Getting Started

Exercises (page10)

library(downloader)

url <- "https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/femaleMiceWeights.csv"

filename <- "femaleMiceWeights.csv"

download(url, destfile=filename)

library(devtools)

dir <- system.file(package="dagdata")

list.files(file.path(dir,"extdata"))

filename <- file.path(dir,"extdata/femaleMiceWeights.csv")

dat <- read.csv(filename)

str(dat)

dat # exercise-1

dat[12,2] # exercise-2

dat $ Bodyweight # exercise-3 select 11th row

dat [11, 2] # exercise-3

Ln <- length(dat $ Bodyweight)

Ln # exercise-4

hf <- dat[13:24,2]

hf\_mean <- mean(hf)

hf\_mean # exercise-5

hf\_sample <- sample(hf, 1)

hf\_sample

set.seed(1)

hf\_sample # exercise-6

Exercises (page13)

msleep <- read.csv("msleep\_ggplot2.csv")

class(msleep) # exercise-1

head(primates) # primates < order

str(msleep)

library(dplyr)

primates <- filter(msleep, order=="Primates")

head(primates)

nrow(primates) # exercise-2

class(primates) # exercise-3

sleep\_p <- primates %>% select(sleep\_total)

class(sleep\_p) # exercise-4

sleep\_pt <- primates %>% select(sleep\_total) %>% unlist

class(sleep\_pt)

mean(sleep\_pt) # exercise-5

sleep\_pt2 <- primates %>% summarize(mean(sleep\_total)) # exercise-6

sleep\_pt2

Inference

Exercises (page30)

library(downloader)

url <- "https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/femaleControlsPopulation.csv"

filename <- basename(url)

download(url, destfile=filename)

x <- unlist( read.csv(filename) )

all\_M <- mean(x) # exercise-1 23.89338

all\_M

set.seed(1)

y <- sample(x, 5)

y

sample\_M <- mean(y)

sample\_M

difA <- abs(sample\_M - all\_M) # exercise-2 0.2706222

difA

set.seed(5)

z <- sample(x, 5)

difB <- abs(mean(z)-mean(x)) # exercise-3 1.433378

difB

# exercise-4 C

set.seed(1)

n <- 1000

samA <- vector("numeric",n)

for (i in 1:n) {

y <- sample(x,5)

samA[i] <- mean(y)

}

dif1 <- abs(samA - all\_M)

mean(dif1 >=1) # exercise-5 0.498

m <- 10000

set.seed(1)

samB <- vector("numeric",m)

for (i in 1:m) {

z <- sample(x,5)

samB[i] <- mean(z)

}

dif2 <- abs(samB - all\_M)

mean(dif2 >=1) # exercise-6 0.4976

set.seed(1)

f <- 1000

samC <- vector("numeric",f)

for (i in 1:n) {

w <- sample(x,50)

samC[i] <- mean(w)

}

dif3 <- abs(samC - all\_M)

mean(dif3 >=1) # exercise-7 0.019

library(rafalib)

mypar(1,2)

hist(samA)

hist(samC)

# exercise-8 C

ex <- mean(samC>=23 & samC<=25) # exercise-9 0.976

pnorm(25, 23.9, 0.43) - pnorm(23, 23.9, 0.43) # exercise-10 0.9765648

Exercises (page33)

library(downloader)

url <- "https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/mice\_pheno.csv"

filename <- basename(url)

download(url, destfile=filename)

dat <- read.csv(filename)

dat <- na.omit( dat )

library(dplyr)

x <- filter(dat,Sex == "M" & Diet == "chow") %>%

select(Bodyweight) %>% unlist

mean(x) # exercize-1

library(rafalib)

popsd(x) # exercize-2

set.seed(1)

X <- sample(x, 25)

mean(X) # exercize-3

set.seed(1)

y <- filter(dat,Sex == "M" & Diet == "hf") %>%

select(Bodyweight) %>% unlist

mean(y) # exercize-4

library(rafalib)

popsd(y) # exercize-5

set.seed(1)

Y <- sample(y, 25)

mean(Y) # exercize-6

popsd(Y)

p <- mean(y)-mean(x)

q <- mean(X)-mean(Y)

r <- abs(p-q) # exercize-7 6.556516

# Female

library(dplyr)

xF <- filter(dat,Sex == "F" & Diet == "chow") %>%

select(Bodyweight) %>% unlist

mean(xF)

library(rafalib)

popsd(xF)

set.seed(1)

XF <- sample(xF, 25)

mean(XF)

set.seed(1)

yF <- filter(dat,Sex == "F" & Diet == "hf") %>%

select(Bodyweight) %>% unlist

mean(yF)

library(rafalib)

popsd(yF)

set.seed(1)

