SLEEP PATTERNS BY DIET

GRACY CHANGIRALA

2024-09-03

Table of Contents

# Loading the libraries and Viewing Data

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

data()  
view(msleep)

*This code loads the ‘tidyverse’, package, which is a collection of packages in R for Data manipulation and visualisation. The ‘data()’ function lists all available datasets in R, and ‘view(msleep)’ opens the ‘msleep’ dataset in a new tab for inspections.*

# Descriptive Statistics on ‘msleep’ Data

# Describe the spread, centrality and variance of the data  
min(msleep$awake)

## [1] 4.1

max(msleep$awake)

## [1] 22.1

range(msleep$awake)

## [1] 4.1 22.1

IQR(msleep$awake)

## [1] 5.9

mean(msleep$awake)

## [1] 13.56747

median(msleep$awake)

## [1] 13.9

var(msleep$awake)

## [1] 19.82106

**These statistics give us the insights into the distribution and variability of the ‘awake’ variable.**

*This block of code calculates basic descriptive statistics for the ‘awake’ variable in the ‘msleep’ dataset:*

**- ‘min()’, ‘max()’, ‘range()’**: *Calculates the minumum, maximum and ranges of the dataset.*

**- ‘IQR()’**: *Interquartile range, showing the spread of the middle 50% of the data.*

**-‘mean()’, ‘median()’“**: *Central tendency measures.*

**- ‘var()’**: *Variance, indicating the spread of the data.*

# Summary Statistics for the Entire Dataset

#Summarise selected variables  
summary(msleep)

## name genus vore order   
## Length:83 Length:83 Length:83 Length:83   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## conservation sleep\_total sleep\_rem sleep\_cycle   
## Length:83 Min. : 1.90 Min. :0.100 Min. :0.1167   
## Class :character 1st Qu.: 7.85 1st Qu.:0.900 1st Qu.:0.1833   
## Mode :character Median :10.10 Median :1.500 Median :0.3333   
## Mean :10.43 Mean :1.875 Mean :0.4396   
## 3rd Qu.:13.75 3rd Qu.:2.400 3rd Qu.:0.5792   
## Max. :19.90 Max. :6.600 Max. :1.5000   
## NA's :22 NA's :51   
## awake brainwt bodywt   
## Min. : 4.10 Min. :0.00014 Min. : 0.005   
## 1st Qu.:10.25 1st Qu.:0.00290 1st Qu.: 0.174   
## Median :13.90 Median :0.01240 Median : 1.670   
## Mean :13.57 Mean :0.28158 Mean : 166.136   
## 3rd Qu.:16.15 3rd Qu.:0.12550 3rd Qu.: 41.750   
## Max. :22.10 Max. :5.71200 Max. :6654.000   
## NA's :27

summary(msleep$sleep\_total)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.90 7.85 10.10 10.43 13.75 19.90

msleep %>%   
 select(sleep\_total, brainwt) %>%  
 summary

## sleep\_total brainwt   
## Min. : 1.90 Min. :0.00014   
## 1st Qu.: 7.85 1st Qu.:0.00290   
## Median :10.10 Median :0.01240   
## Mean :10.43 Mean :0.28158   
## 3rd Qu.:13.75 3rd Qu.:0.12550   
## Max. :19.90 Max. :5.71200   
## NA's :27

**Summary statistics gives and overview of each variable, helping identify trends, outliners, and missing values.**

**- ‘summary(msleep)’**: *Provides summary statistics for all variable in the ‘msleep’ dataset.*

**- ‘summary(msleep$sleep\_total)’**: *Specifically summarizes the ‘sleep\_total’ variable.*

**- ‘select(sleep\_total, brainwt) %>% summary’**: *Selects the ‘sleep\_total’ and ‘brainwt’ variables and then summarizes them.*

