

SAS Statistics Midterm Exam

2016/04/22

* 기본배점은 3점이며 문항에 따로 표시되어 있는 경우는 5점임.

1. SAS의 기본 window에 대한 다음 설명 중 틀린 것은?
 - a. 프로그램 에디터(program editor)는 SAS code를 입력 또는 편집할 때 사용한다.
 - b. 출력(output)창을 통하여서는 프로그램 실행 결과를 확인할 수 있다.
 - c. 로그(log)창에서는 프로그램의 실행 과정에서 나타나는 메시지들을 확인할 수 있다.
 - d. 출력(output)은 HTML 과 list 둘 중 하나를 선택해서 출력하여야 한다.
2. DATA step 에 대한 설명 중 맞는 것을 골라라.
 - a. 항상 "DATA"라는 구문으로 시작한다.
 - b. SAS 데이터셋(dataset)을 생성하거나 수정할 뿐 결과를 출력할 수는 없다.
 - c. run 이 나오면 해당 DATA step의 종료를 의미한다.
 - d. a~c 모두 맞다.
3. "exam3"라는 이름의 permanent SAS library를 정의하고 "D:\users" 폴더에 할당하기 위한 문장은 무엇인가?
 - a. libname 'D:\users' exam3;
 - b. libname exam3 'D:\users';
 - c. filename 'D:\users' exam3;
 - d. libname work.exam3 'D:\users';
4. SAS dataset에 대한 설명 중 틀린 것은? (복수응답 가능)
 - a. descriptor portion과 data portion으로 구분된다.
 - b. 문자형(character)과 수치형(numeric) 값들을 포함할 수 있다.
 - c. 각 data value의 길이는 최대 8자리를 초과할 수 없다.
 - d. 결측치(missing value)는 . (period) 으로 처리된다.
5. 다음 중 SAS dataset의 이름으로 적절하지 않은 것은?
 - a. 2apple
 - b. _max2
 - c. Ab_7
 - d. z123456
6. 1919년 3월 1일을 입력하고자 한다. 03/01/19와 같은 형태로 입력하고자 할 때 쓸 수 있는 적절한 옵션은 무엇인가?
 - a. options year 1900;
 - b. options yearcutoff 1900;
 - c. options yearcutoff=1900;
 - d. 따로 옵션을 지정할 필요 없다.
7. 총 50명의 학생에 대한 성적이 기록된 자료가 있다. 이 중 11번째에서 50번째 학생에 대한 자료만을 출력하고자 할 때 다음 빈칸을 적절히 채워라.

```
options  ;  
proc print data=score;  
run;
```

8. 다음과 같은 출력결과를 얻고자 할 때 적절한 프로그램은 무엇인가?

The SAS System			
Name	Name	Height	Weight
Murray, W	Murray, W	72	168
Almers, C	Almers, C	66	152
Bonaventure, T	Bonaventure, T	61	123
Johnson, R	Johnson, R	63	137
LaMance, K	LaMance, K	71	158

- a. proc print data=exam8;
 id name height weight;
run;
- b. proc print data=exam8;
 var name height weight;
run;
- c. proc print data=exam8;
 var name height weight;
 id name;
run;
- d. proc print data=exam8 noobs;
 var name height weight;
 id name height weight;
run;

9. 나이(age)가 30세 이하이면서 성별(sex)이 남성(M)인 경우만 출력하고자 할 때 적절한 code를 써라.

```
proc print data=exam9;  
    where  ;  
run;
```

10. (5점) exam10 데이터에는 나이(age), 성별(sex), 연봉(money) 변수에 대한 정보가 포함되어 있다. proc print를 이용하여 나이가 30세를 초과하는 사람들에 대해서 연봉의 총합을 출력하되 남성, 여성 두 그룹으로 나누어 출력하는 code를 써라.

11. exam11 데이터에는 height 변수가 포함되어 있다. proc print를 이용하여 이를 출력하되, 이 변수가 cm로 측정되었음을 표시하기 위해 적절한 것은?