YF <- sample(yF, 25)

mean(YF)

popsd(YF)

pF <- mean(yF)-mean(xF)

qF <- mean(XF)-mean(YF)

rF <- abs(pF-qF) # exercise-8 5.487517

rF

# exercise-9 A

Exercises (page126)

# exercise 1 6

boxplot(InsectSprays$count ~ InsectSprays$spray)

# exercise 2 B?

load("~/Kyoto/skew.RData")

head(dat)

summary(dat)

str(dat)

boxplot(dat)

# exercise 3 V4

# exercise 4 V9

library(dplyr)

load("/Users/TOMO/Downloads/UsingR/data/nym.2002.rda")

head(nym.2002)

summary(nym.2002)

str(nym.2002)

boxplot(nym.2002$time ~ nym.2002$gender)

Mtime <- filter(nym.2002, gender=="Male") %>% select(time) %>% unlist

Ftime <- filter(nym.2002, gender=="Female") %>% select(time) %>% unlist

library(rafalib)

mypar(1,2)

hist(Mtime, xlim=c(100,600))

hist(Ftime, xlim=c(100,600), ylim=c(0,250))

# exercise 5 C

Male <- filter(nym.2002, gender=="Male")

Female <- filter(nym.2002, gender=="Female")

cor(Male$age, Male$time) # exercise 6 0.2432273

cor(Female$age, Female$time) # exercise 7 0.2443156

library(rafalib)

mypar(1,2)

plot(Male$age, Male$time, cex=0.5, xlim=c(0,80), xaxp=c(0,80,16))

plot(Female$age, Female$time, cex=0.5, xlim=c(0,80), xaxp=c(0,80,16))

mage <- cut( Male$age, breaks=seq( 5, 80, 5 ), right=FALSE, ordered\_result=TRUE )

boxplot(Male$time~mage)

fage <- cut( Female$age, breaks=seq( 5, 80, 5 ), right=FALSE, ordered\_result=TRUE )

boxplot(Female$time~fage)

# exercise 8 D

# exercise 9 D

# exercise 10 D

Exercises (page134)

data(ChickWeight)

head(ChickWeight)

plot( ChickWeight$Time, ChickWeight$weight, col=ChickWeight$Diet)

chick = reshape(ChickWeight, idvar=c("Chick","Diet"), timevar="Time",direction="wide")

chick = na.omit(chick)

chick$weight.4

c4w<-chick$weight.4

c4w1<-c(cw,3000)

Mc4w<-mean(c4w)

Mc4w1<-mean(c4w1)

Mc4w # 60.15556

Mc4w1 # 124.0652

Mc4w1/Mc4w # exercise 1 2.062407

Mdc4w<-median(c4w) # 61

Mdc4w1<-median(c4w1) # 61

Mdc4w1/Mdc4w # exercise 2 1

SDc4w<-sd(c4w) # 4.279738

SDc4w1<-sd(c4w1) # 433.4771

SDc4w1/SDc4w # exercise 3 101.2859

MADc4w<-mad(c4w) # 4.4478

MADc4w1<-mad(c4w1) # 4.4478

MADc4w1/MADc4w # exercise 4 1

c21w<-chick$weight.21

c21w1<-c(c21w,3000)

Cc4w<-cor(c4w,c21w) # 0.4159499

Cc21w<-cor(c4w1,c21w1) # 0.9861002

Cc21w/Cc4w # exercise 5 2.370719

x<-chick$weight.4[chick$Diet=="1"]

y<-chick$weight.4[chick$Diet=="4"]

t.test(x,y) # p-value = 7.32e-06

wilcox.test(x,y)

x1<-c(x,200)

t.test(x1,y) # p-value = 0.938

t.test(x1,y)$p.value # exercise 6 p-value = 0.9380347

wilcox.test(x1,y) # exercise 7 p-value = 0.0009841

library(rafalib)

mypar(1,3)

boxplot(x,y)

boxplot(x,y+10)

boxplot(x,y+100)

t.test(x,y)$statistic # -5.839339

t.test(x,y+10)$statistic # -13.36722

t.test(x,y+100)$statistic # -81.11819

t.test(x,y+10)$statistic-t.test(x,y+100)$statistic

# exercise 8 67.75097

wilcox.test(x,y)$statistic # 6

wilcox.test(x,y+10)$statistic # 0

wilcox.test(x,y+100)$statistic # 0

wilcox.test(c(1,2,3),c(4,5,6) ) # exercise 9 p-value = 0.1

wilcox.test(c(1,2,3),c(400,500,600) ) # exercise 10 p-value = 0.1