HW\_morbidity\_measures

Elena Marochkina

2024-09-15

Table of Contents

# Read Data

data <- read\_csv("data/carrental.csv")

## Rows: 100 Columns: 5  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (5): id, experience, start, stop, accident  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

glimpse(data)

## Rows: 100  
## Columns: 5  
## $ id <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, …  
## $ experience <dbl> 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1,…  
## $ start <dbl> 351, 128, 40, 79, 53, 61, 120, 20, 186, 105, 129, 302, 86, …  
## $ stop <dbl> 365, 149, 41, 147, 103, 93, 365, 49, 262, 332, 211, 315, 12…  
## $ accident <dbl> 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0,…

# Calculate prevalence for experience

**Prevalence** is the number of people in the sample with the characteristic of interest (experience), divided by the total number of people in the sample.

calculate\_prevalence <- function(data, experience\_flag) {  
 prevalence <- data %>%  
 summarize(  
 prevalence\_all = n(),  
 prevalence = sum(experience == experience\_flag)  
 ) %>%  
 mutate(prevalence = prevalence / prevalence\_all) %>%  
 pull(prevalence)  
}  
  
prevalence\_experienced <- calculate\_prevalence(data, 1)  
prevalence\_not\_experienced <- calculate\_prevalence(data, 0)  
  
print(paste("Prevalence for experienced drivers is equal to", prevalence\_experienced))

## [1] "Prevalence for experienced drivers is equal to 0.32"

print(paste("Prevalence for non-experienced drivers is equal to", prevalence\_not\_experienced))

## [1] "Prevalence for non-experienced drivers is equal to 0.68"

# Calculate incidence proportion(risk) of accident

**Risk** is the number of events (good or bad) in a treated (exposed) or control (nonexposed) group, divided by the number of people in that group.

risk\_calculation <- function(data, general\_flag, experience\_flag) {  
 if (general\_flag == TRUE) {  
 # Calculate risk for the entire dataset  
 risk\_summary <- data %>%  
 summarize(  
 risk\_all = n(),  
 accident\_count = sum(accident == 1)  
 )  
 } else {  
 # Calculate risk for filtered data based on experience\_flag  
 risk\_summary <- data %>%  
 filter(experience == experience\_flag) %>%  
 summarize(  
 risk\_all = n(),  
 accident\_count = sum(accident == 1)  
 )  
 }  
  
 risk\_summary <- risk\_summary %>%  
 mutate(risk = (accident\_count / risk\_all) \* 100)  
  
 risk\_value <- risk\_summary$risk  
  
 return(risk\_value)  
}  
  
risk\_general <- risk\_calculation(data, TRUE, 0)  
risk\_experienced <- risk\_calculation(data, FALSE, 1)  
risk\_not\_experienced <- risk\_calculation(data, FALSE, 0)  
  
print(paste("Incidence proportion(Risk) for all drivers is equal to", risk\_general, paste("%")))

## [1] "Incidence proportion(Risk) for all drivers is equal to 22 %"

print(paste("Incidence proportion(Risk) for experienced drivers is equal to", risk\_experienced, paste("%")))

## [1] "Incidence proportion(Risk) for experienced drivers is equal to 9.375 %"

print(paste("Incidence proportion(Risk) for non-experienced drivers is equal to", risk\_not\_experienced, paste("%")))

## [1] "Incidence proportion(Risk) for non-experienced drivers is equal to 27.9411764705882 %"

# Calculate incidence rate of accident

**Incidence rate** is the ratio of the number of cases to the total time the population is at risk of disease.

incidence\_rate\_calculation <- function(data, experience\_flag = NA) {  
 # Filter the data if experience\_flag is provided  
 if (!is.na(experience\_flag)) {  
 data <- data %>%  
 filter(experience == experience\_flag)  
 }  
   
 # Calculate observation time and incidence rate  
 incidence\_summary <- data %>%  
 mutate(observation\_time = stop - start) %>%  
 summarize(  
 total\_observation\_time = sum(observation\_time),  
 accident\_count = sum(accident == 1)  
 ) %>%  
 mutate(incidence\_rate = accident\_count / total\_observation\_time)  
   
 incidence\_rate <- incidence\_summary$incidence\_rate  
 return(incidence\_rate)  
}  
  
# Example usage:  
incidence\_rate\_general <- incidence\_rate\_calculation(data)  
incidence\_rate\_experienced <- incidence\_rate\_calculation(data, 1)  
incidence\_rate\_not\_experienced <- incidence\_rate\_calculation(data, 0)  
  
   
print(paste("Incidence rate for all drivers is equal to", incidence\_rate\_general))

## [1] "Incidence rate for all drivers is equal to 0.00304624757684852"

print(paste("Incidence rate for experienced drivers is equal to", incidence\_rate\_experienced))

## [1] "Incidence rate for experienced drivers is equal to 0.00317796610169492"

print(paste("Incidence rate for non-experienced drivers is equal to", incidence\_rate\_not\_experienced))

## [1] "Incidence rate for non-experienced drivers is equal to 0.00302644154189232"

# Summarise results and add logical explanation

## Results

1. Prevalence for experience is equal to **0.32**
2. Incidence proportion(Risk) for all drivers is equal to **22 %**
3. Incidence proportion(Risk) for experienced drivers is equal to **9.375 %**
4. Incidence proportion(Risk) for non-experienced drivers is equal to **27.9411764705882 %**
5. Incidence rate for all drivers is equal to **0.00304624757684852**
6. Incidence rate for experienced drivers is equal to **0.00317796610169492**
7. Incidence rate for non-experienced drivers is equal to **0.00302644154189232**

## Overall Interpretation

1. **Prevalence** tells us that only 32% of the drivers are experienced.
2. **Risk (Incidence Proportion)** shows that non-experienced drivers are significantly more likely to be involved in an accident than experienced drivers (28% vs. 9.4%).
3. **Incidence Rate** adjusts for the time of exposure, showing that experienced drivers have a slightly higher accident rate per unit of time than non-experienced drivers. This could suggest that the longer driving time for experienced drivers leads to more accidents, even though they are generally less risky when viewed as a percentage of the group.