의사결정모형/신경망 모형 R코드

나오는 코드 이외에 다른 방법도 있으니 반드시 구글링을 하면서 숙지하세요.

……………………………………………………………………………………………………………………………………………………………………………

1. 신경망 모형

먼저 엑셀,CSV로 된 파일을 R에 입력을 해야한다. 여기에서는 과제를 할때 조사한 데이터를 가지고 신경망 모형을 만들어 볼 것이다.

데이터를 입력받는 R코드는 다음과 같다.

#install.package(xlsx)  
library(xlsx)  
  
data\_file = read.xlsx("C:/Users/madda/Documents/동아리/R문서(개인)/연습자료.xlsx",  
 1,  
 colIndex = 1 : 9,  
 rowIndex = 1: 125,  
 header = T,  
 encoding = "UTF-8")

위의 코드를 설명하면 install.packages는 내장되지 않은 패키지를 설치하는 변수이다. 따라서 이외의 패키지를 설치할때 이용하자.(필자는 이미 설치했으므로 주석처리를 하였다.)

library 함수는 다운로드한 패키지를 언패킹 하는 함수이다. 즉, 압축되어있는 폴더를 열어서 사용한다는 느낌이라고 생각하면 된다. 문제는 xlsx 같은 경우 다른 패키지와 달리 java를 설치해야 library 함수가 돌아간다. 반드시 자바를 설치하고 library를 하자.

data\_file은 엑셀 파일을 저장하고자 하는 변수이름이다. 변수이름은 자신이 알기 쉽게 지정하면 된다.(필자는 data\_file로 저장을 하였음.)

read.xlsx는 엑셀파일을 읽어주는 함수이다.()안에 있는 것은 필수옵션과 선택옵션이 있다. 일단은 read.xlsx을 사용할때 주로 사용하는 옵션 위주로 적어보았다.

()안에 적어야될 것중 반드시 해야하는건 엑셀파일이 있는 pc의 경로를 입력해줘야한다. 위에 코드를 보면 "“안에 경로를 입력해야된다. 해당파일이 있는 경로를 찾는 방법은 <Files,Plots,Packages,Help,Viewer> 창을 보면 Files 창에 들어가서 자신이 있는 엑셀파일이 있는 폴더를 들어가서 More > Set as Working Directory를 눌러주면 console창에 파일경로가 r에 저장된다. 그 경로를 복사해서”"안에 넣은뒤, 파일명과 파일종류를 적어주면 된다.(필자의 경우 연습자료.xlsx이다.)

()옵션 중 1은 엑셀파일의 sheet 위치를 지정하는 것이다. 보통 sheet 1개일시, 1만 적어준다. ()옵션 중 colIndex와 rowIndex는 엑셀파일 데이터 중 불러올 행과 열을 입력하는 곳이다. row - 행을, col은 열을 지정해주는 것으로, 필자의 경우 열의 경우 1 ~ 9(옵션 입력시 1 : 9), 행의 경우 1 ~ 125이다. (숫자 : 숫자의 의미는 숫자 ~ 숫자이다.)

header는 맨 첫 행렬을 데이터로서 볼것인지 아닌지 지정하는 것이다. 데이터가 아닌 변수일 시 T, 아닐시에는 F를 입력하면 된다.

2번째로 중요한 것은 R의 경우 엑셀파일 데이터 중에 한글로 되어 있으면 읽지를 못하므로 변환을 해줘야한다. 따라서 encoding이란 함수를 통해 한글을 읽어줄 수 있게 해야한다. encoding = “UTF-8”을 입력해서 R이 엑셀을 읽을 수 있게 해줘야한다.

이제 자신들이 원하는 엑셀파일을 스스로 지정해서 불러들이도록 한다. 만약 데이터가 없을시 R내부에 저장되어 있는 iris 데이터를 이용해 보자.

……………………………………………………………………………………………………………………………………………………………………………

데이터를 불러들였으면 이제 데이터 가공을 해보자. 데이터 가공의 경우 불러들이기 전에 엑셀에서 결측값을 미리 처리하던가 아니면 R을 이용해서 결측값을 처리해주면 된다.

iris 데이터에 결측값이 있는지 없는지부터 확인하자.

is.na(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## [1,] FALSE FALSE FALSE FALSE FALSE  
## [2,] FALSE FALSE FALSE FALSE FALSE  
## [3,] FALSE FALSE FALSE FALSE FALSE  
## [4,] FALSE FALSE FALSE FALSE FALSE  
## [5,] FALSE FALSE FALSE FALSE FALSE  
## [6,] FALSE FALSE FALSE FALSE FALSE  
## [7,] FALSE FALSE FALSE FALSE FALSE  
## [8,] FALSE FALSE FALSE FALSE FALSE  
## [9,] FALSE FALSE FALSE FALSE FALSE  
## [10,] FALSE FALSE FALSE FALSE FALSE  
## [11,] FALSE FALSE FALSE FALSE FALSE  
## [12,] FALSE FALSE FALSE FALSE FALSE  
## [13,] FALSE FALSE FALSE FALSE FALSE  
## [14,] FALSE FALSE FALSE FALSE FALSE  
## [15,] FALSE FALSE FALSE FALSE FALSE  
## [16,] FALSE FALSE FALSE FALSE FALSE  
## [17,] FALSE FALSE FALSE FALSE FALSE  
## [18,] FALSE FALSE FALSE FALSE FALSE  
## [19,] FALSE FALSE FALSE FALSE FALSE  
## [20,] FALSE FALSE FALSE FALSE FALSE  
## [21,] FALSE FALSE FALSE FALSE FALSE  
## [22,] FALSE FALSE FALSE FALSE FALSE  
## [23,] FALSE FALSE FALSE FALSE FALSE  
## [24,] FALSE FALSE FALSE FALSE FALSE  
## [25,] FALSE FALSE FALSE FALSE FALSE  
## [26,] FALSE FALSE FALSE FALSE FALSE  
## [27,] FALSE FALSE FALSE FALSE FALSE  
## [28,] FALSE FALSE FALSE FALSE FALSE  
## [29,] FALSE FALSE FALSE FALSE FALSE  
## [30,] FALSE FALSE FALSE FALSE FALSE  
## [31,] FALSE FALSE FALSE FALSE FALSE  
## [32,] FALSE FALSE FALSE FALSE FALSE  
## [33,] FALSE FALSE FALSE FALSE FALSE  
## [34,] FALSE FALSE FALSE FALSE FALSE  
## [35,] FALSE FALSE FALSE FALSE FALSE  
## [36,] FALSE FALSE FALSE FALSE FALSE  
## [37,] FALSE FALSE FALSE FALSE FALSE  
## [38,] FALSE FALSE FALSE FALSE FALSE  
## [39,] FALSE FALSE FALSE FALSE FALSE  
## [40,] FALSE FALSE FALSE FALSE FALSE  
## [41,] FALSE FALSE FALSE FALSE FALSE  
## [42,] FALSE FALSE FALSE FALSE FALSE  
## [43,] FALSE FALSE FALSE FALSE FALSE  
## [44,] FALSE FALSE FALSE FALSE FALSE  
## [45,] FALSE FALSE FALSE FALSE FALSE  
## [46,] FALSE FALSE FALSE FALSE FALSE  
## [47,] FALSE FALSE FALSE FALSE FALSE  
## [48,] FALSE FALSE FALSE FALSE FALSE  
## [49,] FALSE FALSE FALSE FALSE FALSE  
## [50,] FALSE FALSE FALSE FALSE FALSE  
## [51,] FALSE FALSE FALSE FALSE FALSE  
## [52,] FALSE FALSE FALSE FALSE FALSE  
## [53,] FALSE FALSE FALSE FALSE FALSE  
## [54,] FALSE FALSE FALSE FALSE FALSE  
## [55,] FALSE FALSE FALSE FALSE FALSE  
## [56,] FALSE FALSE FALSE FALSE FALSE  
## [57,] FALSE FALSE FALSE FALSE FALSE  
## [58,] FALSE FALSE FALSE FALSE FALSE  
## [59,] FALSE FALSE FALSE FALSE FALSE  
## [60,] FALSE FALSE FALSE FALSE FALSE  
## [61,] FALSE FALSE FALSE FALSE FALSE  
## [62,] FALSE FALSE FALSE FALSE FALSE  
## [63,] FALSE FALSE FALSE FALSE FALSE  
## [64,] FALSE FALSE FALSE FALSE FALSE  
## [65,] FALSE FALSE FALSE FALSE FALSE  
## [66,] FALSE FALSE FALSE FALSE FALSE  
## [67,] FALSE FALSE FALSE FALSE FALSE  
## [68,] FALSE FALSE FALSE FALSE FALSE  
## [69,] FALSE FALSE FALSE FALSE FALSE  
## [70,] FALSE FALSE FALSE FALSE FALSE  
## [71,] FALSE FALSE FALSE FALSE FALSE  
## [72,] FALSE FALSE FALSE FALSE FALSE  
## [73,] FALSE FALSE FALSE FALSE FALSE  
## [74,] FALSE FALSE FALSE FALSE FALSE  
## [75,] FALSE FALSE FALSE FALSE FALSE  
## [76,] FALSE FALSE FALSE FALSE FALSE  
## [77,] FALSE FALSE FALSE FALSE FALSE  
## [78,] FALSE FALSE FALSE FALSE FALSE  
## [79,] FALSE FALSE FALSE FALSE FALSE  
## [80,] FALSE FALSE FALSE FALSE FALSE  
## [81,] FALSE FALSE FALSE FALSE FALSE  
## [82,] FALSE FALSE FALSE FALSE FALSE  
## [83,] FALSE FALSE FALSE FALSE FALSE  
## [84,] FALSE FALSE FALSE FALSE FALSE  
## [85,] FALSE FALSE FALSE FALSE FALSE  
## [86,] FALSE FALSE FALSE FALSE FALSE  
## [87,] FALSE FALSE FALSE FALSE FALSE  
## [88,] FALSE FALSE FALSE FALSE FALSE  
## [89,] FALSE FALSE FALSE FALSE FALSE  
## [90,] FALSE FALSE FALSE FALSE FALSE  
## [91,] FALSE FALSE FALSE FALSE FALSE  
## [92,] FALSE FALSE FALSE FALSE FALSE  
## [93,] FALSE FALSE FALSE FALSE FALSE  
## [94,] FALSE FALSE FALSE FALSE FALSE  
## [95,] FALSE FALSE FALSE FALSE FALSE  
## [96,] FALSE FALSE FALSE FALSE FALSE  
## [97,] FALSE FALSE FALSE FALSE FALSE  
## [98,] FALSE FALSE FALSE FALSE FALSE  
## [99,] FALSE FALSE FALSE FALSE FALSE  
## [100,] FALSE FALSE FALSE FALSE FALSE  
## [101,] FALSE FALSE FALSE FALSE FALSE  
## [102,] FALSE FALSE FALSE FALSE FALSE  
## [103,] FALSE FALSE FALSE FALSE FALSE  
## [104,] FALSE FALSE FALSE FALSE FALSE  
## [105,] FALSE FALSE FALSE FALSE FALSE  
## [106,] FALSE FALSE FALSE FALSE FALSE  
## [107,] FALSE FALSE FALSE FALSE FALSE  
## [108,] FALSE FALSE FALSE FALSE FALSE  
## [109,] FALSE FALSE FALSE FALSE FALSE  
## [110,] FALSE FALSE FALSE FALSE FALSE  
## [111,] FALSE FALSE FALSE FALSE FALSE  
## [112,] FALSE FALSE FALSE FALSE FALSE  
## [113,] FALSE FALSE FALSE FALSE FALSE  
## [114,] FALSE FALSE FALSE FALSE FALSE  
## [115,] FALSE FALSE FALSE FALSE FALSE  
## [116,] FALSE FALSE FALSE FALSE FALSE  
## [117,] FALSE FALSE FALSE FALSE FALSE  
## [118,] FALSE FALSE FALSE FALSE FALSE  
## [119,] FALSE FALSE FALSE FALSE FALSE  
## [120,] FALSE FALSE FALSE FALSE FALSE  
## [121,] FALSE FALSE FALSE FALSE FALSE  
## [122,] FALSE FALSE FALSE FALSE FALSE  
## [123,] FALSE FALSE FALSE FALSE FALSE  
## [124,] FALSE FALSE FALSE FALSE FALSE  
## [125,] FALSE FALSE FALSE FALSE FALSE  
## [126,] FALSE FALSE FALSE FALSE FALSE  
## [127,] FALSE FALSE FALSE FALSE FALSE  
## [128,] FALSE FALSE FALSE FALSE FALSE  
## [129,] FALSE FALSE FALSE FALSE FALSE  
## [130,] FALSE FALSE FALSE FALSE FALSE  
## [131,] FALSE FALSE FALSE FALSE FALSE  
## [132,] FALSE FALSE FALSE FALSE FALSE  
## [133,] FALSE FALSE FALSE FALSE FALSE  
## [134,] FALSE FALSE FALSE FALSE FALSE  
## [135,] FALSE FALSE FALSE FALSE FALSE  
## [136,] FALSE FALSE FALSE FALSE FALSE  
## [137,] FALSE FALSE FALSE FALSE FALSE  
## [138,] FALSE FALSE FALSE FALSE FALSE  
## [139,] FALSE FALSE FALSE FALSE FALSE  
## [140,] FALSE FALSE FALSE FALSE FALSE  
## [141,] FALSE FALSE FALSE FALSE FALSE  
## [142,] FALSE FALSE FALSE FALSE FALSE  
## [143,] FALSE FALSE FALSE FALSE FALSE  
## [144,] FALSE FALSE FALSE FALSE FALSE  
## [145,] FALSE FALSE FALSE FALSE FALSE  
## [146,] FALSE FALSE FALSE FALSE FALSE  
## [147,] FALSE FALSE FALSE FALSE FALSE  
## [148,] FALSE FALSE FALSE FALSE FALSE  
## [149,] FALSE FALSE FALSE FALSE FALSE  
## [150,] FALSE FALSE FALSE FALSE FALSE

