

2과목 데이터정보처리입문 (36~60)

출제위원 : 방송대 이기재

출제범위 : 교재 전범위(웹강의 포함)

36. 데이터의 구분에 대한 설명이다. 다음 괄호 안에 알맞은 용어로 짝지어진 것은? (2점)

데이터는 측정수준에 따라서 질적 데이터와 양적 데이터로 구분된다. 여기에서 양적 데이터란 (a)나 (b)에 의해서 측정된 데이터를 말한다.

- ① 명목척도, 순서척도
 ② 명목척도, 구간척도
 ③ 순서척도, 구간척도
 ④ 구간척도, 비율척도

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

(a)-조사, 실험의 계획-데이터의 수집-(b)-분석결과와 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

38. 어느 회사에서 200명의 직원을 표본으로 추출하여 이름, 성별, 나이, 교육수준, 직급, 직무만족도(아주 만족, 만족, 보통, 불만족, 아주 불만족)를 조사하여 데이터로 정리하였다. 다음의 설명 중 옳은 것은 모두 몇 개인가? (3점)

- I. 정리된 데이터는 6개 변수에 대해서 200개 케이스로 이루어졌다.
 II. 데이터에서 직무만족도는 구간척도로 측정되었다.
 III. 성별이나 나이는 명목척도로 측정되었다.

- ① 없음 ② 1개
 ③ 2개 ④ 3개

39. 다음 자료의 평균과 중앙값으로 옳바른 것은? (3점)

1, 7, 3, 3, 6, 4

- ① 4, 3 ② 4.8, 3.5
 ③ 4, 3.5 ④ 4.8, 3

40. 초등학교 5학년생 중 일부를 뽑아서 체중을 조사하여 얻은 데이터에 대한 줄기-잎 그림은 다음과 같다. 이 자료의 사분위수 범위(Inter-quartile range)에 가장 가까운 것은? (3점)

2 | 679
 3 | 233457
 4 | 13345789
 5 | 36
 6 | 1

- ① 35 ② 32
 ③ 15 ④ 10

41. 탐색적 자료분석의 관점에서 살펴볼 때 한 묶음의 자료를 정리하는 숫자로서 다섯숫자요약이란 다음 중 무엇인가? (2점)

- ① 중앙값, 평균, 분산, 제1사분위수, 제3사분위수
 ② 중앙값, 평균, 표준편차, 제1사분위수, 제3사분위수
 ③ 최소값, 최대값, 중앙값, 제1사분위수, 제3사분위수
 ④ 최소값, 최대값, 평균, 제1사분위수, 제3사분위수

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

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

- ① I ② I, II
 ③ I, III ④ I, II, III

43. 다음 설명 중 옳은 것을 모두 고른 것은? (3점)

- I. 사분위수 범위는 범위에 비해서 특이치(outlier)의 영향을 적게 받는다.
 II. 중앙값은 평균에 비해서 특이치(outlier)의 영향을 적게 받는다.
 III. 제3사분위수와 최빈값은 분포의 퍼진 정도에 대한 척도이다.

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

44. 한글 2014에서 다음의 수식에 의한 결과로 알맞은 것은? (3점)

$\frac{\{\bar{X} - \mu\}}{\{\sigma/\sqrt{n}\}} \sim N(0, 1)$

- ① $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ② $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
 ③ $\bar{X} \sim N\left(\mu, \frac{\sigma}{n}\right)$ ④ $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

45. 다음은 한글 2014 사용법에 대한 설명이다. 옳은 설명은 모두 몇 개인가? (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 변환]을 대신할 수 있다.
 III. 신문이나 잡지 등에서 사용되고 있는 단 편집의 기능을 제공하고 있다.

- ① 없음 ② 1개
 ③ 2개 ④ 3개

※ (46~47) 다음 엑셀 워크시트 화면에 대해서 다음 물음에 답하시오.

	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					

46. D2 셀에 =\$B\$2+C2를 입력한 후 드래그&드롭 기능을 이용하여 D5 셀까지 채워 넣었다. 이 때 D5 셀의 결과는? (3점)

- ① 136 ② 159
 ③ 오류 발생 ④ 알 수 없음

47. D6 셀에 =sum(B2:B4, C2)라고 입력하였을 때 그 결과값은 얼마인가? (3점)

- ① 236 ② 266
 ③ 오류 발생 ④ 273

48. 다음 중 키, 몸무게, 성적 등의 연속형 자료에 대한 전반적인 분포를 살펴보고자 할 때 사용할 수 있는 그래프는 모두 몇 개인가? (3점)

- I. 히스토그램 II. 원형 차트
 III. 상자그림 IV. 분산형 차트

- ① 1개 ② 2개
 ③ 3개 ④ 4개 모두 옳음

49. 엑셀에서 데이터를 입력한 후 이웃하지 않은 셀들을 선택하고자 한다. 올바른 방법은? (2점)

