

**2과목 데이터정보처리입문 (36~60)**

출제위원 : 방송대 이기재

출제범위 : 교재 전범위(웹강의 포함)

36. 다음은 통계소프트웨어 R에 대한 설명이다. 옳은 것을 모두 고른 것은? (2점)

- I. 객체지향 프로그래밍 언어로 대화형 통계분석과 그래프 기능이 뛰어나다.  
 II. GUI 환경 하에서 통계분석 및 자료처리가 이루어지므로 쉽게 분석처리를 할 수 있다.  
 III. 미국에서 개발된 소프트웨어로 스프레드시트 형태의 데이터 입력을 취하고 있다.

- ① I  
 ② I, II  
 ③ I, III  
 ④ II, III

37. 다음과 같은 질문에 대한 응답 결과를 데이터로 바꿀 때 그 유형을 적절하게 구분한 것은? (2점)

- 최근 구입한 휴대폰에 대해서 어느 정도 만족하십니까?  
 ① 전혀 만족하지 않는다. ② 만족하지 않는다.  
 ③ 만족한다. ④ 아주 만족한다.

- ① 질적자료 - 명목형 ② 질적자료 - 순서형  
 ③ 양적자료 - 명목형 ④ 양적자료 - 순서형

38. 다음 설명과 가장 관련이 깊은 용어는 무엇인가? (2점)

수집된 설문지의 응답결과나 관측결과 등을 통계적으로 분석할 수 있도록 일정한 원칙에 따라 각 응답에 숫자를 부여하는 과정이다.

- ① 데이터 수집  
 ② 부호화(coding)  
 ③ 데이터 정선(data cleaning)  
 ④ 데이터 디버깅(data debugging)

39. 다음 측정 수준에 대한 설명 중 옳지 않은 것은? (3점)

- ① 구간척도에서 0값은 자의적으로 부여되었으므로 절대적 의미를 가질 수 없다.  
 ② 섭씨온도, 습도, 지능지수 등은 비율척도로 측정된 값이다.  
 ③ 변수의 측정수준은 적합한 통계 분석 기법의 선택에 영향을 미친다.  
 ④ 계란을 무게에 따라 대·중·소로 구분하였다면 순서척도로 측정된 것이다.

40. 다음은 어느 고등학교에서 기말시험 점수를 분석한 결과이다. 두 과목 점수의 산포를 비교하고자 한다. 이 때 유용하게 이용되는 측도는? (3점)

과목	표본평균	표본표준편차
수학	56	13.3
국어	78	19.4

- ① 표본평균 ② 표본표준편차  
 ③ 표준오차 ④ 변이계수

41. 다음 설명 중 옳은 것은? (3점)

- ① 50백분위수는 25백분위수의 두 배이다.  
 ② 다섯숫자요약은 최소값, 중앙값, 평균, 표준편차, 최대값이다.  
 ③ 상자그림에서 상자의 길이는 사분위수범위(IQR)와 같다.  
 ④ 자료 {3, 5, 3, 8, 6}에서 평균, 중앙값, 최빈값은 모두 같다.

42. 다음은 상자그림에 대한 설명이다. 다음 설명 중에서 올바른 것끼리 묶은 것은? (3점)

- I. 데이터에 대한 다섯숫자요약을 이용하여 작성한다.  
 II. 그래프를 통해서 데이터 상의 군집의 존재 여부를 쉽게 살펴볼 수 있다.  
 III. 여러 집단의 데이터 분포를 비교하는 데 유용하다.

- ① I, II ② I, III  
 ③ II, III ④ I, II, III

43. 다음 설명 중 옳은 것을 모두 고른 것은? (3점)

- I. 표준편차는 사분위수 범위에 비해서 특이치(outlier)의 영향을 적게 받는다.  
 II. 중앙값은 평균에 비해서 특이치(outlier)의 영향을 적게 받는다.  
 III. 표준편차와 사분위수 범위는 분포의 퍼진 정도에 대한 측도이다.

