Data used in this tutorial: car.csv

1. Set up your directory

getwd() : shows the working directory

setwd("C:/desktop/data") for Window or setwd(/Users/data) for Mac : change your working directory

2. Creating directories

dir(): lists files in the working directory after setting up your directory.

3. Install packages

Use Tools in RStudio

4. get help

help(xxx): For example, help(reshape) helps to access codebook.

args(xxx): Description of the command

5. read data

The option header=TRUE implies that the first row of mydata consists of variable names rather than actual data.

A blank space is used to separate columns when sep=" " is specified within its argument. Other popular choices include comma (sep=","), and tab (sep="\t").

mydata <- read.csv("e:\\car.csv", header=TRUE, sep=".", row.names=1)

mydata <-read.csv(file.choose(), header =TRUE): select file from your file

6. explore data

summary(mydata): offers basic descriptive statistics

edit(mydata): open data editor

str(mydata): provides the structure

names(mydata): lists variables

head(mydata, n=8): first 8 rows of dataset. You can remove “, n=8” for the default.

tail(mydata, n=8): last 8 rows

mydata[1:5, 1:5]: first 5 rows of data and first 5 columns of data

7. missing data

rowSums(is.na(mydata): number of missing per row

colSums(is.na(mydata)): number of missing per column

mydata1 <-na.omit(mydata): creating a new dataset without missing data

8. Replace data

mydata1[mydata1$Height ==130] <-150 : replace Height=150 if Height=130 in the original dataset

9. Delete data

mydata1[mydata1$Height ==130] <-NULL

10. descriptive statistics

library(pastecs)

stat.desc(mydata)

mean(mydata1$Height)

median(mydata1$Height)

var(mydata1$Height)

sd(mydata1$Height)

max(mydata1$Height)

min(mydata1$Height)

range(mydata1$Height)

quantile(mydata1$Height)

table(mydata1$Height)

11. plot data

plot(mydata$Length, mydata$Weight)

plot(mydata$Height, mydata$Weight)

plot(mydata$Length, mydata$Weight, main="Weight/Length", xlab="Length", ylab="Weight", col="red")

lines(lowess(mydata$Length, mydata$Weight), col="green")

lines(mydata$Length, mydata$Weight, col="green")

hist(mydata$Height)

boxplot(mydata$Height)

library(rgl)

library(car)

scatter3d(z= mydata$Height, x= mydata$Length, y= mydata$Weight, surface = F)

12. linear regression

result = lm(mydata$Weight ~mydata$Height + mydata$Length)

summary(result)