# Creating a summary Table by Grouping

# Create a summary table  
msleep %>%  
 drop\_na(vore) %>%  
 group\_by(vore) %>%  
 summarise(  
 Lower = min(sleep\_total),  
 Average = mean(sleep\_total),  
 Upper = max(sleep\_total),  
 Difference = max(sleep\_total) - min(sleep\_total)  
 ) %>%  
 arrange(Average)

## # A tibble: 4 × 5  
## vore Lower Average Upper Difference  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 herbi 1.9 9.51 16.6 14.7  
## 2 carni 2.7 10.4 19.4 16.7  
## 3 omni 8 10.9 18 10   
## 4 insecti 8.4 14.9 19.9 11.5

**Grouping and summarising data this way helps compare different categories and understand how they differ in terms of sleep patterns.**

*This code creates a summary table for ‘sleep\_total’ based on the ‘vore’(dietary category) variable:*

**- ‘drop\_na(vore)’**: *Remove rows where ‘vore’ is ‘NA’.*

**- ‘group\_by(vore)’**: *‘Groups data by the ’vore’ category.*

**- ‘summarise()’**: *Calculates the minimum, average, maximum, and range (‘Difference’) of ‘sleep\_total’ within each ‘vore’ category.*

**- arrange(Average)’**” *Sorts the results by the average ‘sleep\_total’.*

# Contingency Tables

# Creating contingency tables  
library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

attach(Cars93)  
  
glimpse(Cars93)

## Rows: 93  
## Columns: 27  
## $ Manufacturer <fct> Acura, Acura, Audi, Audi, BMW, Buick, Buick, Buick,…  
## $ Model <fct> Integra, Legend, 90, 100, 535i, Century, LeSabre, R…  
## $ Type <fct> Small, Midsize, Compact, Midsize, Midsize, Midsize,…  
## $ Min.Price <dbl> 12.9, 29.2, 25.9, 30.8, 23.7, 14.2, 19.9, 22.6, 26.…  
## $ Price <dbl> 15.9, 33.9, 29.1, 37.7, 30.0, 15.7, 20.8, 23.7, 26.…  
## $ Max.Price <dbl> 18.8, 38.7, 32.3, 44.6, 36.2, 17.3, 21.7, 24.9, 26.…  
## $ MPG.city <int> 25, 18, 20, 19, 22, 22, 19, 16, 19, 16, 16, 25, 25,…  
## $ MPG.highway <int> 31, 25, 26, 26, 30, 31, 28, 25, 27, 25, 25, 36, 34,…  
## $ AirBags <fct> None, Driver & Passenger, Driver only, Driver & Pas…  
## $ DriveTrain <fct> Front, Front, Front, Front, Rear, Front, Front, Rea…  
## $ Cylinders <fct> 4, 6, 6, 6, 4, 4, 6, 6, 6, 8, 8, 4, 4, 6, 4, 6, 6, …  
## $ EngineSize <dbl> 1.8, 3.2, 2.8, 2.8, 3.5, 2.2, 3.8, 5.7, 3.8, 4.9, 4…  
## $ Horsepower <int> 140, 200, 172, 172, 208, 110, 170, 180, 170, 200, 2…  
## $ RPM <int> 6300, 5500, 5500, 5500, 5700, 5200, 4800, 4000, 480…  
## $ Rev.per.mile <int> 2890, 2335, 2280, 2535, 2545, 2565, 1570, 1320, 169…  
## $ Man.trans.avail <fct> Yes, Yes, Yes, Yes, Yes, No, No, No, No, No, No, Ye…  
## $ Fuel.tank.capacity <dbl> 13.2, 18.0, 16.9, 21.1, 21.1, 16.4, 18.0, 23.0, 18.…  
## $ Passengers <int> 5, 5, 5, 6, 4, 6, 6, 6, 5, 6, 5, 5, 5, 4, 6, 7, 8, …  
## $ Length <int> 177, 195, 180, 193, 186, 189, 200, 216, 198, 206, 2…  
## $ Wheelbase <int> 102, 115, 102, 106, 109, 105, 111, 116, 108, 114, 1…  
## $ Width <int> 68, 71, 67, 70, 69, 69, 74, 78, 73, 73, 74, 66, 68,…  
## $ Turn.circle <int> 37, 38, 37, 37, 39, 41, 42, 45, 41, 43, 44, 38, 39,…  
## $ Rear.seat.room <dbl> 26.5, 30.0, 28.0, 31.0, 27.0, 28.0, 30.5, 30.5, 26.…  
## $ Luggage.room <int> 11, 15, 14, 17, 13, 16, 17, 21, 14, 18, 14, 13, 14,…  
## $ Weight <int> 2705, 3560, 3375, 3405, 3640, 2880, 3470, 4105, 349…  
## $ Origin <fct> non-USA, non-USA, non-USA, non-USA, non-USA, USA, U…  
## $ Make <fct> Acura Integra, Acura Legend, Audi 90, Audi 100, BMW…

