13b3866b90a  
## 5 ce247f71bbfd47b642866dd803de4d8f

#Распространенность "опытных"  
prevalence\_experience <- sum(df$experience == 1) / nrow(df)  
cat("Распространенность “experience”:", prevalence\_experience, "\n")

## Распространенность “experience”: 0.32

#Риск “accident” в общей группе, в группе “experience” и в группе без “experience”  
all\_risk\_accident <- mean(df$accident)  
risk\_accident\_experience <- sum(df$accident[df$experience == 1] == 1) / sum(df$experience == 1)  
risk\_accident\_no\_experience <- sum(df$accident[df$experience == 0] == 1) / sum(df$experience == 0)  
cat("Общий риск ДТП (accident):", all\_risk\_accident, "(", (all\_risk\_accident\*100), "%)", "\n")

## Общий риск ДТП (accident): 0.22 ( 22 %)

cat("Риск ДТП среди водителей со стажем:", round(risk\_accident\_experience,2),"(", round(risk\_accident\_experience\*100,2), "%)", "\n")

## Риск ДТП среди водителей со стажем: 0.09 ( 9.38 %)

cat("Риск ДТП среди водителей без стажа:", round(risk\_accident\_no\_experience,2),"(", round(risk\_accident\_no\_experience\*100,2),"%)", "\n")

## Риск ДТП среди водителей без стажа: 0.28 ( 27.94 %)

#Присоединяем столбец с временем аренды  
df <- df %>% mutate(rent\_time = stop - start)  
head(df, 5)

## id experience start stop accident id\_rent  
## 1 1 0 351 365 0 a3f7d4a39b65b170005aafbbeed05106  
## 2 2 1 128 149 0 5391009fc72643b4f98a7307baf4a4c9  
## 3 3 1 40 41 0 7b00a8676bfbbf3ca1278d66289505e8  
## 4 4 0 79 147 0 80027a04a698b863ba848ed65ae365bf  
## 5 5 0 53 103 0 fb5375101937f15f22b2c5a5a4bb77f7  
## id\_driver rent\_time  
## 1 1a8258e7d2adc48ac67a89a1854278a2 14  
## 2 c8f7037c66966f7979df22c34fc853f0 21  
## 3 97eefbb028cc3f0ad70dba5a3b8f0299 1  
## 4 651e9b286029d0c2e79fa13b3866b90a 68  
## 5 ce247f71bbfd47b642866dd803de4d8f 50

#incidence\_rate(All), incidence\_rate(experience), incidence\_rate(no experience)  
  
IR\_all <- sum(df$accident) / sum(df$rent\_time)  
IR\_exp <- sum(df$accident[df$experience == 1]) / sum(df$rent\_time[df$experience == 1])  
IR\_no\_exp <- sum(df$accident[df$experience == 0]) / sum(df$rent\_time[df$experience == 0])  
  
cat("Общая инцидентность (ДТПшность):", round(IR\_all, 4), "\n")

## Общая инцидентность (ДТПшность): 0.003

cat("Инцидентность среди водителей со стажем:", round(IR\_exp,4), "\n")

## Инцидентность среди водителей со стажем: 0.0032

cat("Инцидентность среди водителей без стажа:", round(IR\_no\_exp,4), "\n")

## Инцидентность среди водителей без стажа: 0.003

#**Конклюжнъ**

#Распространенность “experience” - 32% пользователей сервиса имеют водительский стаж;

#Риск “accident”: общий - 22% вероятности того, что любой арендованный автомобиль попадет в ДТП, группа “experience” - 9,38% вероятности того, что автомобиль с “опытным” водителем попадет в ДТП, группа “no\_experience” - 27.94% вероятности того, что автомобиль с “неопытным” водителем попадет в ДТП, что значительно превышает риск у водителя, имеющего опыт вождения.

#Плотность событий (incidence rate) в общей группе - 0.003 означает, что произойдут примерно 3 аварии на каждые 1000 дней аренды.

#IR среди водителей со стажем - 0.00318;

#IR среди водителей без стажа - 0.00303 Показатель ниже, чем у “опытных” водителей. Это доказывает использование показателей инцидентности является более оправданным для поиска “группы риска”.

#Примечание

#Использование информации об id заказа, водителя и транспорта (и другой) в оценке риска ДТП может быть использовано для выявления некоторых зависимостей. Например, длительности аренды перед ДТП, составления вероятного портрета рискованного водителя, наличия скрытых неисправностей авто и др. Я уверен, что это и многое другое уже интегрировано в систему оценки рисков ДТП при аренде.