**Final Exam (Statistical Computing 1)**

December 15th, 2021 (10:30~11:45 am)

|  |  |  |
| --- | --- | --- |
| 학과: 벤처중소기업학과/정보통계보험수리학과 | 학번: 20161759 | 이름: 최봉수 |

[Part 2] R을 사용하여 아래 질문에 답하시오. 오픈북 시험으로 교재, 수업 노트, 개인 정리 노트는 참고 가능하지만 인터넷 사용은 금지되며 R 코드와 출력 결과를 정리하여 스마트 캠퍼스에 제출하시오.

1. 교육 수준 (education)과 흡연율 (smoking)간의 관련성을 분석하기 위한 연구 가설을 세우고 단계별로 가설을 검정하시오. (25점)

(1) smoking.csv파일을 읽고 아래 표의 빈칸을 채우시오. [5점]

|  |
| --- |
| <Code>  ## 아래 부분에 코드를 작성하시오.  setwd("/Users/bongsu/Downloads/data\ 2")  getwd()  answer1 = read.csv('smoking.csv', fileEncoding = "CP949")  table(answer1) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | heavy | normal | non-smoking | 합계 |
| college | 51 | 92 | 68 | 211 |
| high | 22 | 21 | 9 | 52 |
| middle | 43 | 28 | 21 | 92 |
| 합계 | 116 | 141 | 98 |  |

(2) 독립성 검정을 수행하고 가설 및 검정 결과 그리고 결론을 쓰시오. [10점]

|  |
| --- |
| <Code>  ## 아래 부분에 코드를 작성하시오.  install.packages('gmodels')  library(gmodels)  x = answer1$education  y = answer1$smoking  chisq.test(x,y) |

|  |  |
| --- | --- |
| 교육수준과 흡연율 사이에 관련성이 없다. | 교육수준과 흡연율 사이에 관련성이 있다.. |

가설

검정결과 (빈칸을 채우시오)

|  |  |  |
| --- | --- | --- |
| Chi-squared | df | p-value |
| 18.911 | 4 | 0.0008183 |

|  |
| --- |
| p-value < 0.05이므로 대립가설인 교육수준과 흡연율 사이에는 관련성이 있다고 할 수 있다. |

결론

(3) Maria DB에 접속하여 (1)에서 smoking.csv파일을 읽어 저장한 객체를 “smoking”이라는 이름의 table로 Maria DB에 저장하고 새로 만든 “smoking” table에서 조건 where smoking='3.non-smoking' and education='2.high' 를 만족하는 행을 불러오는 Rㅏ코드와 결과값을 쓰시오. [10점]

|  |
| --- |
| <Code>  ## 아래 부분에 코드를 작성하시오.  install.packages("DBI")  install.packages('RMySQL') # mysql 돌리기  library(DBI)  library(RMySQL) # 맥에서 안돼서 다른방법찾을 찾음  drv <- dbDriver("MySQL")  conn <- dbConnect(drv, username="scott", password="tiger", dbname ="work", host="localhost")  dbSendQuery(conn, "SET NAMES utf8;") # 콘솔에 한글 나오게 하기  dbSendQuery(conn, "SET CHARACTER SET utf8;") # 콘솔에 한글 나오게 하기  dbSendQuery(conn, "SET character\_set\_connection=utf8;") # 콘솔에 한글 나오게 하기  answer1  dbWriteTable(conn, "smoking2", answer1)  dbGetQuery(conn, "select \* from smorking where smoking=3.non-smoking and education=2.high")  <Result>  ## 결과 값을 복사해서 붙여 넣으시오. |

2. “user\_data.csv” 파일을 읽고 아래 질문에 답하시오. (15점)

(1) 주어진 “user\_data.csv” 파일을 user\_data 이름으로 읽고 house\_type이 1이면 “단독주택”, 2이면 “다세대주택”, 3이면 “아파트”, 4이면 “오피스텔”로 바꾸고 age\_group이라는 새로운 변수를 만들어서 age가 40미만이면 “40세미만”, 40이상이면 “40세이상”으로 저장하시오. 아래 <Result> 부분의 table() 함수를 실행하여 결과를 붙여 넣으시오 [10점]

|  |
| --- |
| <Code>  ## 아래 부분에 코드를 쓰시오  user\_data <- read.csv('user\_data.csv', fileEncoding = "CP949")  user\_data  user\_data$house\_type = factor(user\_data$house\_type,labels = c('단독주택', '다세대주택', '아파트', '오피스텔'))  user\_data$age\_group = ifelse(user\_data$age >= 40, "40대 이상", "40대 미만")  <Result>  ## 실행 결과를 붙여 넣으시오  table(user\_data$age\_group,user\_data$house\_type)  table(user\_data$age\_group, user\_data$house\_type)    단독주택 다세대주택 아파트 오피스텔  40대 미만 0 1 2 25  40대 이상 32 46 19 275 |

(2) ggplot2 라이브러리를 이용하여 house type별 age를 box plot으로 그리시오. 작성된 코드는 <Code>에 쓰고 box plot 그림은 <Result>에 붙여 넣으시오 (그림에서 x 축은 house\_type, y 축은 age으로 하여 네개의 Box를 그리면 됨) [5점]

|  |
| --- |
| <Code>  ## 아래 부분에 ggplot함수를 이용한 코드를 쓰시오  library('ggplot2')  library(datasets)  user\_data <- read.csv('user\_data.csv', fileEncoding = "CP949")  user\_data$house\_type = factor(user\_data$house\_type,labels = c('단독주택', '다세대주택', '아파트', '오피스텔'))  par  answer2 = ggplot(user\_data, aes(x=user\_data$house\_type, y=user\_data$age)) + geom\_boxplot() + scale\_y\_continuous(name = "age") + scale\_x\_discrete(name = "house\_type")  x11()  answer2  <Result>  ## box plot을 붙여 넣으시오. |

3. R의 객체 지향 언어 (OOP)와 관련해서 참조 클래스 (RC Object)를 만들고 다음을 호출하시오 (20점)

(1) Class1이라는 참조 클래스 (RC Object)를 만들고 score라는 변수를 numeric 데이터 타입으로 설정하시오. Class1 클래스의 score라는 변수에 99를 넣는 생성자를 실행시켜 c1이라는 객체를 만든 다음 c1$score를 실행시키시오. [10점]

|  |
| --- |
| <Code>  ## 아래 부분에 코드를 작성하시오.  answer3 = setRefClass('Class1', fields=list(score="numeric"))  c1 = answer3$new(score=99)  c1$score  <Result>  ## 결과 값을 복사해서 붙여 넣으시오.  c1$score  [1] 99 |

(2) Class1을 상속받은 Class2라는 참조 클래스 (RC Object)를 만드시오. Class2에 getCredit()이라는 함수를 만들고 score가 90점 이상이면 A를 출력하고 90미만이면 B를 출력하도록 하시오. Class2 클래스의 score라는 변수에 89를 넣는 생성자를 실행시켜 c2라는 객체를 만든 다음 c2$getCredit()함수를 실행시키시오. [10점]

|  |
| --- |
| <Code>  ## 아래 부분에 코드를 작성하시오.  answer3 = setRefClass('Class1', fields=list(score="numeric"))  answer3\_1 = setRefClass("Class2",  contains='Class1',  method=list(  getCredit = function(x) {  if(score > 90) {  "A"  } else {  "B"  }  }  ))  c2 = answer3\_1$new(score=89)  c2$getCredit()  <Result>  ## 결과 값을 복사해서 붙여 넣으시오.  c2$getCredit()  [1] "B" |