

평가방법: 포트폴리오 – R 기본 및 통계 기본

평가일시	2017년 09월 24일		
과정명	SBA 빅데이터 사이언스 실무자 양성과정		
교과목	R기본 및 통계기본	담당교사	임 동 조 (인)
단원명	R기본 및 통계기본		

구분 (능력단위요소)	문 항	점 수
	<p>(문항 1~ 문항 10) 3점, (문항 11~ 문항 14) 5점</p> <p>1. dplyr 설치 후, 호출하여 R에서 사용하려고 합니다. 빈칸에 알맞은 함수를 아래에서 골라주세요. _____ (dplyr)</p> <p>(1) attach (2) import (3) install (4) library</p> <p>(정답) _____</p> <p>2. 데이터 프레임 만들기</p> <pre>test_1 <- _____(name = c('hanna','joy','tei','cabin','mike','beky','mimi','daniel','hans','rucy') , class = c(1, 2, 3, 3, 4, 3, 3, 2, 1, 2))</pre> <pre>test_2 <- _____(name = c('hanna','mimi','daniel','mike','beky','joy','tei','cabin','hans','rucy') , english = c(55, 50, 45, 60, 70, 80, 90, 85, NA, 60) , math = c(70, 83, NA, 95, 80, NA, 45, 60, 70, 80))</pre> <p>(1) data.frame (2) as.data.frame (3) frame (4) as.frame</p> <p>(정답) _____</p>	50/50점

3. name 컬럼을 기준으로 데이터를 결합하려고 합니다. 빈 칸에 알맞지 않은 것을 골라 주세요.

df <- _____(test_1, test_2, by = "name")

- (1) merge
- (2) inner_join
- (3) left_join
- (4) join

(정답) _____

4. 파생변수 추가하기

영어점수가 70점 이상이면 'PASS', 그렇지 않으면 'FAIL'

수학점수가 70점 이상이면 'PASS', 그렇지 않으면 'FAIL'

df %>%

_____ (eng_result = ifelse(english >= 70, 'PASS', 'FAIL')
,math_result = ifelse(math >= 70, 'PASS', 'FAIL'))

- (1) select
- (2) arrange
- (3) mutate
- (4) summarise

(정답) _____

5. 'gridExtra' 설치하기

_____('gridExtra')

- (1) install.package
- (2) installed.package
- (3) install.packages
- (4) installed.packages

(정답) _____

6. tr_mod.csv 파일 읽어오기

df <- _____("tr_mod.csv")

- (1) read.csv
- (2) read.table
- (3) read.spss
- (4) read_excel

(정답) _____

7. df 행열 확인하기

_____(df)

- (1) dim
- (2) sum
- (3) ncol
- (4) nrow

(정답) _____

8. df 컬럼명을 확인 할 수 있는 함수 모두 고르시오.

_____(df)

- (1) renames
- (2) colnames
- (3) rownames
- (4) names

(정답) _____

9. df 컬럼별 결측치 갯수를 확인하기

_____(is.na(df))

- (1) sum
- (2) table
- (3) colSums
- (4) dim

(정답) _____

10. 빈도표는 변수의 각 값들이 몇 개씩 존재하는지, 데이터의 개수를 나타낸 표이다.

_____(df\$Pclass) # Pclass 별 인원 파악하기

- (1) cor.test
- (2) t.test
- (3) prob.table
- (4) table

(정답) _____

11. 데이터 탐색 시 사용되는 함수가 아닌 것은?

- (1) str
- (2) attach
- (3) names
- (4) dim

(정답) _____

12. 빈칸을 채워주세요.(순서대로 번호를 적어 주세요. ex> 1,2,3) (5점)

df %>%

_____(!is.na(english) & !is.na(math)) %>% # 조건에 맞는 데이터만
필터링(NA값 제외.)

_____(class) %>% # class별로 그룹지어 나누기

_____(mean(english), mean(math)) # 그룹별로 요약하기

- (1) mutate
- (2) group_by
- (3) filter
- (4) summarise

(정답) _____

13. ip , app, device, os, channel 5개의 변수를 이용하여 is_attributed을 예측하는 모델을 만들어보자

```
model <- glm(_____, family = binomial,  
data = _____)
```

```
new_df <- data.frame( ip = c(2948, 2684, 2685, 2686)  
, app = c(19 , 286, 21, 23)  
, device = c(0, 1, 1, 2)  
, os = c(13 , 25, 13, 13)  
, channel = c(213, 256, 122, 459))
```

(정답) _____

14. 위에서 만들어진 model을 통해 new_df의 is_attributed를 예측해 보자.

```
pred <- _____(model, newdata = new_df, type = "response")
```

- (1) rpart
- (2) lm
- (3) sample
- (4) predict

(정답) _____

총점