table(is.na(iris))

##   
## FALSE   
## 750

is.na()는 결측값이 있는지 없는지를 알려주는 함수이다. ()안에는 데이터가 저장된 변수를 입력하면 된다.

(모든 함수의 ()안에는 데이터와 그 함수의 옵션 명령을 넣어주면 된다.)

위에 두 구문의 차이는 is.na()을 했을때는 각 n by n에 따라 각각의 위치에 NA이 없을 경우 FALSE 있으면 TURE를 표기한다.

table()을 쓰게되면 FALSE, TRUE를 테이블 형식으로 보여준다. 따라서 후자를 선택하는 것이 더 편하다.

table(is.na(data\_file))

##   
## FALSE TRUE   
## 1115 1

다른 방법으로는 complete.cases() 함수를 이용한다.

complete.cases(data\_file)

## [1] FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [12] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [23] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [34] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [45] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [56] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [67] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [78] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [89] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [100] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [111] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [122] TRUE TRUE TRUE

위를 보면 NA값이 있는 행에 대하여 T/F로 나타내 주고 있다. 이를 통해 어디에 NA값이 있는지를 확인할 수 있다.

결과를 보면 NA값이 1개 존재한다. 따라서 이 결측값을 처리해줘야 데이터 분석 과정으로 들어갈 수 있다.

결측값을 처리하는 경우는 결측값이 들어간 행을 빼줘도 되거나, 자료의 형태에 따라 대체되는 값이 달라진다.

연속형일 경우 보통 결측값을 제거한 평균을 결측값을 넣어주거나, 최빈값으로 대체한다.

범주형일 경우 결측값을 의미하는 범주를 만들어서 그 값으로 대체한다.

#case 1. 행,열 선택/없애기  
data\_file[1:5,2:4]

## 성별 연령 직업  
## 1 2 18 0  
## 2 1 18 0  
## 3 2 19 1  
## 4 1 19 0  
## 5 1 20 0

data\_file[2:3, ]

## ID 성별 연령 직업 통신사 가입년수 사용료 부족한서비스 남는서비스  
## 2 A2 1 18 0 1 3 1.5 3 1  
## 3 A3 2 19 1 3 8 10.0 5 4

data\_file[-1:-100, 2:3]

## 성별 연령  
## 101 1 50  
## 102 1 0  
## 103 1 0  
## 104 1 0  
## 105 1 0  
## 106 2 0  
## 107 1 0  
## 108 1 0  
## 109 1 0  
## 110 1 0  
## 111 1 20  
## 112 1 22  
## 113 1 23  
## 114 2 22  
## 115 2 22  
## 116 2 20  
## 117 2 20  
## 118 1 23  
## 119 2 20  
## 120 2 22  
## 121 2 24  
## 122 2 22  
## 123 2 20  
## 124 1 0

첫번째 코드를 보면 ‘data가 들어간 변수이름[]’ 형태를 통해 데이터를 뽑아낼 수 있다. []안에 보면 [row,col]의 형태이다. row위치에 데이터를 보고싶은 위치의 값을 넣으면 행의 그 위치를 뽑아낸다. 따라서 data\_file[1:5,2:4]는 1~5행, 2~4열을 뽑아낸다.

두번째 코드에서 [x,y]에서 y값에 빈칸이 들어가 있다. 이는 오류가 아니라 열 전체를 뽑으라는 의미이다. 따라서 어느 행,열 전체를 보고 싶으면 그 칸에는 빈칸을 입력하면 된다.

3번째 코드에서 [x,y]에 -값이 들어갈 경우, 그 값을 빼고 나머지의 데이터를 수출하라는 의미이다. 위에서 보면 1~100행을 뺀 행, 2~3열을 호출하라는 의미이다.

이를 통해서 결측값이 들어있는 값을 제거해보자.

#결측값 위치를 찾기  
data\_file[!complete.cases(data\_file),]

## ID 성별 연령 직업 통신사 가입년수 사용료 부족한서비스 남는서비스  
## 1 A1 2 18 0 NA 3 5.5 4 3

#결측값 위치 제거  
#case 1.  
data\_file2 <- data\_file[complete.cases(data\_file),]  
data\_file2

