중간고사 REPORT

201031441 최관현

1. 데이터 정의

이 데이터는 한국인의 신체의 수치를 분석한 데이터이다. 칼럼수는 148개이며 로우는 6421건이 있다.

1. R 스크립트

DF <- read.csv("D:/R/myData.csv",header=T) #excel파일을 csv로 형식변환한 상태

ncol(DF) #148

nrow(DF) #6421

DF\_my <- DF[,c(1,2,7,18,46,57,84)]

class(DF\_my)

colnames(DF\_my) <- c("Sex","Age","BMI","Height","Weight","Waist","Leg")

library(doBy)

Data1 <- summaryBy(DF\_my$Weight ~ DF\_my$Age, na.rm=T, DF\_my)

Data2 <- summaryBy(DF\_my$Waist ~ DF\_my$Age, na.rm=T, DF\_my)

Data3 <- summaryBy(DF\_my$Weight ~ DF\_my$Sex, na.rm=T, DF\_my)

plot(Data1)

plot(Data2)

plot(Data3)

library(plyr)

Data4 <- split(DF\_my$BMI,DF\_my$Sex)

sapply(Data4.mean) #sapply를 이용하여 split으로 나이별로 BMI를 계산햇던 것을 평균내기

barplot(sapply(Data4,mean))

Data5 <- as.data.frame(sapply(Data4,mean))

Freq.Data4 <- prop.table(as.data.frame(sapply(Data4,mean))) # 상대분포 구하기

sum(Freq.Data4) # 모두 더해서 1이 나오는 것을 확인

Freq.Data4 <- t(Freq.Data4) #전치행렬 사용하기 보기 쉽게 만들어 준다

Data6 <- rbind(Data5,Freq.Data4)

Data6 <- addmargins(Data6, margin = 2)

# 성별별 BMI 평균과 상대도수분포표를 구했다.

# 하나의 데이터로 묶어주어 도수분포표를 만들었다.

Weight <- cut(DF\_my$Weight,breaks=4)

FreqOfWeight <- table(Weight)

FreqOfWeight <- rbind(FreqOfWeight,prop.table(FreqOfWeight))

rownames(FreqOfWeight)[2] <- "RelativeFreq"

# 상대도수를 구하고 2번째 행의 이름을 지어준다

FreqOfWeight <- rbind(FreqOfWeight,(CumuFreq <- cumsum(FreqOfWeight[2,])))

rownames(FreqOfWeight)[3] <- "CumuFreq"

# 누적상대도수를 구하고 3번째 행의 이름을 지어준다

rownames(FreqOfWeight) <- c("도수","상대도수","누적도수")

FreqOfWeight <- addmargins(FreqOfWeight,margin=2)

FreqOfWeight

# 총 무게에 따른 4구역으로 나눈 누적상대분포표까지 완성

# 그렇다면 시각화를 시작하여

# 성별별 무게 평균을 이용하여 barplot 나타내기

# barplot은 명목형 변수를 시각화하기 좋은 그래프다

VIZ1 <- Data5[-2,-4]

VIZ1 <- t(data.frame(VIZ1))

rownames(VIZ1)[1] <- "Sex"

VIZ1

barplot(VIZ1)

# 아까전의 무게를 4구역으로 나눈 FreqOfWeight를 그래프로

FreqOfWeight

# 상대도수와 누적도수를 제거

library(ggplot2)

library(ggthemes)

VIZ2 <- t(data.frame(FreqOfWeight[1,]))

rownames(VIZ2)[1] <- "Weight"

class(VIZ2)

VIZ2 <- as.data.frame(VIZ2)

# 다른방향으로 다시 분석

range(DF$Sex,na.rm = T)

# 나이가 제일 적은 것과 많은 것을 출력, na.rm=T 인자를 넣어줘야 결측치를 제외

# 제일 나이가 어린것의 정보를 출력해보고 싶다

subset(DF,subset=(Sex==15))

# Hegiht 칼럼을 2번째 자리까지 구해서 반올림한걸 숫자를 센 것

V <- table(round(DF$Height,digits=2))

library(ggplot2)

qplot(Length,data=DF,geom="bar")