table(Origin)

## Origin  
## USA non-USA   
## 48 45

table(AirBags, Origin)

## Origin  
## AirBags USA non-USA  
## Driver & Passenger 9 7  
## Driver only 23 20  
## None 16 18

addmargins(table(AirBags, Origin))

## Origin  
## AirBags USA non-USA Sum  
## Driver & Passenger 9 7 16  
## Driver only 23 20 43  
## None 16 18 34  
## Sum 48 45 93

table(AirBags, Origin)

## Origin  
## AirBags USA non-USA  
## Driver & Passenger 9 7  
## Driver only 23 20  
## None 16 18

prop.table(table(AirBags, Origin))\*100

## Origin  
## AirBags USA non-USA  
## Driver & Passenger 9.677419 7.526882  
## Driver only 24.731183 21.505376  
## None 17.204301 19.354839

round(prop.table(table(AirBags, Origin), 2)\*100)

## Origin  
## AirBags USA non-USA  
## Driver & Passenger 19 16  
## Driver only 48 44  
## None 33 40

**Contingency tables are useful for exploring relationships between categorical variable.**

**- attach(Cars93)**: *Makes the variables in the Cars93 dataset accessible directly.*

**- glimpse(Cars93)**: *Provides a quick look at the Cars93 dataset structure.*

**- table(Origin)**: *Creates a frequency table for the Origin variable.*

**- table(AirBags, Origin)**: *Creates a contingency table showing the relationship between AirBags and Origin.*

**- addmargins(table(AirBags, Origin))**: *Adds row and column sums (margins) to the contingency table.*

**- prop.table()**: *Converts the contingency table into proportions, showing the percentage of each combination.*

# Pivoting Data for Summary

Cars93 %>%  
 group\_by(Origin, AirBags) %>%  
 summarise(number = n()) %>%  
 pivot\_wider(names\_from = Origin,  
 values\_from = number)

## `summarise()` has grouped output by 'Origin'. You can override using the  
## `.groups` argument.

## # A tibble: 3 × 3  
## AirBags USA `non-USA`  
## <fct> <int> <int>  
## 1 Driver & Passenger 9 7  
## 2 Driver only 23 20  
## 3 None 16 18

**Turning helps in reshaping information, making it more straightforward to picture and analyze bunches next to each other.**

**This code block groups the Cars93 dataset by Origin and AirBags, counts the number of occurrences (n()), and then uses pivot\_wider to reshape the data:**

**- group\_by(Origin, AirBags)**: *Groups the data by Origin and AirBags.*

**- summarise(number = n())**: *Counts the number of cars in each group.*

**- pivot\_wider()**: *Converts the long-format data into a wider format where the Origin categories become separate columns.*

# Conclusion

*In the wake of running the above code blocks, you ought to have a complete synopsis and depiction of the msleep and Cars93 datasets. The examination ought to give experiences into the circulation, focal inclination, fluctuation, and connections inside the information.*

*This sort of examination is essential in information science, assisting you with understanding your information prior to applying more complicated models or representations.*