# P38 Exercises

# Q1; Answer: 67.6%

# Q2; Answer: 96.1%

# Q3; Answer: 99.4%

# Q4; Answer: 67.7%

# Q5; Answer: 93.7%

# Q6; Answer: 98.6%

# Q7; Answer: C)

# Q8; Answer: B)

# Q9; Answer: 30.95

# Q10; Answer: 4.28

# Q11; Answer: A)

# Q12; Answer: 12.8%

# Q13; Answer: C)

# 以下、解答を導き出す根拠として使用したRソースコードを記載する

# 準備の部

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 )

# Q1

hist(dat$Bodyweight)

mean(dat$Bodyweight) # 28.84705

var(dat$Bodyweight) # 39.19297

sd(dat$Bodyweight) # 6.260429

Mean <- mean(dat$Bodyweight)

SD <- sd(dat$Bodyweight)

Mean - SD

Mean + SD

A <- which(dat$Bodyweight >= Mean - SD)

B <- which(dat$Bodyweight <= Mean + SD)

length(which(is.element(A, B) == TRUE)) / nrow(dat) # Answer: 67.6%

# Q2

Mean - 2\*SD

Mean + 2\*SD

A2 <- which(dat$Bodyweight >= Mean - 2\*SD)

B2 <- which(dat$Bodyweight <= Mean + 2\*SD)

length(which(is.element(A2, B2) == TRUE)) / nrow(dat) # Answer: 96.1%

# Q3

Mean - 3\*SD

Mean + 3\*SD

A3 <- which(dat$Bodyweight >= Mean - 3\*SD)

B3 <- which(dat$Bodyweight <= Mean + 3\*SD)

length(which(is.element(A3, B3) == TRUE)) / nrow(dat) # Answer: 99.4%

# Q4

library(rafalib)

length(which(dat$Sex=="M" & dat$Diet=="chow") == TRUE)

mean(dat$Sex=="M" & dat$Diet=="chow")

sd(dat$Sex=="M" & dat$Diet=="chow")

popsd(dat$Sex=="M" & dat$Diet=="chow")

Dat2 <- dat[ dat$Sex=="M" & dat$Diet=="chow", ]

hist(Dat2$Bodyweight)

Mean2 <- mean(Dat2$Bodyweight)

PopSD2 <- popsd(Dat2$Bodyweight)

A4 <- which(Dat2$Bodyweight >= Mean - PopSD2)

B4 <- which(Dat2$Bodyweight <= Mean + PopSD2)

length(which(is.element(A4, B4) == TRUE)) / nrow(Dat2) # Answer: 67.7%

# Q5

A5 <- which(Dat2$Bodyweight >= Mean - 2\* PopSD2)

B5 <- which(Dat2$Bodyweight <= Mean + 2\* PopSD2)

length(which(is.element(A5, B5) == TRUE)) / nrow(Dat2) # Answer: 93.7%

# Q6

A6 <- which(Dat2$Bodyweight >= Mean - 3\* PopSD2)

B6 <- which(Dat2$Bodyweight <= Mean + 3\* PopSD2)

length(which(is.element(A6, B6) == TRUE)) / nrow(Dat2) # Answer: 98.6%

# Q7

qqnorm(dat$Bodyweight) # Q1-3の条件(全データでプロット)

qqline(dat$Bodyweight)

Dat2 <- dat[ dat$Sex=="M" & dat$Diet=="chow", ]

# Q4-6の条件(Sex=="M" & Diet=="chow)

qqnorm(Dat2$Bodyweight)

qqline(Dat2$Bodyweight)

# Answer: C)

# Q8

# 4群の体重の平均値、標準偏差を求め、QQプロットを描画

Dat1 <- dat[ dat$Sex=="M" & dat$Diet=="hf", ]

length(Dat1$Bodyweight)

mean(Dat1$Bodyweight)

sd(Dat1$Bodyweight)

popsd(Dat1$Bodyweight)

qqnorm(Dat1$Bodyweight)

qqline(Dat1$Bodyweight)

Dat2 <- dat[ dat$Sex=="M" & dat$Diet=="chow", ]

length(Dat2$Bodyweight)

mean(Dat2$Bodyweight)

sd(Dat2$Bodyweight)

popsd(Dat2$Bodyweight)

qqnorm(Dat2$Bodyweight)

qqline(Dat2$Bodyweight)

Dat3 <- dat[ dat$Sex=="F" & dat$Diet=="hf", ]

length(Dat3$Bodyweight)

mean(Dat3$Bodyweight)

sd(Dat3$Bodyweight)

popsd(Dat3$Bodyweight)

qqnorm(Dat3$Bodyweight)

qqline(Dat3$Bodyweight)

Dat4 <- dat[ dat$Sex=="F" & dat$Diet=="chow", ]

length(Dat4$Bodyweight)

mean(Dat4$Bodyweight)

sd(Dat4$Bodyweight)

popsd(Dat4$Bodyweight)

qqnorm(Dat4$Bodyweight)

qqline(Dat4$Bodyweight)

# Answer: B)

#Q9

library(dplyr)

set.seed(1);

y <- filter(dat, Sex=="M" & Diet=="chow") %>% select(Bodyweight) %>% unlist

avgs <- replicate(10000, mean( sample(y, 25)))

mean(avgs)

hist(avgs)

qqnorm(avgs)

qqline(avgs)

# Answer: 30.95

#Q10

set.seed(1);

y <- filter(dat, Sex=="M" & Diet=="chow") %>% select(Bodyweight) %>% unlist

avgsd <- replicate(10000, popsd( sample(y, 25)))

mypar(1,2)

mean(avgsd)

hist(avgsd)

qqnorm(avgsd)

qqline(avgsd)

# Answer: 4.28

#Q11

popsd(y) # A) 4.420501

popsd(avgs)/sqrt(25) # B) 0.1673638

sqrt(25)/popsd(y) # C) 1.131094

popsd(y)/sqrt(25) # D) 0.8841001

# Answer: A)

#Q12

set.seed(1);

y <- filter(dat, Sex=="M" & Diet=="chow") %>% select(Bodyweight) %>% unlist

avgsd <- replicate(10000, popsd( sample(y, 25)))

mypar(1,2)

mean(avgsd)

hist(avgsd)

qqnorm(avgsd)

qqline(avgsd)

SDbelow3.5 <- which(avgsd<=3.5)

length(SDbelow3.5 == TRUE) / length(avgsd) # Answer: 12.0%

#Q13

x=seq(0.0001, 0.9999, length=300)

curve(dnorm(x) ,-4,4,type="l")

curve(dt(x, 3),-4,4,type="l")

curve(dt(x, 10),-4,4,type="l")

curve(dt(x, 30),-4,4,type="l")

curve(dt(x, 100),-4,4,type="l")

# 描画した図より、自由度30以上のときに正規分布とよく一致する

# Answer: C)