

2과목 데이터정보처리입문 (36~60)

출제위원 : 방송대 이기재

출제범위 : 교재 전범위(웹강의 포함)

36. 일반적인 데이터분석 절차이다. ()안에 순서대로 가장 적합한 것은? (2점)

(a) - 조사, 실험의 계획 - 데이터의 수집 - (b) - 분석 결과의 평가

- ① a: 문제의 정의 b: 설문지 평가
- ② a: 설문지 작성 b: 데이터의 정리, 분석
- ③ a: 전문가와의 상담 b: 데이터의 분석
- ④ a: 문제의 정의 b: 데이터의 정리, 분석

37. 다음은 측정 수준에 대한 설명이다. 옳은 것을 모두 고른 것은? (3점)

- I. 순서척도에서 각 조사단위에 부여된 숫자는 구분을 목적으로 부여된 기호에 불과하다.
- II. 점수, 순위, 습도, 지능지수 등은 비율척도로 측정된 값이다.
- III. 구간척도에서 0값은 자의적으로 부여되었으므로 절대적 의미를 가질 수 없다.

- ① I ② II
- ③ III ④ II, III

38. 우리 대학교에 재학 중인 학생들을 대상으로 통계조사를 실시하였다. 다음에 제시된 변수를 조사하였는데, 이 중 명목척도로 측정된 변수는 몇 개인가? (3점)

- 소속 학과
- 나이
- 성별(남, 여)
- 소속 지역대학
- 학교 교과과정 만족도(아주 만족, 만족, 보통, 불만족, 아주 불만족)

- ① 1개 ② 2개
- ③ 3개 ④ 4개

※ (39~40) 다음 줄기-잎 그림을 보고 물음에 답하시오.

```

3 | 0
4 | 138
5 | 122379
6 | 147
7 | 47
    
```

39. 줄기-잎 그림으로 주어진 데이터에 대한 중앙값과 최빈값(mode)은 각각 얼마인가? (3점)

- ① 51, 52
- ② 51, 53
- ③ 52, 52
- ④ 53, 52

40. 주어진 데이터에 대한 범위는 얼마인가? (2점)

- ① 77
- ② 47
- ③ 44
- ④ 알 수 없음

41. 다음은 어느 지역에서 임의로 추출한 납세자 7명의 연간 소득 자료(단위: 100만원)이다.

45, 35, 50, 60, 23, 32, 850

위의 값 중 850은 85가 잘못 입력된 것이라고 밝혀졌다. 850을 85로 바꾸더라도 그 값이 변하지 않는 것은? (3점)

- ① 표본평균
- ② 중앙값
- ③ 범위
- ④ 표본표준편차

42. 다음 설명 중 옳은 것을 모두 고른 것은? (3점)

- I. 중앙값은 50% 백분위수이다.
- II. 다섯숫자요약은 최소값, 제1사분위수, 평균, 제3사분위수, 최대값이다.
- III. 상자그림에서 상자의 길이는 사분위수범위(IQR)와 같다.

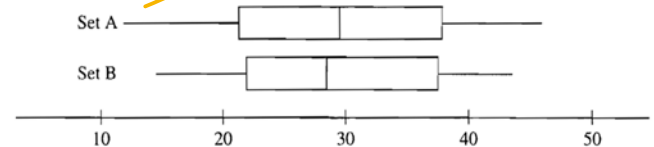
- ① I, II
- ② I, III
- ③ II, III
- ④ III

43. 다음은 변이계수(coefficient of variation)에 대한 설명이다. 옳은 것을 모두 고른 것은? (3점)

- I. 성인 집단과 신생아 집단의 체중 데이터에 대한 산포 정도를 비교하고자 할 때 사용할 수 있다.
- II. 표준편차를 평균으로 나누어 구한다.
- III. 계산이 간편하고, 극단값의 영향을 거의 받지 않는다.

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

44. 다음은 데이터 A와 B를 분석하여 요약한 상자그림이다. 다음 설명 중에서 옳은 것을 모두 고른 것은? (3점)



- I. 데이터 A의 중앙값은 데이터 B의 중앙값보다 크다.
- II. 데이터 A의 제1사분위수는 데이터 B의 제1사분위수보다 작다.
- III. 데이터 A의 사분위수 범위는 데이터 B의 사분위수 범위보다 작다.

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

45. 다음은 한글 2014 사용법에 대한 설명이다. 옳은 설명을 모두 고른 것은? (3점)

- I. 복사나 이동을 위한 블록 지정에서 F3은 칸 단위 블록 설정, F4는 줄 단위 블록을 지정할 수 있다.
- II. [F9] 키를 이용하면 [입력]-[한자로 변화]을 대신할 수 있다.
- III. 문서를 작성하면서 그림을 삽입할 수 있다.

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

46. 수식편집기창을 이용하여 다음 수식을 작성하고자 한다.

$$\chi^2 = \sum \frac{(\text{관측도수} - \text{기대도수})^2}{\text{기대도수}}$$

$$= \sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

수식편집창의 다음 ()안의 ㉠, ㉡, ㉢에 적합한 것은? (3점)

(㉠) ~ (㉡) = ~ sum { (관측도수 - 기대도수)^2 } (㉢) 기대도수 #
(㉡) = ~ sum { (f_{ij}) - {hat f}_{ij})^2 } (㉢) {hat f}_{ij}

- ① ㉠: chi, ㉡: &, ㉢: vert
② ㉠: chi, ㉡: &, ㉢: over
③ ㉠: theta, ㉡: #, ㉢: vert
④ ㉠: theta, ㉡: &, ㉢: over