## ID 성별 연령 직업 통신사 가입년수 사용료 부족한서비스 남는서비스  
## 2 A2 1 18 0 1 3 1.5 3 1  
## 3 A3 2 19 1 3 8 10.0 5 4  
## 4 A4 1 19 0 1 8 4.5 5 3  
## 5 A5 1 20 0 1 3 3.5 3 1  
## 6 A6 1 20 0 2 3 7.5 3 2  
## 7 A7 1 20 0 2 3 1.5 3 2  
## 8 A8 1 20 0 1 3 3.5 3 2  
## 9 A9 1 20 0 2 3 6.5 1 2  
## 10 A10 1 20 0 1 3 5.5 3 1  
## 11 A11 1 20 0 2 3 4.5 3 1  
## 12 A12 1 20 0 2 3 3.5 3 2  
## 13 A13 2 20 0 3 8 3.5 3 4  
## 14 A14 1 20 0 1 8 4.5 3 1  
## 15 A15 2 20 0 1 3 3.5 3 4  
## 16 A16 1 20 0 2 10 7.5 4 2  
## 17 A17 1 20 0 2 3 6.5 4 2  
## 18 A18 1 20 0 1 8 4.5 3 4  
## 19 A19 2 20 0 2 8 8.5 3 2  
## 20 A20 1 20 0 2 3 6.5 4 3  
## 21 A21 2 20 0 2 10 5.5 3 4  
## 22 A22 1 20 0 2 8 4.5 3 1  
## 23 A23 1 20 0 2 8 4.5 3 1  
## 24 A24 2 20 0 3 3 8.5 5 3  
## 25 A25 2 20 0 1 3 8.5 3 5  
## 26 A26 2 20 1 1 8 7.5 3 3  
## 27 A27 1 21 0 1 8 5.5 3 1  
## 28 A28 1 21 0 1 3 4.5 3 1  
## 29 A29 2 21 0 3 3 7.5 5 1  
## 30 A30 1 21 0 1 10 6.5 5 1  
## 31 A31 1 21 0 1 8 5.5 3 1  
## 32 A32 1 21 0 1 10 4.5 3 1  
## 33 A33 1 21 0 1 10 6.5 3 2  
## 34 A34 1 21 0 2 10 3.5 3 1  
## 35 A35 1 21 0 1 8 5.5 3 2  
## 36 A36 1 21 0 1 10 3.5 3 2  
## 37 A37 1 21 0 1 10 3.5 5 5  
## 38 A38 1 21 0 1 8 3.5 3 1  
## 39 A39 2 21 0 2 3 3.5 3 2  
## 40 A40 2 21 0 1 8 5.5 3 2  
## 41 A41 2 21 0 2 3 3.5 3 2  
## 42 A42 2 21 0 2 8 4.5 3 2  
## 43 A43 1 21 0 3 3 6.5 5 1  
## 44 A44 2 21 0 2 3 5.5 3 1  
## 45 A45 2 21 0 1 3 5.5 3 5  
## 46 A46 1 21 0 2 3 3.5 3 1  
## 47 A47 2 21 0 1 3 3.5 3 1  
## 48 A48 1 22 0 1 8 7.5 3 3  
## 49 A49 2 22 0 2 3 5.5 3 2  
## 50 A50 1 22 0 1 3 6.5 5 5  
## 51 A51 1 22 0 3 3 6.5 1 4  
## 52 A52 1 22 0 1 3 6.5 5 1  
## 53 A53 2 22 0 1 8 5.5 3 2  
## 54 A54 2 22 1 3 3 7.5 5 2  
## 55 A55 1 22 0 3 3 6.5 3 5  
## 56 A56 2 22 0 1 3 10.0 5 2  
## 57 A57 2 22 0 1 8 4.5 5 2  
## 58 A58 2 22 0 3 8 6.5 5 4  
## 59 A59 1 22 0 1 8 8.5 3 5  
## 60 A60 1 23 0 2 8 3.5 3 2  
## 61 A61 1 23 0 3 3 6.5 3 2  
## 62 A62 1 23 0 2 3 3.5 3 2  
## 63 A63 1 23 0 2 3 4.5 3 1  
## 64 A64 1 23 0 2 3 7.5 5 5  
## 65 A65 1 23 0 2 3 4.5 3 2  
## 66 A66 2 24 0 3 8 3.5 5 2  
## 67 A67 1 24 0 3 3 3.5 3 1  
## 68 A68 2 24 0 1 3 5.5 3 4  
## 69 A69 1 24 0 1 8 4.5 3 2  
## 70 A70 2 24 0 2 10 3.5 3 2  
## 71 A71 2 25 1 1 3 6.5 3 2  
## 72 A72 2 25 0 1 10 7.5 4 4  
## 73 A73 1 25 1 1 8 7.5 5 5  
## 74 A74 1 25 0 2 10 6.5 5 1  
## 75 A75 2 25 1 3 3 6.5 5 1  
## 76 A76 2 26 1 2 3 3.5 3 2  
## 77 A77 2 26 0 2 10 6.5 4 1  
## 78 A78 1 26 0 2 8 7.5 4 1  
## 79 A79 1 26 1 2 8 3.5 3 1  
## 80 A80 2 26 1 3 3 4.5 3 2  
## 81 A81 1 26 1 1 3 6.5 3 3  
## 82 A82 2 26 1 3 3 6.5 5 2  
## 83 A83 1 26 1 1 3 6.5 3 2  
## 84 A84 1 26 0 1 8 5.5 5 5  
## 85 A85 1 27 1 2 3 8.5 3 4  
## 86 A86 2 27 1 3 3 7.5 4 1  
## 87 A87 2 27 1 3 3 6.5 3 4  
## 88 A88 1 29 1 1 3 5.5 3 1  
## 89 A89 2 30 1 1 3 7.5 3 1  
## 90 A90 1 30 0 2 3 5.5 5 5  
## 91 A91 2 31 1 1 10 9.5 3 1  
## 92 A92 2 34 1 1 8 5.5 1 4  
## 93 A93 1 36 1 1 3 3.5 3 4  
## 94 A94 1 36 1 1 10 5.5 4 3  
## 95 A95 1 36 1 1 3 6.5 3 2  
## 96 A96 1 39 1 2 8 3.5 1 3  
## 97 A97 1 40 1 1 10 3.5 1 1  
## 98 A98 1 40 1 1 10 8.5 3 2  
## 99 A99 1 49 1 1 8 3.5 3 1  
## 100 A100 1 49 1 1 3 6.5 5 3  
## 101 A101 1 50 1 1 8 6.5 3 1  
## 102 A102 1 0 0 1 3 5.5 5 5  
## 103 A103 1 0 0 1 8 7.5 5 1  
## 104 A104 1 0 1 1 10 6.5 3 2  
## 105 A105 1 0 0 2 8 6.5 5 5  
## 106 A106 2 0 0 1 3 5.5 3 4  
## 107 A107 1 0 1 2 8 6.5 5 1  
## 108 A108 1 0 0 2 3 6.5 4 4  
## 109 A109 1 0 1 3 8 7.5 3 4  
## 110 A110 1 0 0 3 3 3.5 3 1  
## 111 A111 1 20 0 1 3 5.5 3 2  
## 112 A112 1 22 0 2 3 4.5 3 4  
## 113 A113 1 23 0 2 3 3.5 3 2  
## 114 A114 2 22 0 2 3 6.5 3 2  
## 115 A115 2 22 0 1 8 3.5 3 4  
## 116 A116 2 20 0 2 3 6.5 5 5  
## 117 A117 2 20 0 1 3 4.5 3 2  
## 118 A118 1 23 0 3 3 5.5 5 2  
## 119 A119 2 20 0 2 3 4.5 3 1  
## 120 A120 2 22 0 1 8 3.5 3 2  
## 121 A121 2 24 0 2 3 1.5 3 2  
## 122 A122 2 22 0 1 3 3.5 3 2  
## 123 A123 2 20 0 3 3 5.5 3 4  
## 124 A124 1 0 0 1 10 6.5 3 1

#case 2.  
data\_file[-1,]