# 막대그래프를 이용해 나타내보았다.

# 분할표를 이용하기, 나이대별 길이와 무게를 나타냄

# table(DF$Sex,DF$Sex) \*\*\*\*계산안됨\*\*\*\*

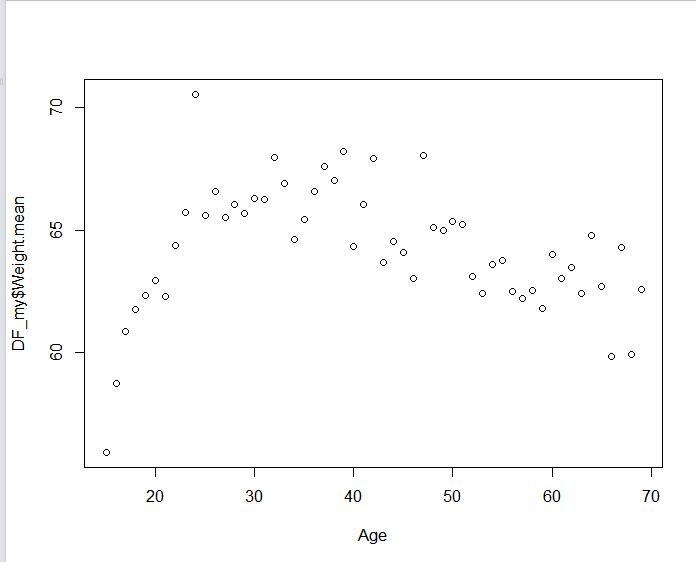
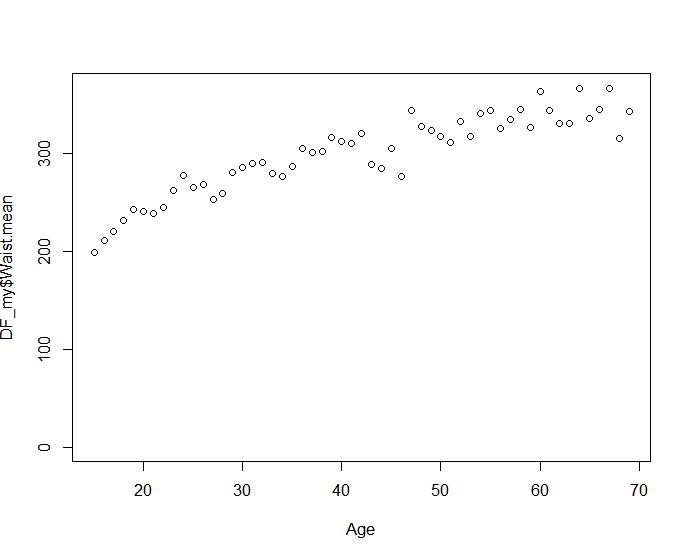
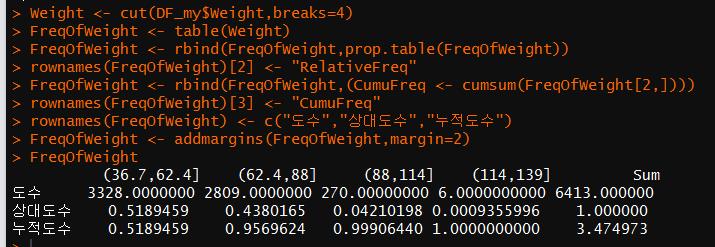
# 다시 몸무게를 20단계로 나눔

VV <- table(cut(DF$Height,breaks = 20))

library(ggplot2)

ggplot(data=DF, aes(x=DF$Height, y=DF$Weight))+geom\_point()

1. 결과



수치상으로 확인을 했을 때 우리나라 신체 사이즈가 서구화가 이루어졌다는 것을 확인을 하였고, PDF 파일에 나오는 자료를 같이 참조했을 때, 예전에 비해 키나 몸무게가 많이 증가하는 것을 볼 수 있었다.

1. 해석과 결론(의사결정)

참고사항 :  
4.2 제7차 한국인 인체치수 조사 결과 요약

4.2.1 평균키의 변화

ㅇ 1979년 제1차부터 2015년 제7차 한국인인체치수조사까지 결과에 따르면 35년간 한 국인의 체형은 키는 커지고 다리길이도 길어지는 등 서구형 체형으로 변해가고 있 으며 2000년대 이후 체형의 서구화경향은 작은 폭으로 진행 중

ㅇ 평균키 : 2015년 현재 20대 평균키는 남자 173.9㎝, 여자 160.9㎝ 30대 평균키는 남 자 173.1㎝, 여자 160.2㎝, 40대 평균키는 남자 170.4㎝, 여자 157.0㎝ 로 나타남. 1979년 이후 전체 연령대에서 남자는 5㎝~7.6㎝, 여자는 3.7~6.5㎝ 가량 평균키가 커졌으며, 남녀 모두 30대 전반(30~34세)에서 가장 큰 변화를 보이며 남자는 약 7.6 ㎝ 가 커진 173.7㎝, 여자는 약 6.5㎝가 커진 160.2 ㎝로 나타남

4.2.2 신체비례의 변화

ㅇ 키에 대한 다리길이(inseam, 안다리길이)의 비율은 2004년, 2010년도에 비해 20세 이후 성인여자 전체에서 유의미하게 커져, 다리길이 비율이 길어짐.

ㅇ 복부비만의 지표가 되는 허리둘레는 1979년 이후 전체 연령대에서 남자는 3.6㎝~10.4 ㎝, 여자는 3.1~5.5㎝ 가량 커짐. 비교적 식생활이 비슷한 2004년과 비교할 때도 남녀 모두 전체적으로 허리둘레가 커졌으며, 특히 남자는 20~34세에서 2.3㎝~2.8㎝, 여자는 17세~49세에 거쳐 2.2㎝~5.6㎝ 증가하여 복부비만이 광범위한 연령대에서 진행되고 있는 것으로 나타남.

다음과 같은 사실도 참고하여 보면,

문제는 작은 치수 옷의 재고가 늘었으며, 큰 치수의 옷은 빨리 매진되었다는 것이다.

다음 수치와 그래프들을 참고삼아 첫째, 앞으로 우리 회사의 S, M, L, XL 의 옷의 치수에 대한 재정의가 필요하다고 결론을 지었다. 우리나라의 서구화된 체형에 따라 전체적인 현 우리나라 사람들에 체형에 맞는 각 S,M,L 사이즈로 바꾸어야한다. 둘째, XXL 사이즈의 필요성을 느끼게 되어 XXL 사이즈의 생산을 계획할 것이다. 종종 아웃라이어가 큰 것들을 관찰할 수 있었다. 셋째, 우리나라의 체형이 많이 서구화가 일어나서 사이즈의 전체적인 재정의가 필요하다는 것을 알 수 있었다. 그렇기 때문에 작은 사이즈보다 큰 사이즈의 생산을 늘려야 한다.