- ① I, II ② II  
 ③ II, III ④ I, II, III

44. 다음 중 연속형 자료의 분포를 살펴보고자 할 때 사용할 수 있는 그래프로 짝 지어진 것은? (2점)

- ① 상자그림, 원형 차트  
 ② 줄기-잎 그림, 히스토그램  
 ③ 상자그림, 막대그림  
 ④ 줄기-잎 그림, 시계열그림

45. 다음은 한글 2014 사용법에 대한 설명이다. 옳은 설명은 모두 몇 개인가? (3점)

- I. [F9] 키를 이용하면 [입력]-[한자로 변화]을 대신할 수 있다.  
 II. 신문이나 잡지 등에서 사용되고 있는 다단 편집의 기능을 제공하고 있다.  
 III. 문서를 작성하면서 그림을 삽입할 수 있다.

- ① 없음 ② 1개  
 ③ 2개 ④ 3개

46. 한글 2014의 수식편집기창을 이용하여 다음 수식을 작성하고자 한다.

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{아닐 때} \end{cases}$$

수식편집창의 다음 ( )안의 a, b에 적합한 것은? (3점)

$$f(x) = (a) \{ 1, \& 0 \leq x \leq 1 (b) 0, \& \text{아닐 때} \}$$

- ① a=ver, b=& ② a=cases, b=&  
 ③ a=ver, b=# ④ a=cases, b=#

※ (47~48) 다음 엑셀 워크시트 화면에 대해서 다음 물음에 답하시오.

N17					
	A	B	C	D	E
1	번호	통계학	수학		
2	1	71	65		
3	2	63	58		
4	3	74	92		
5	4	84	88		
6					

47. D2 셀에 =B2+\$C\$2를 입력한 후 드래그&amp;드롭 기능을 이용하여 D5 셀까지 채워 넣었다. 이 때 D5 셀의 결과는? (3점)

- ① 136 ② 149  
 ③ 172 ④ 오류 발생

48. D5 셀에 =IF(SUM(B5:C5)&lt;150, "불합격", "합격")를 입력하였을 때 그 결과는? (3점)

- ① 152 ② 76  
 ③ 합격 ④ 불합격

49. 함수  $f(x) = -2x^3 - 2x^2 + x + 7$ 의 개형을 (-3, 3) 범위에서 그리고자 한다. 차트마법사의 차트 종류 입력상자에서 선택할 차트 종류로 알맞은 것은? (2점)

- ① 세로 막대형이나 가로 막대형  
 ② 영역형  
 ③ 분산형  
 ④ 방사형

50. 복리로 계산될 때 연이율  $r$ , 기간이  $n$ 인 경우 현재 금액  $p$ 에 대한 일정 기간 후의 원리합계는  $G = p \times (1 + r)^n$  로 계산할 수 있다. 원금이 1,000만원인 경우에 연 2.5%와 연 3.0%로 향후 20년까지 경과하였을 때 원리합계가 얼마인지 계산하고자 한다. 채워넣기를 이용해서 계산한다고 할 때 B5 셀에 알맞은 함수식은? (3점)

	A	B	C	D
1	원금	10,000,000		
2	연이율	2.50%	3.00%	
3				
4	경과년수	원리합계1	원리합계2	
5	1			
6	2			
7	3			
8	4			
9	5			
10	6			
11	7			

- ①  $=BS\$1*(1+B\$2)^A\$5$   
 ②  $=BS\$1*(1+B\$2)^A\$5$   
 ③  $=BS\$1*(1+B\$2)^A\$5$   
 ④  $=BS\$1*(1+B\$2)^A\$5$

51. 유권자 500명을 대상으로 이번 대통령선거에서 어느 후보자를 지지하는지 조사하였다. 그 결과를 분석하여 각 후보자별 지지율을 그래프로 나타내고자 한다. 적합한 그래프끼리 묶인 것은? (3점)

구분	응답자 수	지지율
A 후보	210	42.0%
B 후보	123	24.6%
C 후보	105	21.0%
D 후보	62	12.4%
합계	500	100%

- ① 막대그래프, 원그래프  
 ② 히스토그램, 원그래프  
 ③ 막대그래프, 분산형 차트  
 ④ 원그래프, 분산형 차트

52. 다음은 엑셀의 차트 기능에 대한 설명이다. 옳은 것을 모두 고른 것은? (3점)

- I. 대통령 선거에서 각 후보의 지지율을 원그래프로 나타내면 유용하다.  
 II. 두 변수 사이의 산점도는 차트 유형 중 방사형을 선택하여 그린다.  
 III. 히스토그램은 차트 마법사의 기본 유형으로 제공되고 있다.

