고급R활용 과제

# 1장 연습문제

## 4번 문제 : 자료 1.1을 이용하여 다음에 답하세요.

### (1) 남녀별 값 m은 1로, f는 2로 변환하고, 각각 도수를 구하시요.

setwd("D:/Work\_R/R-Project")  
insurance.rawdata <- read.table("insurance.txt", header=T)  
insurance.data <- insurance.rawdata  
  
head( insurance.data )

## id sex job religion edu amount salary  
## 1 1 m 1 1 3 7.0 110  
## 2 2 m 2 1 4 12.0 135  
## 3 3 f 2 3 5 8.5 127  
## 4 4 f 3 3 5 5.0 150  
## 5 5 m 1 3 3 4.5 113  
## 6 6 m 2 1 2 3.5 95

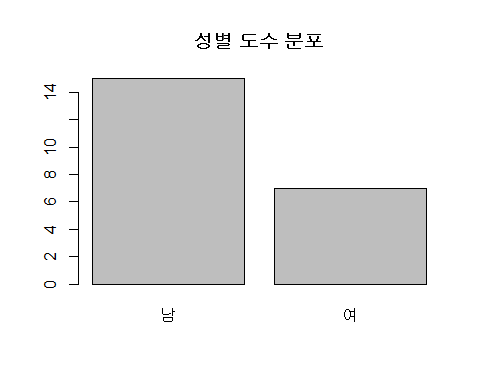
str( insurance.data)

## 'data.frame': 22 obs. of 7 variables:  
## $ id : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ sex : Factor w/ 2 levels "f","m": 2 2 1 1 2 2 2 1 1 2 ...  
## $ job : int 1 2 2 3 1 2 3 3 2 1 ...  
## $ religion: int 1 1 3 3 3 1 2 2 3 3 ...  
## $ edu : int 3 4 5 5 3 2 4 4 4 5 ...  
## $ amount : num 7 12 8.5 5 4.5 3.5 4 4 4.5 17 ...  
## $ salary : int 110 135 127 150 113 95 102 122 110 200 ...

index <- which(insurance.data$sex == "f")  
sex2 <- rep( 1, length(insurance.data$sex) )  
sex2[ index ] <- 2  
  
insurance.data$sex2 <- factor( sex2, level=c(1:2), labels=c("남", "여"))  
  
sex.table <- table( insurance.data$sex2 )  
sex.table

##   
## 남 여   
## 15 7

barplot( sex.table, main="성별 도수 분포" )



### (2) 변수 "종교"에서 1과 2를 1로, 3은 2로 변환한 새로운 변수를 만들고, 각 값에 적합한 라벨을 지정하시요.

#install.packages("car")  
library(car)  
insurance.data$religion2 <- recode( insurance.data$religion, "2=1;3=2" )  
head( insurance.data, n=7 )

## id sex job religion edu amount salary sex2 religion2  
## 1 1 m 1 1 3 7.0 110 남 1  
## 2 2 m 2 1 4 12.0 135 남 1  
## 3 3 f 2 3 5 8.5 127 여 2  
## 4 4 f 3 3 5 5.0 150 여 2  
## 5 5 m 1 3 3 4.5 113 남 2  
## 6 6 m 2 1 2 3.5 95 남 1  
## 7 7 m 3 2 4 4.0 102 남 1

str( insurance.data)

## 'data.frame': 22 obs. of 9 variables:  
## $ id : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ sex : Factor w/ 2 levels "f","m": 2 2 1 1 2 2 2 1 1 2 ...  
## $ job : int 1 2 2 3 1 2 3 3 2 1 ...  
## $ religion : int 1 1 3 3 3 1 2 2 3 3 ...  
## $ edu : int 3 4 5 5 3 2 4 4 4 5 ...  
## $ amount : num 7 12 8.5 5 4.5 3.5 4 4 4.5 17 ...  
## $ salary : int 110 135 127 150 113 95 102 122 110 200 ...  
## $ sex2 : Factor w/ 2 levels "남","여": 1 1 2 2 1 1 1 2 2 1 ...  
## $ religion2: num 1 1 2 2 2 1 1 1 2 2 ...

insurance.data$religion2 <- factor( insurance.data$religion2, level=c(1:2), labels=c("유", "무"))  
head( insurance.data, n=7 )

## id sex job religion edu amount salary sex2 religion2  
## 1 1 m 1 1 3 7.0 110 남 유  
## 2 2 m 2 1 4 12.0 135 남 유  
## 3 3 f 2 3 5 8.5 127 여 무  
## 4 4 f 3 3 5 5.0 150 여 무  
## 5 5 m 1 3 3 4.5 113 남 무  
## 6 6 m 2 1 2 3.5 95 남 유  
## 7 7 m 3 2 4 4.0 102 남 유

### (3) 보험금애글 로그변화한 새로운 변수를 만드시요.

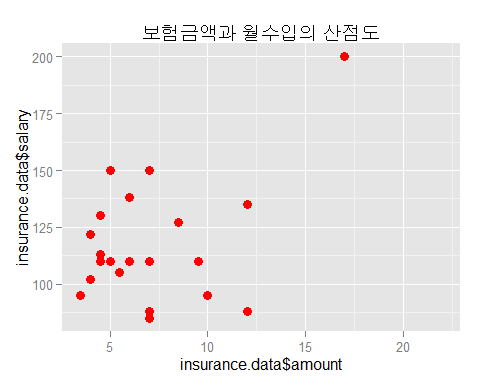
insurance.data$amount\_log <- log(insurance.data$amount)  
head( insurance.data )

## id sex job religion edu amount salary sex2 religion2 amount\_log  
## 1 1 m 1 1 3 7.0 110 남 유 1.945910  
## 2 2 m 2 1 4 12.0 135 남 유 2.484907  
## 3 3 f 2 3 5 8.5 127 여 무 2.140066  
## 4 4 f 3 3 5 5.0 150 여 무 1.609438  
## 5 5 m 1 3 3 4.5 113 남 무 1.504077  
## 6 6 m 2 1 2 3.5 95 남 유 1.252763

### (4) 보험금액과 월수입의 산점도 및 (3)에서 로그변환한 변수와 월수입의 산점도를 그리시오.

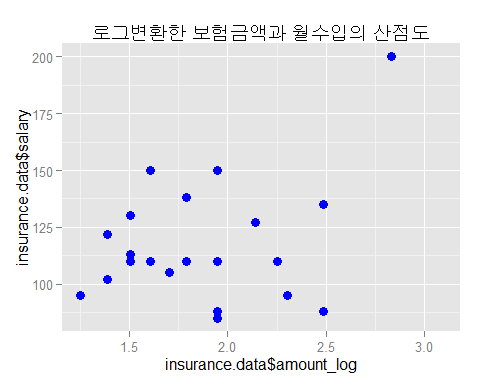
#install.packages("ggplot2")  
library(ggplot2)  
ggplot(data=insurance.data, aes(x=insurance.data$amount, y=insurance.data$salary)) +   
 geom\_point(shape=19, size=3, colour="red") +  
 ggtitle("보험금액과 월수입의 산점도")

## Warning: Removed 1 rows containing missing values (geom\_point).



ggplot(data=insurance.data, aes(x=insurance.data$amount\_log, y=insurance.data$salary)) +   
 geom\_point(shape=19, size=3, colour="blue") +  
 ggtitle("로그변환한 보험금액과 월수입의 산점도")

## Warning: Removed 1 rows containing missing values (geom\_point).



### (1) 템플릿