## ID 성별 연령 직업 통신사 가입년수 사용료 부족한서비스 남는서비스  
## 2 A2 1 18 0 1 3 1.5 3 1  
## 3 A3 2 19 1 3 8 10.0 5 4  
## 4 A4 1 19 0 1 8 4.5 5 3  
## 5 A5 1 20 0 1 3 3.5 3 1  
## 6 A6 1 20 0 2 3 7.5 3 2  
## 7 A7 1 20 0 2 3 1.5 3 2  
## 8 A8 1 20 0 1 3 3.5 3 2  
## 9 A9 1 20 0 2 3 6.5 1 2  
## 10 A10 1 20 0 1 3 5.5 3 1  
## 11 A11 1 20 0 2 3 4.5 3 1  
## 12 A12 1 20 0 2 3 3.5 3 2  
## 13 A13 2 20 0 3 8 3.5 3 4  
## 14 A14 1 20 0 1 8 4.5 3 1  
## 15 A15 2 20 0 1 3 3.5 3 4  
## 16 A16 1 20 0 2 10 7.5 4 2  
## 17 A17 1 20 0 2 3 6.5 4 2  
## 18 A18 1 20 0 1 8 4.5 3 4  
## 19 A19 2 20 0 2 8 8.5 3 2  
## 20 A20 1 20 0 2 3 6.5 4 3  
## 21 A21 2 20 0 2 10 5.5 3 4  
## 22 A22 1 20 0 2 8 4.5 3 1  
## 23 A23 1 20 0 2 8 4.5 3 1  
## 24 A24 2 20 0 3 3 8.5 5 3  
## 25 A25 2 20 0 1 3 8.5 3 5  
## 26 A26 2 20 1 1 8 7.5 3 3  
## 27 A27 1 21 0 1 8 5.5 3 1  
## 28 A28 1 21 0 1 3 4.5 3 1  
## 29 A29 2 21 0 3 3 7.5 5 1  
## 30 A30 1 21 0 1 10 6.5 5 1  
## 31 A31 1 21 0 1 8 5.5 3 1  
## 32 A32 1 21 0 1 10 4.5 3 1  
## 33 A33 1 21 0 1 10 6.5 3 2  
## 34 A34 1 21 0 2 10 3.5 3 1  
## 35 A35 1 21 0 1 8 5.5 3 2  
## 36 A36 1 21 0 1 10 3.5 3 2  
## 37 A37 1 21 0 1 10 3.5 5 5  
## 38 A38 1 21 0 1 8 3.5 3 1  
## 39 A39 2 21 0 2 3 3.5 3 2  
## 40 A40 2 21 0 1 8 5.5 3 2  
## 41 A41 2 21 0 2 3 3.5 3 2  
## 42 A42 2 21 0 2 8 4.5 3 2  
## 43 A43 1 21 0 3 3 6.5 5 1  
## 44 A44 2 21 0 2 3 5.5 3 1  
## 45 A45 2 21 0 1 3 5.5 3 5  
## 46 A46 1 21 0 2 3 3.5 3 1  
## 47 A47 2 21 0 1 3 3.5 3 1  
## 48 A48 1 22 0 1 8 7.5 3 3  
## 49 A49 2 22 0 2 3 5.5 3 2  
## 50 A50 1 22 0 1 3 6.5 5 5  
## 51 A51 1 22 0 3 3 6.5 1 4  
## 52 A52 1 22 0 1 3 6.5 5 1  
## 53 A53 2 22 0 1 8 5.5 3 2  
## 54 A54 2 22 1 3 3 7.5 5 2  
## 55 A55 1 22 0 3 3 6.5 3 5  
## 56 A56 2 22 0 1 3 10.0 5 2  
## 57 A57 2 22 0 1 8 4.5 5 2  
## 58 A58 2 22 0 3 8 6.5 5 4  
## 59 A59 1 22 0 1 8 8.5 3 5  
## 60 A60 1 23 0 2 8 3.5 3 2  
## 61 A61 1 23 0 3 3 6.5 3 2  
## 62 A62 1 23 0 2 3 3.5 3 2  
## 63 A63 1 23 0 2 3 4.5 3 1  
## 64 A64 1 23 0 2 3 7.5 5 5  
## 65 A65 1 23 0 2 3 4.5 3 2  
## 66 A66 2 24 0 3 8 3.5 5 2  
## 67 A67 1 24 0 3 3 3.5 3 1  
## 68 A68 2 24 0 1 3 5.5 3 4  
## 69 A69 1 24 0 1 8 4.5 3 2  
## 70 A70 2 24 0 2 10 3.5 3 2  
## 71 A71 2 25 1 1 3 6.5 3 2  
## 72 A72 2 25 0 1 10 7.5 4 4  
## 73 A73 1 25 1 1 8 7.5 5 5  
## 74 A74 1 25 0 2 10 6.5 5 1  
## 75 A75 2 25 1 3 3 6.5 5 1  
## 76 A76 2 26 1 2 3 3.5 3 2  
## 77 A77 2 26 0 2 10 6.5 4 1  
## 78 A78 1 26 0 2 8 7.5 4 1  
## 79 A79 1 26 1 2 8 3.5 3 1  
## 80 A80 2 26 1 3 3 4.5 3 2  
## 81 A81 1 26 1 1 3 6.5 3 3  
## 82 A82 2 26 1 3 3 6.5 5 2  
## 83 A83 1 26 1 1 3 6.5 3 2  
## 84 A84 1 26 0 1 8 5.5 5 5  
## 85 A85 1 27 1 2 3 8.5 3 4  
## 86 A86 2 27 1 3 3 7.5 4 1  
## 87 A87 2 27 1 3 3 6.5 3 4  
## 88 A88 1 29 1 1 3 5.5 3 1  
## 89 A89 2 30 1 1 3 7.5 3 1  
## 90 A90 1 30 0 2 3 5.5 5 5  
## 91 A91 2 31 1 1 10 9.5 3 1  
## 92 A92 2 34 1 1 8 5.5 1 4  
## 93 A93 1 36 1 1 3 3.5 3 4  
## 94 A94 1 36 1 1 10 5.5 4 3  
## 95 A95 1 36 1 1 3 6.5 3 2  
## 96 A96 1 39 1 2 8 3.5 1 3  
## 97 A97 1 40 1 1 10 3.5 1 1  
## 98 A98 1 40 1 1 10 8.5 3 2  
## 99 A99 1 49 1 1 8 3.5 3 1  
## 100 A100 1 49 1 1 3 6.5 5 3  
## 101 A101 1 50 1 1 8 6.5 3 1  
## 102 A102 1 0 0 1 3 5.5 5 5  
## 103 A103 1 0 0 1 8 7.5 5 1  
## 104 A104 1 0 1 1 10 6.5 3 2  
## 105 A105 1 0 0 2 8 6.5 5 5  
## 106 A106 2 0 0 1 3 5.5 3 4  
## 107 A107 1 0 1 2 8 6.5 5 1  
## 108 A108 1 0 0 2 3 6.5 4 4  
## 109 A109 1 0 1 3 8 7.5 3 4  
## 110 A110 1 0 0 3 3 3.5 3 1  
## 111 A111 1 20 0 1 3 5.5 3 2  
## 112 A112 1 22 0 2 3 4.5 3 4  
## 113 A113 1 23 0 2 3 3.5 3 2  
## 114 A114 2 22 0 2 3 6.5 3 2  
## 115 A115 2 22 0 1 8 3.5 3 4  
## 116 A116 2 20 0 2 3 6.5 5 5  
## 117 A117 2 20 0 1 3 4.5 3 2  
## 118 A118 1 23 0 3 3 5.5 5 2  
## 119 A119 2 20 0 2 3 4.5 3 1  
## 120 A120 2 22 0 1 8 3.5 3 2  
## 121 A121 2 24 0 2 3 1.5 3 2  
## 122 A122 2 22 0 1 3 3.5 3 2  
## 123 A123 2 20 0 3 3 5.5 3 4  
## 124 A124 1 0 0 1 10 6.5 3 1

data\_file[!complete.cases(data\_file),] 에서 ‘!’는 나오는 결과값의 반대를 출력하라는 의미이다.’!’가 없으면 그냥 코드에 대한 값들을 출력하고 “!”이 있으면 코드에 대한 값들을 제외하고 전부 출력이 된다.

두 개의 실행값을 일치한다. 다만 case2의 경우에는 NA값이 어디있는지 아는 경우에 하는 경우이고, case1의 경우에는 NA값이 어디있는지 모르는 경우에 사용하면 좋다.

#결측값을 다른 값으로 바꿀 때  
data\_file3 <- data\_file  
  
data\_file3[is.na(data\_file3)] <- 9  
data\_file3

## ID 성별 연령 직업 통신사 가입년수 사용료 부족한서비스 남는서비스  
## 1 A1 2 18 0 9 3 5.5 4 3  
## 2 A2 1 18 0 1 3 1.5 3 1  
## 3 A3 2 19 1 3 8 10.0 5 4  
## 4 A4 1 19 0 1 8 4.5 5 3  
## 5 A5 1 20 0 1 3 3.5 3 1  
## 6 A6 1 20 0 2 3 7.5 3 2  
## 7 A7 1 20 0 2 3 1.5 3 2  
## 8 A8 1 20 0 1 3 3.5 3 2  
## 9 A9 1 20 0 2 3 6.5 1 2  
## 10 A10 1 20 0 1 3 5.5 3 1  
## 11 A11 1 20 0 2 3 4.5 3 1  
## 12 A12 1 20 0 2 3 3.5 3 2  
## 13 A13 2 20 0 3 8 3.5 3 4  
## 14 A14 1 20 0 1 8 4.5 3 1  
## 15 A15 2 20 0 1 3 3.5 3 4  
## 16 A16 1 20 0 2 10 7.5 4 2  
## 17 A17 1 20 0 2 3 6.5 4 2  
## 18 A18 1 20 0 1 8 4.5 3 4  
## 19 A19 2 20 0 2 8 8.5 3 2  
## 20 A20 1 20 0 2 3 6.5 4 3  
## 21 A21 2 20 0 2 10 5.5 3 4  
## 22 A22 1 20 0 2 8 4.5 3 1  
## 23 A23 1 20 0 2 8 4.5 3 1  
## 24 A24 2 20 0 3 3 8.5 5 3  
## 25 A25 2 20 0 1 3 8.5 3 5  
## 26 A26 2 20 1 1 8 7.5 3 3  
## 27 A27 1 21 0 1 8 5.5 3 1  
## 28 A28 1 21 0 1 3 4.5 3 1  
## 29 A29 2 21 0 3 3 7.5 5 1  
## 30 A30 1 21 0 1 10 6.5 5 1  
## 31 A31 1 21 0 1 8 5.5 3 1  
## 32 A32 1 21 0 1 10 4.5 3 1  
## 33 A33 1 21 0 1 10 6.5 3 2  
## 34 A34 1 21 0 2 10 3.5 3 1  
## 35 A35 1 21 0 1 8 5.5 3 2  
## 36 A36 1 21 0 1 10 3.5 3 2  
## 37 A37 1 21 0 1 10 3.5 5 5  
## 38 A38 1 21 0 1 8 3.5 3 1  
## 39 A39 2 21 0 2 3 3.5 3 2  
## 40 A40 2 21 0 1 8 5.5 3 2  
## 41 A41 2 21 0 2 3 3.5 3 2  
## 42 A42 2 21 0 2 8 4.5 3 2  
## 43 A43 1 21 0 3 3 6.5 5 1  
## 44 A44 2 21 0 2 3 5.5 3 1  
## 45 A45 2 21 0 1 3 5.5 3 5  
## 46 A46 1 21 0 2 3 3.5 3 1  
## 47 A47 2 21 0 1 3 3.5 3 1  
## 48 A48 1 22 0 1 8 7.5 3 3  
## 49 A49 2 22 0 2 3 5.5 3 2  
## 50 A50 1 22 0 1 3 6.5 5 5  
## 51 A51 1 22 0 3 3 6.5 1 4  
## 52 A52 1 22 0 1 3 6.5 5 1  
## 53 A53 2 22 0 1 8 5.5 3 2  
## 54 A54 2 22 1 3 3 7.5 5 2  
## 55 A55 1 22 0 3 3 6.5 3 5  
## 56 A56 2 22 0 1 3 10.0 5 2  
## 57 A57 2 22 0 1 8 4.5 5 2  
## 58 A58 2 22 0 3 8 6.5 5 4  
## 59 A59 1 22 0 1 8 8.5 3 5  
## 60 A60 1 23 0 2 8 3.5 3 2  
## 61 A61 1 23 0 3 3 6.5 3 2  
## 62 A62 1 23 0 2 3 3.5 3 2  
## 63 A63 1 23 0 2 3 4.5 3 1  
## 64 A64 1 23 0 2 3 7.5 5 5  
## 65 A65 1 23 0 2 3 4.5 3 2  
## 66 A66 2 24 0 3 8 3.5 5 2  
## 67 A67 1 24 0 3 3 3.5 3 1  
## 68 A68 2 24 0 1 3 5.5 3 4  
## 69 A69 1 24 0 1 8 4.5 3 2  
## 70 A70 2 24 0 2 10 3.5 3 2  
## 71 A71 2 25 1 1 3 6.5 3 2  
## 72 A72 2 25 0 1 10 7.5 4 4  
## 73 A73 1 25 1 1 8 7.5 5 5  
## 74 A74 1 25 0 2 10 6.5 5 1  
## 75 A75 2 25 1 3 3 6.5 5 1  
## 76 A76 2 26 1 2 3 3.5 3 2  
## 77 A77 2 26 0 2 10 6.5 4 1  
## 78 A78 1 26 0 2 8 7.5 4 1  
## 79 A79 1 26 1 2 8 3.5 3 1  
## 80 A80 2 26 1 3 3 4.5 3 2  
## 81 A81 1 26 1 1 3 6.5 3 3  
## 82 A82 2 26 1 3 3 6.5 5 2  
## 83 A83 1 26 1 1 3 6.5 3 2  
## 84 A84 1 26 0 1 8 5.5 5 5  
## 85 A85 1 27 1 2 3 8.5 3 4  
## 86 A86 2 27 1 3 3 7.5 4 1  
## 87 A87 2 27 1 3 3 6.5 3 4  
## 88 A88 1 29 1 1 3 5.5 3 1  
## 89 A89 2 30 1 1 3 7.5 3 1  
## 90 A90 1 30 0 2 3 5.5 5 5  
## 91 A91 2 31 1 1 10 9.5 3 1  
## 92 A92 2 34 1 1 8 5.5 1 4  
## 93 A93 1 36 1 1 3 3.5 3 4  
## 94 A94 1 36 1 1 10 5.5 4 3  
## 95 A95 1 36 1 1 3 6.5 3 2  
## 96 A96 1 39 1 2 8 3.5 1 3  
## 97 A97 1 40 1 1 10 3.5 1 1  
## 98 A98 1 40 1 1 10 8.5 3 2  
## 99 A99 1 49 1 1 8 3.5 3 1  
## 100 A100 1 49 1 1 3 6.5 5 3  
## 101 A101 1 50 1 1 8 6.5 3 1  
## 102 A102 1 0 0 1 3 5.5 5 5  
## 103 A103 1 0 0 1 8 7.5 5 1  
## 104 A104 1 0 1 1 10 6.5 3 2  
## 105 A105 1 0 0 2 8 6.5 5 5  
## 106 A106 2 0 0 1 3 5.5 3 4  
## 107 A107 1 0 1 2 8 6.5 5 1  
## 108 A108 1 0 0 2 3 6.5 4 4  
## 109 A109 1 0 1 3 8 7.5 3 4  
## 110 A110 1 0 0 3 3 3.5 3 1  
## 111 A111 1 20 0 1 3 5.5 3 2  
## 112 A112 1 22 0 2 3 4.5 3 4  
## 113 A113 1 23 0 2 3 3.5 3 2  
## 114 A114 2 22 0 2 3 6.5 3 2  
## 115 A115 2 22 0 1 8 3.5 3 4  
## 116 A116 2 20 0 2 3 6.5 5 5  
## 117 A117 2 20 0 1 3 4.5 3 2  
## 118 A118 1 23 0 3 3 5.5 5 2  
## 119 A119 2 20 0 2 3 4.5 3 1  
## 120 A120 2 22 0 1 8 3.5 3 2  
## 121 A121 2 24 0 2 3 1.5 3 2  
## 122 A122 2 22 0 1 3 3.5 3 2  
## 123 A123 2 20 0 3 3 5.5 3 4  
## 124 A124 1 0 0 1 10 6.5 3 1