- ① I  
 ② I, II  
 ③ II, III  
 ④ I, II, III

53. 다음은 데이터를 불러와 분석하는 R 프로그램의 일부이다. 다음 설명 중에서 옳지 않은 것은? (3점)

```
> score <- scan("c:/data/t3.txt")
Read 20 items
> stem(score)
```

- ① "c:/data/t3.txt"에 저장되어 있는 데이터를 읽어서 변수명을 score로 하였다.  
 ② 데이터는 20개의 관측치로 구성되어 있다.  
 ③ stem(score) 명령에 의하여 상자그림을 얻을 수 있다.  
 ④ stem(score) 명령은 연속형 자료의 분포 모양을 파악하고자 할 때 사용된다.

54. 다음은 상관관계수에 대한 설명이다. 옳바른 것을 모두 고른 것은? (3점)

- I. 상관관계수는 0과 1사이의 값을 갖는다.  
 II. 상관관계수는 두 변수 사이의 선형 연관성의 정도를 나타낸다.  
 III. 상관관계수가 0의 근처일수록 두 변수 사이의 연관성이 없다는 뜻이다.

- ① I, II  
 ② II  
 ③ II, III  
 ④ I, II, III

55. 다음과 같은 텍스트 데이터 파일을 읽고자 할 때 괄호 안에 알맞은 R 명령은? (3점)

```
ex.data = ( )("c:/data/ex8-1.txt", header=T)
```

ex8-1 - 메모장

파일(F)	편집(E)	서식(O)	보기(V)	도움말(H)
id	sex	age	edu	salary
1	1	21	2	150
2	2	22	1	100
3	1	33	2	200
4	2	33	3	220
5	1	28	2	170
6	1	41	3	300
7	2	39	2	290
8	1	32	3	220
9	2	44	1	370
10	1	55	3	410

- ① scan  
 ② read.table  
 ③ data.frame  
 ④ read.xlsx

56. 읽은 데이터 파일 ex.data에 대해서 다음과 같은 R 명령의 결과는 무엇인가? (3점)

```
> head(ex.data)
```

- ① 데이터세트의 처음 6개 케이스를 출력한다.  
 ② 데이터세트의 변수명을 출력한다.  
 ③ 전체 데이터세트를 출력한다.  
 ④ 각 변수의 측정척도를 보여준다.

57. 성별(sex)로 salary 변수에 대한 상자그림을 그리고자 한다. R 스크립트에 입력할 명령어는? (3점)

- ① boxt(salary~sex)  
 ② box(sex~salary)  
 ③ boxplot(salary~sex)  
 ④ boxplot(sex~salary)

58. 다음 R 명령 수행결과는? (3점)

```
> x = c(-1, 0, 1)
> y = c(1, 2, 3)
> z = c(4, 5, 6)
> dframe = data.frame(x, y, z)
> dframe[2, 3]
```

- ① 1  
 ② 3  
 ③ 5  
 ④ 6

59. 다음과 같은 R 명령어를 입력한 결과로 알맞은 것은? (3점)

```
> x=(-3:3)
> y=x[x>1]
> sum(y^2)
```

- ① 4  
 ② 9  
 ③ 13  
 ④ 14

60. 다음과 같은 R 명령어를 입력한 결과로 알맞은 것은? (3점)

```
hist(rnorm(100))
```

- ① 평균이 0, 분산이 1인 정규분포로부터 난수 100개를 생성하여 히스토그램 작성  
 ② 평균이 0, 분산이 1인 정규분포 함수를 그래프로 나타냄  
 ③ 평균이 0, 분산이 1인 정규분포로부터 난수 100개를 생성하여 줄기-잎 그림 작성  
 ④ 평균이 100, 분산이 1인 정규분포에 대한 히스토그램 작성