※ (47~49) 다음은 엑셀 워크시트 화면이다. 다음 물음에 답하시오

	D2				
	A	B	C	D	E
1	번호	통계학	수학		
2	1	71	65		
3	2	63	58		
4	3	74	88		
5	4	65	87		
6	5	86	92		
7	6	87	80		
8	7	84	79		
9	8	77	70		
10	9	85	77		
11	10	59	45		
12	과목평균				
13					

47. D2 셀에 =B\$2+C2를 입력한 후 드래그&드롭 기능을 이용하여 D11 셀까지 채워 넣었다. 이 때 D6 셀의 결과는? (2점)

- ① 오류 발생
② 알 수 없음
③ 136
④ 163

48. E7 셀에 =sum(B6, C7)라고 입력하였을 때 그 결과값은 얼마인가? (2점)

- ① 166
② 167
③ 오류 발생
④ 345

49. D2 셀에 =IF(AVERAGE(B2:C2)<70, "불합격", "합격")를 입력한 후 드래그&드롭 기능을 이용하여 D11 셀까지 채워 넣었다. 이 때 D10 셀의 결과는? (3점)

- ① 81
② 162
③ 합격
④ 불합격

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

- I. 대통령 선거에서 각 후보의 지지율은 원형 차트로 나타내면 유용하다.
II. 두 변수 사이의 산점도를 그리고자 한다면 차트 유형 중 분산형을 선택한다.
III. 히스토그램은 차트 마법사의 기본 유형으로 제공된다.

- ① I
② I, II
③ II, III
④ I, II, III

51. 유권자 500명을 대상으로 정당별 지지도를 조사하였다. 그 결과를 분석하여 각 정당별 지지율을 그래프로 나타내고자 한다. 적합한 그래프끼리 묶인 것은? (3점)

구분	응답자 수	지지율
A 당	230	46.0%
B 당	123	24.6%
C 당	105	21.0%
기타	32	6.4%
합계	500	100%

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

52. 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 정규분포로부터 난수 100개를 생성하여 상자그림 작성
③ 평균이 0, 분산이 1인 정규분포로부터 난수 100개를 생성하여 줄기-잎 그림 작성
④ 평균이 100, 분산이 1인 정규분포에 대한 상자그림 작성

53. 다음은 R 프로그램의 일부이다. 이 프로그램을 통해서 얻는 분석 결과는? (3점)

```
> math <- c(66, 64, 48, 78, 60, 90, 50, 66, 70)
> physics <- c(70, 69, 46, 84, 64, 92, 52, 68, 72)
> plot(math, physics)
```

- ① 두 변수 math와 physics 사이의 상관계수를 구한다.
② 두 변수 math와 physics 사이의 산점도를 그린다.
③ 두 변수 math와 physics에 대한 줄기-잎 그림을 그린다.
④ 두 변수 math와 physics 사이의 단순회귀식을 적합한다.

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

- I. 상관계수는 -1과 1사이의 값을 갖는다.
II. 상관계수는 두 변수 사이의 선형 연관성의 정도를 나타낸다.
III. 상관계수가 -1에 가까워 갈수록 두 변수 사이의 연관성이 없다는 뜻이다.

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

※ (55~57) 다음과 같이 "c:/data/ex_lkj.txt"에 저장되어 있는 텍스트 파일을 읽어 분석하고자 한다. 다음 물음에 답하여라.

ex_lkj - 메모장				
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)				
id	sex	height	weight	
1	1	184.0	76.4	
2	2	160.3	57.2	
3	1	179.3	74.2	
4	1	176.2	68.2	
5	2	156.4	51.4	
6	2	162.1	54.2	
7	1	176.2	68.2	
8	2	158.1	52.4	
9	2	165.2	61.2	
10	1	186.0	73.4	

55. 위의 텍스트 파일을 읽어 통계분석하고자 한다. 괄호 안에 알맞은 R 명령은? (3점)

```
phy.data = ( )("c:/data/ex_lkj.txt", header=T)
```

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

56. 읽은 데이터 파일 phy.data에 대해서 다음과 같은 R 명령의 결과는 무엇인가? (2점)

```
> head(phy.data)
```

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

57. attach(phy.data) 명령을 입력한 후 그룹변수인 성별(sex)의 값에 따라 관심변수 height의 표준편차(standard deviation)를 구하는 R 명령은? (3점)

- ① sapply(height, sex, mean)
- ② sapply(mean, sex, height)
- ③ tapply(height, sex, sd)
- ④ tapply(sd, sex, height)

58. attach(phy.data) 명령을 입력한 후 성별(sex)에 따라 weight에 대한 상자그림을 그리고자 한다. R 스크립트에 입력할 명령어는? (3점)

- ① box(weight ~ sex)
- ② stem(weight ~ sex)
- ③ boxplot(sex ~ weight)
- ④ boxplot(weight ~ sex)

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

```
> x=2*(-4:4)
> y=x[x>2]
> mean=mean(y)
> sum((y-mean)^2)
```

- ① 0
- ② 5
- ③ 8
- ④ 10

60. 다음은 줄기-잎 그림에 대한 설명이다. 옳지 않은 것은? (3점)

- ① 연속형 자료의 분포를 살펴보고자 할 때 유용하다.
- ② 그래프를 통해서 데이터의 전반적인 분포와 함께 구체적인 자료값도 알 수 있다.
- ③ 그래프를 통해서 관측치의 범위, 분포의 모양, 집중도 등을 쉽게 알 수 있다.
- ④ 관측치가 많을 때 여러 개의 계급 구간을 나누어 도수분포표를 작성한 후에 그린다.

②
SKIP 1

$2 \times (-4, -2, 0, 2, 4)$
 $(-8, -4, 0, 4, 8)$

$-8, -4, -2, 0, 2, 4, 6, 8$
 $468 > 2$

mean = 6

4 68

-6

$(-2, 0, -2) \times 2$
 $\sum(404) = 8$