위와 같이 이용하면 na값이 9로 바뀐다. 9대신 원하는 값으로 바꾸면 된다. 필자의 경우 NA값이 1개 밖에 없으므로 결측값을 없애는 방향으로 간다.

……………………………………………………………………………………………………………………………………………………………………………

결측값을 제거했으면 신경망 분석을 하기 위해서 검증집단과 테스트집단으로 나누어 보자.

(비율은 7:3으로)

먼저 iris 데이터를 보자.

idx <- sample(2,  
 nrow(iris),  
 replace = TRUE,  
 prob =c(0.7,0.3))  
  
  
iris\_file2\_train <- iris[idx==2,]  
iris\_file2\_test <- iris[idx==1,]  
iris\_file2\_test

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 7 4.6 3.4 1.4 0.3 setosa  
## 8 5.0 3.4 1.5 0.2 setosa  
## 9 4.4 2.9 1.4 0.2 setosa  
## 10 4.9 3.1 1.5 0.1 setosa  
## 13 4.8 3.0 1.4 0.1 setosa  
## 15 5.8 4.0 1.2 0.2 setosa  
## 16 5.7 4.4 1.5 0.4 setosa  
## 18 5.1 3.5 1.4 0.3 setosa  
## 19 5.7 3.8 1.7 0.3 setosa  
## 20 5.1 3.8 1.5 0.3 setosa  
## 21 5.4 3.4 1.7 0.2 setosa  
## 22 5.1 3.7 1.5 0.4 setosa  
## 23 4.6 3.6 1.0 0.2 setosa  
## 24 5.1 3.3 1.7 0.5 setosa  
## 27 5.0 3.4 1.6 0.4 setosa  
## 28 5.2 3.5 1.5 0.2 setosa  
## 29 5.2 3.4 1.4 0.2 setosa  
## 30 4.7 3.2 1.6 0.2 setosa  
## 31 4.8 3.1 1.6 0.2 setosa  
## 34 5.5 4.2 1.4 0.2 setosa  
## 35 4.9 3.1 1.5 0.2 setosa  
## 37 5.5 3.5 1.3 0.2 setosa  
## 39 4.4 3.0 1.3 0.2 setosa  
## 40 5.1 3.4 1.5 0.2 setosa  
## 42 4.5 2.3 1.3 0.3 setosa  
## 43 4.4 3.2 1.3 0.2 setosa  
## 44 5.0 3.5 1.6 0.6 setosa  
## 46 4.8 3.0 1.4 0.3 setosa  
## 47 5.1 3.8 1.6 0.2 setosa  
## 48 4.6 3.2 1.4 0.2 setosa  
## 50 5.0 3.3 1.4 0.2 setosa  
## 51 7.0 3.2 4.7 1.4 versicolor  
## 52 6.4 3.2 4.5 1.5 versicolor  
## 53 6.9 3.1 4.9 1.5 versicolor  
## 54 5.5 2.3 4.0 1.3 versicolor  
## 55 6.5 2.8 4.6 1.5 versicolor  
## 56 5.7 2.8 4.5 1.3 versicolor  
## 57 6.3 3.3 4.7 1.6 versicolor  
## 58 4.9 2.4 3.3 1.0 versicolor  
## 60 5.2 2.7 3.9 1.4 versicolor  
## 61 5.0 2.0 3.5 1.0 versicolor  
## 62 5.9 3.0 4.2 1.5 versicolor  
## 63 6.0 2.2 4.0 1.0 versicolor  
## 64 6.1 2.9 4.7 1.4 versicolor  
## 66 6.7 3.1 4.4 1.4 versicolor  
## 67 5.6 3.0 4.5 1.5 versicolor  
## 69 6.2 2.2 4.5 1.5 versicolor  
## 70 5.6 2.5 3.9 1.1 versicolor  
## 71 5.9 3.2 4.8 1.8 versicolor  
## 72 6.1 2.8 4.0 1.3 versicolor  
## 73 6.3 2.5 4.9 1.5 versicolor  
## 74 6.1 2.8 4.7 1.2 versicolor  
## 75 6.4 2.9 4.3 1.3 versicolor  
## 77 6.8 2.8 4.8 1.4 versicolor  
## 79 6.0 2.9 4.5 1.5 versicolor  
## 80 5.7 2.6 3.5 1.0 versicolor  
## 83 5.8 2.7 3.9 1.2 versicolor  
## 85 5.4 3.0 4.5 1.5 versicolor  
## 86 6.0 3.4 4.5 1.6 versicolor  
## 88 6.3 2.3 4.4 1.3 versicolor  
## 91 5.5 2.6 4.4 1.2 versicolor  
## 94 5.0 2.3 3.3 1.0 versicolor  
## 95 5.6 2.7 4.2 1.3 versicolor  
## 97 5.7 2.9 4.2 1.3 versicolor  
## 98 6.2 2.9 4.3 1.3 versicolor  
## 99 5.1 2.5 3.0 1.1 versicolor  
## 101 6.3 3.3 6.0 2.5 virginica  
## 102 5.8 2.7 5.1 1.9 virginica  
## 103 7.1 3.0 5.9 2.1 virginica  
## 104 6.3 2.9 5.6 1.8 virginica  
## 105 6.5 3.0 5.8 2.2 virginica  
## 107 4.9 2.5 4.5 1.7 virginica  
## 108 7.3 2.9 6.3 1.8 virginica  
## 109 6.7 2.5 5.8 1.8 virginica  
## 112 6.4 2.7 5.3 1.9 virginica  
## 113 6.8 3.0 5.5 2.1 virginica  
## 114 5.7 2.5 5.0 2.0 virginica  
## 115 5.8 2.8 5.1 2.4 virginica  
## 117 6.5 3.0 5.5 1.8 virginica  
## 118 7.7 3.8 6.7 2.2 virginica  
## 120 6.0 2.2 5.0 1.5 virginica  
## 121 6.9 3.2 5.7 2.3 virginica  
## 123 7.7 2.8 6.7 2.0 virginica  
## 124 6.3 2.7 4.9 1.8 virginica  
## 125 6.7 3.3 5.7 2.1 virginica  
## 126 7.2 3.2 6.0 1.8 virginica  
## 128 6.1 3.0 4.9 1.8 virginica  
## 129 6.4 2.8 5.6 2.1 virginica  
## 131 7.4 2.8 6.1 1.9 virginica  
## 132 7.9 3.8 6.4 2.0 virginica  
## 133 6.4 2.8 5.6 2.2 virginica  
## 134 6.3 2.8 5.1 1.5 virginica  
## 135 6.1 2.6 5.6 1.4 virginica  
## 136 7.7 3.0 6.1 2.3 virginica  
## 137 6.3 3.4 5.6 2.4 virginica  
## 138 6.4 3.1 5.5 1.8 virginica  
## 140 6.9 3.1 5.4 2.1 virginica  
## 144 6.8 3.2 5.9 2.3 virginica  
## 147 6.3 2.5 5.0 1.9 virginica  
## 148 6.5 3.0 5.2 2.0 virginica  
## 149 6.2 3.4 5.4 2.3 virginica