- a. var height 'Height in cm';
- b. id height 'Height in cm';
- c. label height 'Height in cm';
- d. print height 'Height in cm';

12. 1270.32라는 data value가 입력되어 있고 이것은 달러를 나타내는 수치라고 할 때, 이를 \$를 표시하기 위하여 적절한 포맷은? (복수응답 가능)

- a. dollar8.2
- b. dollar8.1
- c. comma8.1
- d. dollar10.1

13. (5점) 'exam14.dat' 파일이 'C:\users' 폴더에 위치되어 있다. 파일 안에는 하나의 변수에 대한 자료가 입력되어 있으며 1~4열에 변수에 대한 수치가 위치하고 있다고 할 때, 본 파일을 불러들여 적절한 SAS dataset으로 저장하기 위한 프로그램을 작성하여라.

14~15. 어떤 class의 중간, 기말시험 성적이 입력된 자료를 다음과 같이 불러들여 exam14라는 SAS dataset으로 저장하려 한다.

```
data exam14;  
  infile exam14;  
  input mid 1-3 final 5-7;
```

```
run;
```

14. 중간, 기말시험의 평균점수를 나타내는 변수를 생성하여라.

15. 기말시험점수가 80점 이상이면 'Excellent', 그 이하면 'Good'이라는 메시지를 포함하는 변수를 생성하여라.

16. 다음 연산자(operator)들의 표현 중 잘못된 것은? (좌:연산, 우:SAS 문법) (복수응답 가능)

- a. $1+1 \Rightarrow 1+1$
- b. $2^3 \Rightarrow 2*3$
- c. $\frac{4}{7} \Rightarrow 4|7$
- d. $X \neq 2 \Rightarrow X^{\wedge}=2$

17. means procedure를 이용하여 요약통계량을 출력하고자 한다. 다음 중 대상변수로 적절하지 않은 것은?

- a. 나이
- b. 키
- c. 성별
- d. 혈압

18. 다음과 같은 출력결과를 주는 프로그램을 작성하여라.

The MEANS Procedure		
Variable	Mean	Median
Age	46.70000	49.50000
Height	66.95000	66.00000
Weight	174.65000	163.00000

```
proc means data=exam18  ;
run;
```

19. class와 by에 대한 설명 중 옳은 것을 골라라 (복수응답 가능).

- a. 둘 다 subgroup을 만들어 작업을 수행할 수 있게 해 준다.
- b. 둘 다 sort에 의해 그룹화를 위한 기준변수로 정렬시켜 주는 것이 필요하다.
- c. 기준변수는 문자형 변수로 선택하는 것이 일반적으로 좋다.
- d. 둘 이상의 기준 변수를 설정하는 것도 가능하다.

20. freq 프로시저를 이용하여 sex, actlevel 두 문자형(character)변수들에 대하여 각각 도수분포표를 출력하는 문장은?

- a. tables sex actlevel;
- b. tables sex*actlevel;
- c. by sex actlevel; tables sex actlevel;
- d. by sex actlevel; tables sex*actlevel;

21. 다음과 같은 출력결과를 얻고자 한다. 적절히 빈칸을 채워라.

The FREQ Procedure				
Frequency Col Pct	Table of Sex by ActLevel			
	Sex	ActLevel		
		HIGH	LOW	MOD
	F	4 57.14	4 57.14	3 42.86
	M	3 42.86	3 42.86	4 57.14
	Total	7	7	7

```
proc freq data=exam21;
 ;
run;
```

22. 다음과 같은 출력결과를 얻고자 한다. 적절히 빈칸을 채워라.

The FREQ Procedure						
Sex	ActLevel	Fee	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	HIGH	124.80	2	9.52	2	9.52
F	HIGH	149.75	2	9.52	4	19.05
F	LOW	85.20	2	9.52	6	28.57
F	LOW	124.80	1	4.76	7	33.33
F	LOW	149.75	1	4.76	8	38.10
F	MOD	124.80	1	4.76	9	42.86
F	MOD	149.75	2	9.52	11	52.38
M	HIGH	85.20	2	9.52	13	61.90
M	HIGH	124.80	1	4.76	14	66.67
M	LOW	124.80	2	9.52	16	76.19
M	LOW	149.75	1	4.76	17	80.95
M	MOD	124.80	1	4.76	18	85.71
M	MOD	149.75	3	14.29	21	100.00

```
proc freq data=exam22;
```

```
    _____ ;
```

```
run;
```

