**# HW6**

**# Wenhui Yang**

**# 2/16/2019**

**# Exercise 1**

set.seed(pi)

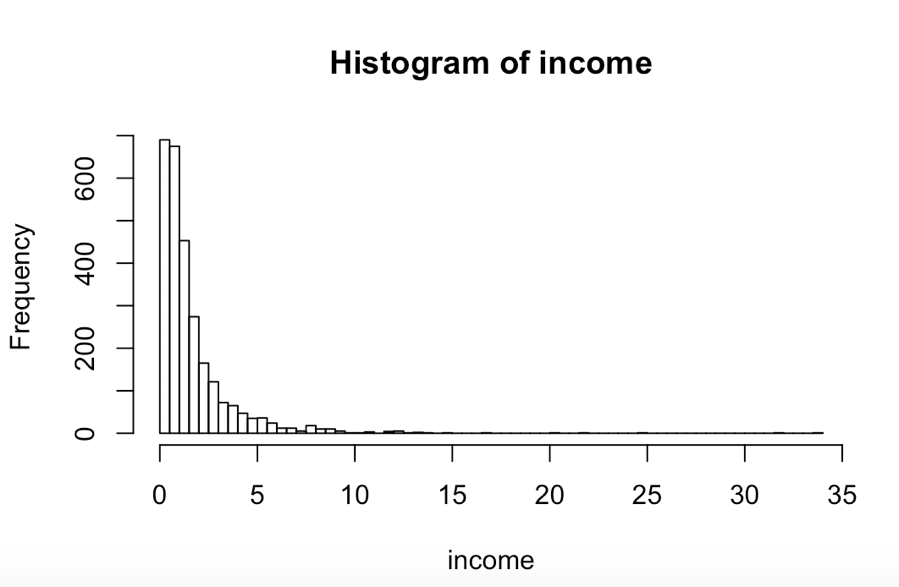
r <- rnorm(2754, 0, 1)

income <- exp(r)

hist(income, breaks=100)

#

quant <- quantile(income, c(0.25, 0.5, 0.75, 0.95))



> quant

25% 50% 75% 95%

0.4992276 1.0122402 1.9345836 5.2582808

#

income.level <- rep(NA,2754)

income.level[income <= quant[1]] <- "VL"

income.level[income > quant[1] & income <= quant[2]] <-"L"

income.level[income > quant[2] & income <= quant[3]] <-"M"

income.level[income > quant[3] & income <= quant[4]] <-"H"

income.level[income > quant[4]] <- "VH"

#

factor\_income.level <- factor(income.level, ordered = TRUE, levals <- c("VL","L","M","H","VH") )

levels(factor\_income.level) <- c("Very Low", "Low", "Middle", "High", "Very High")

summary(factor\_income.level)

> summary(factor\_income.level)

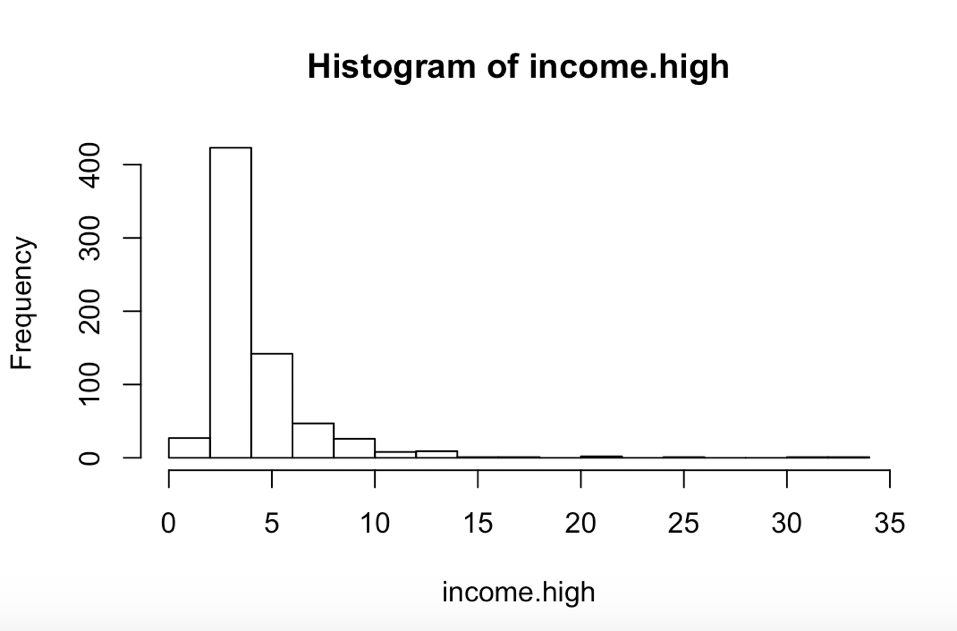
Very Low Low Middle High Very High

689 688 688 551 138

#

income.high <- income[factor\_income.level >= "High"]

hist(income.high, breaks = 20)



#

mean(income[factor\_income.level >= "Middle" & factor\_income.level <= "High"])

[1] 2.14511

#

averageVH <- mean(income[factor\_income.level == "Very High"])

averageH <- mean(income[factor\_income.level == "High"])

difference <- averageH – averageVH

difference

> difference

[1] -5.285512

**# Exercise 2**

industry <- sample(c("Manufacture", "Service", "IT"), 100, replace=TRUE, prob=c(0.3,0.5, 0.2))

stock <- rep(NA,100)

stock[industry == "Manufacture"] <- rnorm(sum(industry=="Manufacture"), 3, 2)

stock[industry == "Service"] <- rnorm(sum(industry=="Service"), 2, 4)

stock[industry == "IT"] <- rnorm(sum(industry=="IT"), 8, 8)

factor.industry <- factor(industry)

#

summary(factor.industry)

> summary(factor.industry)

IT Manufacture Service

23 28 49

#

mean(stock[factor.industry == "Manufacture"])

mean(stock[factor.industry == "Service"])

mean(stock[factor.industry == "IT"])

sd(stock[factor.industry == "Manufacture"])

sd(stock[factor.industry == "Service"])

sd(stock[factor.industry == "IT"])

> mean(stock[factor.industry == "Manufacture"])

[1] 2.437437

> mean(stock[factor.industry == "Service"])

[1] 2.014805

> mean(stock[factor.industry == "IT"])

[1] 5.477804

> sd(stock[factor.industry == "Manufacture"])

[1] 2.190272

> sd(stock[factor.industry == "Service"])

[1] 4.03468

> sd(stock[factor.industry == "IT"])

[1] 7.459002