iris\_file2\_train

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa  
## 11 5.4 3.7 1.5 0.2 setosa  
## 12 4.8 3.4 1.6 0.2 setosa  
## 14 4.3 3.0 1.1 0.1 setosa  
## 17 5.4 3.9 1.3 0.4 setosa  
## 25 4.8 3.4 1.9 0.2 setosa  
## 26 5.0 3.0 1.6 0.2 setosa  
## 32 5.4 3.4 1.5 0.4 setosa  
## 33 5.2 4.1 1.5 0.1 setosa  
## 36 5.0 3.2 1.2 0.2 setosa  
## 38 4.9 3.6 1.4 0.1 setosa  
## 41 5.0 3.5 1.3 0.3 setosa  
## 45 5.1 3.8 1.9 0.4 setosa  
## 49 5.3 3.7 1.5 0.2 setosa  
## 59 6.6 2.9 4.6 1.3 versicolor  
## 65 5.6 2.9 3.6 1.3 versicolor  
## 68 5.8 2.7 4.1 1.0 versicolor  
## 76 6.6 3.0 4.4 1.4 versicolor  
## 78 6.7 3.0 5.0 1.7 versicolor  
## 81 5.5 2.4 3.8 1.1 versicolor  
## 82 5.5 2.4 3.7 1.0 versicolor  
## 84 6.0 2.7 5.1 1.6 versicolor  
## 87 6.7 3.1 4.7 1.5 versicolor  
## 89 5.6 3.0 4.1 1.3 versicolor  
## 90 5.5 2.5 4.0 1.3 versicolor  
## 92 6.1 3.0 4.6 1.4 versicolor  
## 93 5.8 2.6 4.0 1.2 versicolor  
## 96 5.7 3.0 4.2 1.2 versicolor  
## 100 5.7 2.8 4.1 1.3 versicolor  
## 106 7.6 3.0 6.6 2.1 virginica  
## 110 7.2 3.6 6.1 2.5 virginica  
## 111 6.5 3.2 5.1 2.0 virginica  
## 116 6.4 3.2 5.3 2.3 virginica  
## 119 7.7 2.6 6.9 2.3 virginica  
## 122 5.6 2.8 4.9 2.0 virginica  
## 127 6.2 2.8 4.8 1.8 virginica  
## 130 7.2 3.0 5.8 1.6 virginica  
## 139 6.0 3.0 4.8 1.8 virginica  
## 141 6.7 3.1 5.6 2.4 virginica  
## 142 6.9 3.1 5.1 2.3 virginica  
## 143 5.8 2.7 5.1 1.9 virginica  
## 145 6.7 3.3 5.7 2.5 virginica  
## 146 6.7 3.0 5.2 2.3 virginica  
## 150 5.9 3.0 5.1 1.8 virginica

idx 변수에서 보면 sample은 랜덤하게 수를 뽑아주는 함수이다. sample의 옵션을 보면,

2 는 랜덤하게 뽑는 수의 집합을 의미한다. 2일시, 1과 2중 랜덤하게 뽑는다.

nrow()는 행의 개수를 의미한다. 즉 행의 개수만큼 1,2를 랜던하게 뽑는다는 것이다.

replace는 복원수출을 하느냐 안하느냐를 선택하는 것이다.

prob는 랜덤하게 뽑는 수를 어떤 비율로 뽑을 것인지 적는것이다. 랜덤하게 뽑느 수의 집합의 개수만큼 적으면 된다. c()의 의미는 백터를 만들어주는 함수이다. 즉, c(0.7,0.3)은 1은 0.7, 2는 0.3비율만큼 뽑게된다.

여기서 정규화 할대 주위해야 될면은 어떤 목적으로 분석을 하느냐임으로 종속변수에 따라 나눠지는 방법이 다르므로 데이터 분리를 해줘야 될 수도 있다.

예를들어 iris의 데이터중 종(species)에 따라 무언가를 보고 싶을때는 위처럼 나누면 안된다.

iris\_train2 <- iris[c(sample(1:50,35), sample(51:100,35),sample(101:150,35)),]  
iris\_test2 <- iris[-c(sample(1:50,35), sample(51:100,35),sample(101:150,35)),]  
iris\_test2

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 9 4.4 2.9 1.4 0.2 setosa  
## 10 4.9 3.1 1.5 0.1 setosa  
## 11 5.4 3.7 1.5 0.2 setosa  
## 21 5.4 3.4 1.7 0.2 setosa  
## 25 4.8 3.4 1.9 0.2 setosa  
## 30 4.7 3.2 1.6 0.2 setosa  
## 31 4.8 3.1 1.6 0.2 setosa  
## 32 5.4 3.4 1.5 0.4 setosa  
## 35 4.9 3.1 1.5 0.2 setosa  
## 38 4.9 3.6 1.4 0.1 setosa  
## 39 4.4 3.0 1.3 0.2 setosa  
## 41 5.0 3.5 1.3 0.3 setosa  
## 42 4.5 2.3 1.3 0.3 setosa  
## 54 5.5 2.3 4.0 1.3 versicolor  
## 58 4.9 2.4 3.3 1.0 versicolor  
## 59 6.6 2.9 4.6 1.3 versicolor  
## 62 5.9 3.0 4.2 1.5 versicolor  
## 64 6.1 2.9 4.7 1.4 versicolor  
## 68 5.8 2.7 4.1 1.0 versicolor  
## 70 5.6 2.5 3.9 1.1 versicolor  
## 71 5.9 3.2 4.8 1.8 versicolor  
## 72 6.1 2.8 4.0 1.3 versicolor  
## 75 6.4 2.9 4.3 1.3 versicolor  
## 78 6.7 3.0 5.0 1.7 versicolor  
## 81 5.5 2.4 3.8 1.1 versicolor  
## 84 6.0 2.7 5.1 1.6 versicolor  
## 89 5.6 3.0 4.1 1.3 versicolor  
## 96 5.7 3.0 4.2 1.2 versicolor  
## 103 7.1 3.0 5.9 2.1 virginica  
## 107 4.9 2.5 4.5 1.7 virginica  
## 109 6.7 2.5 5.8 1.8 virginica  
## 110 7.2 3.6 6.1 2.5 virginica  
## 111 6.5 3.2 5.1 2.0 virginica  
## 116 6.4 3.2 5.3 2.3 virginica  
## 119 7.7 2.6 6.9 2.3 virginica  
## 120 6.0 2.2 5.0 1.5 virginica  
## 126 7.2 3.2 6.0 1.8 virginica  
## 127 6.2 2.8 4.8 1.8 virginica  
## 128 6.1 3.0 4.9 1.8 virginica  
## 138 6.4 3.1 5.5 1.8 virginica  
## 141 6.7 3.1 5.6 2.4 virginica  
## 149 6.2 3.4 5.4 2.3 virginica  
## 150 5.9 3.0 5.1 1.8 virginica