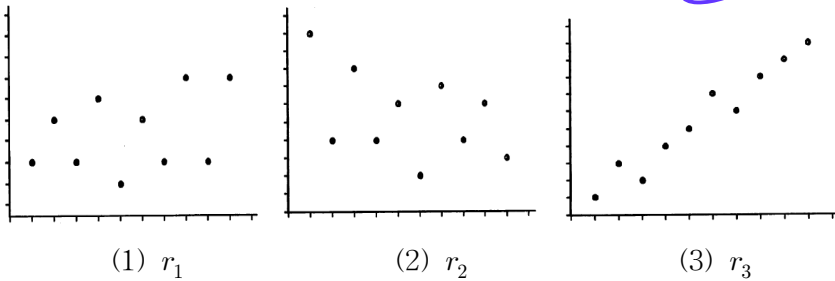
- ① CTRL 키를 누른 상태에서 마우스로 원하는 셀들을 선택한다.
- ② TAB 키를 누른 상태에서 마우스로 원하는 셀들을 선택한다.
- ③ SHIFT 키를 누른 상태에서 마우스로 원하는 셀들을 선택한다.
- ④ ALT 키를 누른 상태에서 마우스로 원하는 셀들을 선택한다.

50. 다음과 같은 지지 정당별 도수분포를 그래프로 나타내고자 한다. 적합한 그래프끼리 묶인 것은? (3점)

구분	도수	백분율
A당	19	38.0%
B당	17	34.0%
C당	14	28.0%
합계	50	100%

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

51. 두 변수 사이의 산점도가 다음과 같이 주어졌을 때 상관계수를 크기 순으로 바르게 나열한 것은? (3점)



- ① $r_1 < r_2 < r_3$
- ② $r_2 < r_1 < r_3$
- ③ $r_2 < r_3 < r_1$
- ④ $r_3 < r_2 < r_1$

52. 두 변수들 사이의 관계를 살펴보기 위해 산점도를 그리고자 한다. 차트마법사의 차트 종류 입력상자에서 선택할 차트 종류는? (2점)

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

53. 일반적으로 축전지의 전압은 시간이 지남에 따라 관계식 $V = 10e^{-0.5t}$ 에 따라 떨어진다고 한다. 엑셀 수식을 이용하여 시간을 0부터 10초까지 변화시키면서 전압을 계산하고자 한다. B2 셀에 알맞은 수식을 입력한 후 채우기 기능을 이용하여 계산하고자 한다. 알맞은 함수식은? (3점)

	A	B	C
1	시간(t)	전압	
2	0	10.000	
3	1	6.065	
4	2	3.679	
5	3	2.231	
6	4	1.353	
7	5	0.821	
8	6	0.498	
9	7	0.302	
10	8	0.183	
11	9	0.111	
12	10	0.067	

- ① $=10*EXP(-0.5*A2)$
- ② $=10*EXP(-0.5*$A$2)$
- ③ $=10*LOG(-0.5*A2)$
- ④ $=10*E(-0.5*$A$2)$

54. 다음은 상관계수(correlation coefficient)에 대한 설명이다. 옳은 것을 모두 고른 것은? (3점)

- I. 두 변수의 상관계수가 0 근처의 값을 나타낼 때 상관관계가 없다고 한다.
- II. 상관계수는 0과 1사이의 값을 갖는다.
- III. 두 변수 간에는 강한 관계가 존재함에도 불구하고 상관계수가 0에 가깝게 나올 수 있다.

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

※ (55~59) 다음과 같이 "c:/data/ex_lkj.txt"에 저장되어 있는 텍스트 파일을 읽어 분석하고자 한다. 다음 물음에 답하여라.

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. 앞서 읽어 온 데이터 객체 phy.data의 변수들을 R에서 직접 사용하고자 한다. 유용한 명령은? (3점)

- ① load(phy.data)
- ② detach(phy.data)
- ③ unload(phy.data)
- ④ attach(phy.data)

57. 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. 성별(sex) 구분에 따라 height에 대한 상자그림을 그리고자 한다. R 스크립트에 입력할 명령어는? (3점)

- ① boxt(phy.data\$height ~ phy.data\$sex)
- ② box(phy.data\$sex ~ phy.data\$height)
- ③ boxplot(phy.data\$height ~ phy.data\$sex)
- ④ boxplot(phy.data\$sex ~ phy.data\$height)

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

```
> x=(-2:2)
> sum(x^2)
```

- ① 0
- ② 1
- ③ 5
- ④ 10

60. 평균이 0, 분산이 1인 정규분포로부터 난수 100개를 생성하여 히스토그램을 작성하고자 할 때 알맞은 명령은? (3점)

- ① hist(pnorm(100))
- ② stem(pnorm(100))
- ③ hist(rnorm(100))
- ④ stem(rnorm(100))