23. exam23에는 2016년 1~4월 월별 매출(sales)이 기록되어 있다 (단위:만원). 2016년 월별 누적매출액을 기록한 변수 cum_sales를 생성하고자 할 때 적절히 빈칸을 채워라.

```
data new;
```

```
    set exam23;
```

```
    _____ ;
```

```
run;
```

24. 위 문항에서 2015년 총 매출액이 9400만원이었다고 할 때, 이를 합산한 2016년도 월별누적매출액을 계산하기 위한 프로그램을 완성하여라.

```
data new;
```

```
    set exam23;
```

```
    _____
```

```
run;
```

25. (5점) exam25에는 총 50명에 대한 총 합산성적(total : 100점만점)이 기록되어 있다. 90점 이상이면 A, 80점 이상 90점 미만이면 B, 나머지는 C 학점을 할당하여 새로운 변수를 만들고자 한다. 50명 중 각각 25명, 15명, 10명이 A,B,C에 해당한다고 할 때 적절히 프로그램을 완성하되 가능하면 효율적으로 (계산량이 적도록) 하여라.

```
data new;
```

```
set exam25;
```

```
run;
```

26. 위 문항에서 재수강생(retake : T가 재수강생, F는 초수강생)이 있어 이 학생들은 데이터셋에서 제외하고자 한다. 적절한 문장은? (복수응답 가능)

- a. if retake='T' then delete;
- b. if retake='F';
- c. drop retake='T';
- d. drop retake;

27. exam27에는 age, sex, height, weight 네 개의 변수가 있다. 이 중 age, sex만 포함된 새로운 데이터셋을 만들고자 한다. 적절한 것은? (복수응답 가능)

- a. data new(drop=height weight);
set exam27;
run;
- b. data new(keep=age sex);
set exam27;
run;
- c. data new;
set exam27;
keep age sex;
run;
- d. data new;
set exam27;
keep=age sex;
run;

28. (5점) 다음 프로그램은 무엇을 계산하기 위한 것인지 간단히 설명하여라.

```
data exam28;
```

```
amount=1000;
```

```
rate=0.003;
```

```
do month=1 to 12;
```

```
earned+(amount+earned)*rate;
```

```
end;
```

```
run;
```

29. (5점) 다음을 계산하기 위한 프로그램을 작성하여라.

$$\sum_{i=1}^5 \sum_{j=1}^i (2j+1)$$

30. do while과 do until 문에 대한 설명 중 틀린 것은?

- a. do while은 주어진 조건이 만족되는 한 계속 실행된다.
- b. do until은 무조건 한 번 이상은 실행된다.
- c. do while은 무조건 한 번 이상은 실행된다.
- d. index를 이용한 반복문과 함께 쓸 수 있다.

31. 다음과 같은 raw data를 읽어들이기 위한 input 문을 완성하여라.

↓-----10-----20-----	
EVANS	DONNY 112 29,996.63
HELMS	LISA 105 18,567.23
HIGGINS	JOHN 111 25,309.00
LARSON	AMY 113 32,696.78
MOORE	MARY 112 28,945.89
POWELL	JASON 103 35,099.50
RILEY	JUDY 111 25,309.00

```
data empinfo;
  infile empdata;
  input _____ ;
run;
```

32. 다음과 같은 raw data를 읽어들이기 위한 input 문을 완성하여라.

Raw Data File Credit

1	---	+	---	10	---	+	---	20
MALE	27	1	8	0	0			
FEMALE	29	3	14	5	10			
FEMALE	34	2	10	3	3			
MALE	35	2	12	4	8			
FEMALE	36	4	16	3	7			
MALE	21	1	5	0	0			
MALE	25	2	9	2	1			
FEMALE	21	1	4	2	6			
MALE	38	3	11	4	3			
FEMALE	30	3	5	1	0			

data credit;

infile credit ;

input ;

run;

33. 다음 설명 중 틀린 것은?

- fixed-file로 입력되어 있는 파일을 column input 방식으로 읽을 수 있으면 formatted input 방식으로 읽을 수 있다.
- fixed-file로 입력되어 있는 파일은 특정 data 영역을 반복해서 읽는 것도 가능하다.
- list input은 공백(blank)이나 다른 구분자(delimiter)로 data value들이 모두 구분되어야 정상적으로 읽을 수 있다.
- list input으로 nonstandard numeric data를 읽을 때는 적절한 informat을 지정하여야 한다.