iris\_train2

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 48 4.6 3.2 1.4 0.2 setosa  
## 29 5.2 3.4 1.4 0.2 setosa  
## 44 5.0 3.5 1.6 0.6 setosa  
## 16 5.7 4.4 1.5 0.4 setosa  
## 12 4.8 3.4 1.6 0.2 setosa  
## 23 4.6 3.6 1.0 0.2 setosa  
## 35 4.9 3.1 1.5 0.2 setosa  
## 11 5.4 3.7 1.5 0.2 setosa  
## 13 4.8 3.0 1.4 0.1 setosa  
## 45 5.1 3.8 1.9 0.4 setosa  
## 26 5.0 3.0 1.6 0.2 setosa  
## 31 4.8 3.1 1.6 0.2 setosa  
## 33 5.2 4.1 1.5 0.1 setosa  
## 25 4.8 3.4 1.9 0.2 setosa  
## 27 5.0 3.4 1.6 0.4 setosa  
## 36 5.0 3.2 1.2 0.2 setosa  
## 34 5.5 4.2 1.4 0.2 setosa  
## 43 4.4 3.2 1.3 0.2 setosa  
## 8 5.0 3.4 1.5 0.2 setosa  
## 47 5.1 3.8 1.6 0.2 setosa  
## 40 5.1 3.4 1.5 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 22 5.1 3.7 1.5 0.4 setosa  
## 49 5.3 3.7 1.5 0.2 setosa  
## 20 5.1 3.8 1.5 0.3 setosa  
## 1 5.1 3.5 1.4 0.2 setosa  
## 9 4.4 2.9 1.4 0.2 setosa  
## 19 5.7 3.8 1.7 0.3 setosa  
## 39 4.4 3.0 1.3 0.2 setosa  
## 30 4.7 3.2 1.6 0.2 setosa  
## 37 5.5 3.5 1.3 0.2 setosa  
## 14 4.3 3.0 1.1 0.1 setosa  
## 50 5.0 3.3 1.4 0.2 setosa  
## 21 5.4 3.4 1.7 0.2 setosa  
## 61 5.0 2.0 3.5 1.0 versicolor  
## 63 6.0 2.2 4.0 1.0 versicolor  
## 97 5.7 2.9 4.2 1.3 versicolor  
## 91 5.5 2.6 4.4 1.2 versicolor  
## 94 5.0 2.3 3.3 1.0 versicolor  
## 85 5.4 3.0 4.5 1.5 versicolor  
## 82 5.5 2.4 3.7 1.0 versicolor  
## 69 6.2 2.2 4.5 1.5 versicolor  
## 64 6.1 2.9 4.7 1.4 versicolor  
## 81 5.5 2.4 3.8 1.1 versicolor  
## 73 6.3 2.5 4.9 1.5 versicolor  
## 98 6.2 2.9 4.3 1.3 versicolor  
## 88 6.3 2.3 4.4 1.3 versicolor  
## 74 6.1 2.8 4.7 1.2 versicolor  
## 67 5.6 3.0 4.5 1.5 versicolor  
## 86 6.0 3.4 4.5 1.6 versicolor  
## 66 6.7 3.1 4.4 1.4 versicolor  
## 60 5.2 2.7 3.9 1.4 versicolor  
## 65 5.6 2.9 3.6 1.3 versicolor  
## 53 6.9 3.1 4.9 1.5 versicolor  
## 90 5.5 2.5 4.0 1.3 versicolor  
## 68 5.8 2.7 4.1 1.0 versicolor  
## 70 5.6 2.5 3.9 1.1 versicolor  
## 57 6.3 3.3 4.7 1.6 versicolor  
## 99 5.1 2.5 3.0 1.1 versicolor  
## 72 6.1 2.8 4.0 1.3 versicolor  
## 56 5.7 2.8 4.5 1.3 versicolor  
## 100 5.7 2.8 4.1 1.3 versicolor  
## 55 6.5 2.8 4.6 1.5 versicolor  
## 52 6.4 3.2 4.5 1.5 versicolor  
## 89 5.6 3.0 4.1 1.3 versicolor  
## 80 5.7 2.6 3.5 1.0 versicolor  
## 71 5.9 3.2 4.8 1.8 versicolor  
## 96 5.7 3.0 4.2 1.2 versicolor  
## 79 6.0 2.9 4.5 1.5 versicolor  
## 139 6.0 3.0 4.8 1.8 virginica  
## 144 6.8 3.2 5.9 2.3 virginica  
## 105 6.5 3.0 5.8 2.2 virginica  
## 128 6.1 3.0 4.9 1.8 virginica  
## 133 6.4 2.8 5.6 2.2 virginica  
## 115 5.8 2.8 5.1 2.4 virginica  
## 129 6.4 2.8 5.6 2.1 virginica  
## 131 7.4 2.8 6.1 1.9 virginica  
## 124 6.3 2.7 4.9 1.8 virginica  
## 116 6.4 3.2 5.3 2.3 virginica  
## 142 6.9 3.1 5.1 2.3 virginica  
## 119 7.7 2.6 6.9 2.3 virginica  
## 111 6.5 3.2 5.1 2.0 virginica  
## 130 7.2 3.0 5.8 1.6 virginica  
## 136 7.7 3.0 6.1 2.3 virginica  
## 108 7.3 2.9 6.3 1.8 virginica  
## 120 6.0 2.2 5.0 1.5 virginica  
## 103 7.1 3.0 5.9 2.1 virginica  
## 146 6.7 3.0 5.2 2.3 virginica  
## 135 6.1 2.6 5.6 1.4 virginica  
## 127 6.2 2.8 4.8 1.8 virginica  
## 125 6.7 3.3 5.7 2.1 virginica  
## 117 6.5 3.0 5.5 1.8 virginica  
## 113 6.8 3.0 5.5 2.1 virginica  
## 140 6.9 3.1 5.4 2.1 virginica  
## 149 6.2 3.4 5.4 2.3 virginica  
## 122 5.6 2.8 4.9 2.0 virginica  
## 118 7.7 3.8 6.7 2.2 virginica  
## 110 7.2 3.6 6.1 2.5 virginica  
## 148 6.5 3.0 5.2 2.0 virginica  
## 114 5.7 2.5 5.0 2.0 virginica  
## 134 6.3 2.8 5.1 1.5 virginica  
## 143 5.8 2.7 5.1 1.9 virginica  
## 132 7.9 3.8 6.4 2.0 virginica  
## 138 6.4 3.1 5.5 1.8 virginica

위에 보면 종별로 50개씩 나열이 되어있기 때문에 50개 중 35개씩(70%)를 각각 나눠서 저장했다.(코드는 1 ~ 50개중 35개 뽑기를 총 3번한 것, replace = F가 기본값이다.) 전체 비율은 7:3으로 지정하였다.

이처럼 목적에 따라 비율을 나눠줘야 원하는 값에 맞춰 분석을 정확하게 할 수 있다.

다음은 data\_file 데이터를 성별에 따라 나눠어 보자.

#data\_file2를 사용한다.  
data\_file2\_male <- data\_file2[data\_file2$성별 == 1,]  
data\_file2\_female <- data\_file2[data\_file2$성별 == 2,]  
  
#dplyr 패키지를 사용한다.  
#dplyr패키지 다운 안받았을시, install.packages("dplyr")할 것  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

data\_file2\_train <- rbind(sample\_frac(data\_file2\_male,0.7),sample\_frac(data\_file2\_female,0.7))  
data\_file2\_test <- rbind(sample\_frac(data\_file2\_male,0.7),sample\_frac(data\_file2\_female,0.7))  
  
data\_file2\_train

## ID 성별 연령 직업 통신사 가입년수 사용료 부족한서비스 남는서비스  
## 1 A103 1 0 0 1 8 7.5 5 1  
## 2 A102 1 0 0 1 3 5.5 5 5  
## 3 A79 1 26 1 2 8 3.5 3 1  
## 4 A30 1 21 0 1 10 6.5 5 1  
## 5 A36 1 21 0 1 10 3.5 3 2  
## 6 A37 1 21 0 1 10 3.5 5 5  
## 7 A107 1 0 1 2 8 6.5 5 1  
## 8 A7 1 20 0 2 3 1.5 3 2  
## 9 A96 1 39 1 2 8 3.5 1 3  
## 10 A5 1 20 0 1 3 3.5 3 1  
## 11 A104 1 0 1 1 10 6.5 3 2  
## 12 A34 1 21 0 2 10 3.5 3 1  
## 13 A22 1 20 0 2 8 4.5 3 1  
## 14 A16 1 20 0 2 10 7.5 4 2  
## 15 A55 1 22 0 3 3 6.5 3 5  
## 16 A94 1 36 1 1 10 5.5 4 3  
## 17 A17 1 20 0 2 3 6.5 4 2  
## 18 A59 1 22 0 1 8 8.5 3 5  
## 19 A8 1 20 0 1 3 3.5 3 2  
## 20 A81 1 26 1 1 3 6.5 3 3  
## 21 A60 1 23 0 2 8 3.5 3 2  
## 22 A10 1 20 0 1 3 5.5 3 1  
## 23 A124 1 0 0 1 10 6.5 3 1  
## 24 A9 1 20 0 2 3 6.5 1 2  
## 25 A108 1 0 0 2 3 6.5 4 4  
## 26 A14 1 20 0 1 8 4.5 3 1  
## 27 A113 1 23 0 2 3 3.5 3 2  
## 28 A12 1 20 0 2 3 3.5 3 2  
## 29 A62 1 23 0 2 3 3.5 3 2  
## 30 A32 1 21 0 1 10 4.5 3 1  
## 31 A67 1 24 0 3 3 3.5 3 1  
## 32 A88 1 29 1 1 3 5.5 3 1  
## 33 A105 1 0 0 2 8 6.5 5 5  
## 34 A61 1 23 0 3 3 6.5 3 2  
## 35 A4 1 19 0 1 8 4.5 5 3  
## 36 A52 1 22 0 1 3 6.5 5 1  
## 37 A28 1 21 0 1 3 4.5 3 1  
## 38 A100 1 49 1 1 3 6.5 5 3  
## 39 A74 1 25 0 2 10 6.5 5 1  
## 40 A50 1 22 0 1 3 6.5 5 5  
## 41 A85 1 27 1 2 3 8.5 3 4  
## 42 A118 1 23 0 3 3 5.5 5 2  
## 43 A23 1 20 0 2 8 4.5 3 1  
## 44 A18 1 20 0 1 8 4.5 3 4  
## 45 A101 1 50 1 1 8 6.5 3 1  
## 46 A111 1 20 0 1 3 5.5 3 2  
## 47 A48 1 22 0 1 8 7.5 3 3  
## 48 A95 1 36 1 1 3 6.5 3 2  
## 49 A33 1 21 0 1 10 6.5 3 2  
## 50 A43 1 21 0 3 3 6.5 5 1  
## 51 A51 1 22 0 3 3 6.5 1 4  
## 52 A73 1 25 1 1 8 7.5 5 5  
## 53 A64 1 23 0 2 3 7.5 5 5  
## 54 A29 2 21 0 3 3 7.5 5 1  
## 55 A106 2 0 0 1 3 5.5 3 4  
## 56 A116 2 20 0 2 3 6.5 5 5  
## 57 A24 2 20 0 3 3 8.5 5 3  
## 58 A117 2 20 0 1 3 4.5 3 2  
## 59 A115 2 22 0 1 8 3.5 3 4  
## 60 A68 2 24 0 1 3 5.5 3 4  
## 61 A76 2 26 1 2 3 3.5 3 2  
## 62 A92 2 34 1 1 8 5.5 1 4  
## 63 A25 2 20 0 1 3 8.5 3 5  
## 64 A39 2 21 0 2 3 3.5 3 2  
## 65 A53 2 22 0 1 8 5.5 3 2  
## 66 A26 2 20 1 1 8 7.5 3 3  
## 67 A41 2 21 0 2 3 3.5 3 2  
## 68 A86 2 27 1 3 3 7.5 4 1  
## 69 A44 2 21 0 2 3 5.5 3 1  
## 70 A57 2 22 0 1 8 4.5 5 2  
## 71 A56 2 22 0 1 3 10.0 5 2  
## 72 A89 2 30 1 1 3 7.5 3 1  
## 73 A40 2 21 0 1 8 5.5 3 2  
## 74 A114 2 22 0 2 3 6.5 3 2  
## 75 A72 2 25 0 1 10 7.5 4 4  
## 76 A70 2 24 0 2 10 3.5 3 2  
## 77 A91 2 31 1 1 10 9.5 3 1  
## 78 A71 2 25 1 1 3 6.5 3 2  
## 79 A21 2 20 0 2 10 5.5 3 4  
## 80 A13 2 20 0 3 8 3.5 3 4  
## 81 A122 2 22 0 1 3 3.5 3 2  
## 82 A54 2 22 1 3 3 7.5 5 2  
## 83 A120 2 22 0 1 8 3.5 3 2  
## 84 A77 2 26 0 2 10 6.5 4 1  
## 85 A82 2 26 1 3 3 6.5 5 2  
## 86 A47 2 21 0 1 3 3.5 3 1

data\_file2\_test

## ID 성별 연령 직업 통신사 가입년수 사용료 부족한서비스 남는서비스  
## 1 A16 1 20 0 2 10 7.5 4 2  
## 2 A81 1 26 1 1 3 6.5 3 3  
## 3 A5 1 20 0 1 3 3.5 3 1  
## 4 A51 1 22 0 3 3 6.5 1 4  
## 5 A38 1 21 0 1 8 3.5 3 1  
## 6 A104 1 0 1 1 10 6.5 3 2  
## 7 A46 1 21 0 2 3 3.5 3 1  
## 8 A62 1 23 0 2 3 3.5 3 2  
## 9 A94 1 36 1 1 10 5.5 4 3  
## 10 A60 1 23 0 2 8 3.5 3 2  
## 11 A110 1 0 0 3 3 3.5 3 1  
## 12 A74 1 25 0 2 10 6.5 5 1  
## 13 A10 1 20 0 1 3 5.5 3 1  
## 14 A111 1 20 0 1 3 5.5 3 2  
## 15 A18 1 20 0 1 8 4.5 3 4  
## 16 A9 1 20 0 2 3 6.5 1 2  
## 17 A95 1 36 1 1 3 6.5 3 2  
## 18 A28 1 21 0 1 3 4.5 3 1  
## 19 A36 1 21 0 1 10 3.5 3 2  
## 20 A118 1 23 0 3 3 5.5 5 2  
## 21 A65 1 23 0 2 3 4.5 3 2  
## 22 A90 1 30 0 2 3 5.5 5 5  
## 23 A37 1 21 0 1 10 3.5 5 5  
## 24 A67 1 24 0 3 3 3.5 3 1  
## 25 A99 1 49 1 1 8 3.5 3 1  
## 26 A73 1 25 1 1 8 7.5 5 5  
## 27 A27 1 21 0 1 8 5.5 3 1  
## 28 A69 1 24 0 1 8 4.5 3 2  
## 29 A23 1 20 0 2 8 4.5 3 1  
## 30 A97 1 40 1 1 10 3.5 1 1  
## 31 A59 1 22 0 1 8 8.5 3 5  
## 32 A55 1 22 0 3 3 6.5 3 5  
## 33 A103 1 0 0 1 8 7.5 5 1  
## 34 A35 1 21 0 1 8 5.5 3 2  
## 35 A85 1 27 1 2 3 8.5 3 4  
## 36 A31 1 21 0 1 8 5.5 3 1  
## 37 A98 1 40 1 1 10 8.5 3 2  
## 38 A102 1 0 0 1 3 5.5 5 5  
## 39 A109 1 0 1 3 8 7.5 3 4  
## 40 A88 1 29 1 1 3 5.5 3 1  
## 41 A11 1 20 0 2 3 4.5 3 1  
## 42 A107 1 0 1 2 8 6.5 5 1  
## 43 A124 1 0 0 1 10 6.5 3 1  
## 44 A33 1 21 0 1 10 6.5 3 2  
## 45 A84 1 26 0 1 8 5.5 5 5  
## 46 A20 1 20 0 2 3 6.5 4 3  
## 47 A113 1 23 0 2 3 3.5 3 2  
## 48 A12 1 20 0 2 3 3.5 3 2  
## 49 A48 1 22 0 1 8 7.5 3 3  
## 50 A2 1 18 0 1 3 1.5 3 1  
## 51 A50 1 22 0 1 3 6.5 5 5  
## 52 A4 1 19 0 1 8 4.5 5 3  
## 53 A8 1 20 0 1 3 3.5 3 2  
## 54 A77 2 26 0 2 10 6.5 4 1  
## 55 A121 2 24 0 2 3 1.5 3 2  
## 56 A15 2 20 0 1 3 3.5 3 4  
## 57 A42 2 21 0 2 8 4.5 3 2  
## 58 A123 2 20 0 3 3 5.5 3 4  
## 59 A72 2 25 0 1 10 7.5 4 4  
## 60 A24 2 20 0 3 3 8.5 5 3  
## 61 A3 2 19 1 3 8 10.0 5 4  
## 62 A53 2 22 0 1 8 5.5 3 2  
## 63 A58 2 22 0 3 8 6.5 5 4  
## 64 A75 2 25 1 3 3 6.5 5 1  
## 65 A44 2 21 0 2 3 5.5 3 1  
## 66 A122 2 22 0 1 3 3.5 3 2  
## 67 A120 2 22 0 1 8 3.5 3 2  
## 68 A106 2 0 0 1 3 5.5 3 4  
## 69 A76 2 26 1 2 3 3.5 3 2  
## 70 A39 2 21 0 2 3 3.5 3 2  
## 71 A54 2 22 1 3 3 7.5 5 2  
## 72 A70 2 24 0 2 10 3.5 3 2  
## 73 A114 2 22 0 2 3 6.5 3 2  
## 74 A87 2 27 1 3 3 6.5 3 4  
## 75 A45 2 21 0 1 3 5.5 3 5  
## 76 A71 2 25 1 1 3 6.5 3 2  
## 77 A80 2 26 1 3 3 4.5 3 2  
## 78 A66 2 24 0 3 8 3.5 5 2  
## 79 A86 2 27 1 3 3 7.5 4 1  
## 80 A117 2 20 0 1 3 4.5 3 2  
## 81 A57 2 22 0 1 8 4.5 5 2  
## 82 A25 2 20 0 1 3 8.5 3 5  
## 83 A56 2 22 0 1 3 10.0 5 2  
## 84 A82 2 26 1 3 3 6.5 5 2  
## 85 A19 2 20 0 2 8 8.5 3 2  
## 86 A49 2 22 0 2 3 5.5 3 2

먼저 data\_file2[data\_file2$성별 == 1,]에 대해 설명하자면,

data\_file2는 변수 이름이고, data\_file2[]는 data\_file2를 테이블형식으로 받아온다는 것이다.

[]안에 내용을 보면, ’data\_file2$성별 == 1’에서

y)라고 예제를 들어 설명하면,> y of x(x의 y)라는 뜻으로 data\_file2의 성별 데이터를 의미한다.

“==”의 의미는 A == B로 설명을 들면 A가 B랑 같나? 라는 의미이다. 반드시 B가 비교대상이 되어야한다.(A == B 와 B == A는 컴퓨터가 읽는 의미가 달라지기 때문이다. 의미가 같을지는 모를지라도 컴퓨터가 보기에는 해석의 순서가 달라진다. 비교대상이 A인지 B인지 달라기지 때문에 순서에 유의하자.)

즉 data\_file2[data\_file2$성별 == 1,]를 설명하자면, data\_file2 중에서 data\_file2의 성별이 1인 데이터 값 row에 그 외 나머지 부분을 전체 호출하라는 의미이다.

(data[x,y]의 형태가 data\_file2[data\_file2$성별 == 1,]이기 때문이다.)

다음으로는 dplyr package에 포함된 함수 중 sample\_frac은 sample\_frac(data = , prob = , replace = ) 형태로 쓰이는데, data 중 prob 비율대로 replace의 여부에 따라 랜덤하게 데이터를 뽑는 함수이다.

rbind는 데이터 내용이 다른 두 데이터프레임의 내용을 합쳐주는 것이다. rbind의 r는 row의 줄인말로, row기준으로 데이터를 합쳐준다. cbind는 col의 줄일말로, col기준으로 데이터를 합쳐준다.

……………………………………………………………………………………………………………………………………………………………………………

다음 과정을 나가전에 이해되었는지 확인하기 위해 문제 하나 풀어보자.

1. iris 데이터 중 Sepal.width이 3.0이상인 집단과 미만인 집단으로 나눈 뒤, 각 종들의 비율이 어느정도인지 iris\_sepal.width\_over3.0, iris\_sepal.width\_less3.0으로 각각 저장하여 호출하시오. hint : nrow() : row의 총 개수를 알려주는 함수, 글자는 " "써줘야 인식한다.

ex) iris$Species == “setosa”

그리고 row의 이름을 바꾸는 함수는 rownames(data) <- c(“이름”,“이름”…)을 row의 개수에 맞게 입력하면 된다.

ex ) rownames(x) <- c(“setosa”, “versicolor”,“virginica”)

## iris\_sepal.width\_less3.0 iris\_sepal.width\_over3.0  
## setosa 0.03508772 0.516129  
## versicolor 0.59649123 0.172043  
## virginica 0.36842105 0.311828