

# 빅데이터의 이해, 시물레이션

2017학년도 2 학기

4 학년 2 교시

※ 정답 하나만을 골라 반드시 컴퓨터용 사인펜으로 OMR 답안지에 표기할 것.	학 과		감독관	인
	학 번	-	성 명	

1과목	빅 데이터의 이해 (1~25)
출제위원 : 방송대 이궁희	
출제범위 : 교재 1~10장 (R 프로그램 실습 제외)	

1. 빅데이터의 3V에 해당하는 것을 모두 고르시오. (4점)

- ㉠ 데이터의 가격
- ㉡ 데이터의 다양성
- ㉢ 데이터의 정확도
- ㉣ 데이터의 규모
- ㉤ 데이터의 속도

- ① ㉠, ㉡, ㉢
- ② ㉡, ㉢, ㉣
- ③ ㉡, ㉣, ㉤
- ④ ㉢, ㉣, ㉤

2. 1 페타바이트(PB)는 몇 메가바이트(MB)인가? (2점)

- ①  $10^3$ MB
- ②  $10^6$ MB
- ③  $10^9$ MB
- ④  $10^{12}$ MB

3. 다음 중 정형 데이터로 가장 바른 것은? (3점)

- ① 페이스북의 사진
- ② 유튜브 동영상
- ③ 신문기사
- ④ 소매점의 거래데이터

4. “데이터 과학은 통계학, ( )과 관련 주제 전문지식이 종합된 분야이다.”에서 ( )에 가장 적합한 단어는? (4점)

- ① 수리능력
- ② 해킹기술
- ③ 인지력
- ④ 경영기술

5. 다음 중 데이터과학자가 가져야할 하드스킬로 가장 바른 것은? (4점)

- ① 데이터를 처리, 분석할 수 있는 R프로그램 작성 능력
- ② 데이터의 특성을 설명, 전달할 수 있는 스토리텔링 능력
- ③ 데이터에 있는 가치를 발견할 수 있는 통찰력
- ④ 데이터 관련 다른 전문가와 소통, 협력할 수 있는 능력

6. 기업이 빅데이터를 의사결정에 활용하는 수준은 4가지 단계로 구분된다. 이중 최선의 해결책을 찾는 단계로 가장 적합한 것은? (3점)

- ① 비즈니스 인텔리전스 단계
- ② 통찰력의 창출단계
- ③ 애널리틱스 기반 의사결정 단계
- ④ 정보의 관리 단계

7. 필립 코틀러 등이 말하는 디지털 시대의 마케팅 트렌드로 가장 적당한 것은? (3점)

- ① 고객의 취향과 요구를 감지, 대응하는데 주력하는 마케팅
- ② 물건을 만드는데 주력하는 마케팅
- ③ 새로운 고객을 확보하는데 주력하는 마케팅
- ④ 단일 채널을 통해 고객을 상대하는 마케팅

8. 분산 시스템에 대한 설명 중 바른 것을 모두 고르시오. (4점)

- ㉠ 컴퓨터를 네트워크에 한 대 추가 연결하는 것으로 전체 분산 시스템의 성능을 간단히 향상할 수 있다.
- ㉡ 저성능 서버를 여러 대 연결하여 만든 분산시스템이 비슷한 성능의 고성능 서버 한 대보다 저렴하다.
- ㉢ 분산 시스템에서는 연결된 컴퓨터들 중 한두 대가 고장을 일으킨다고 해도 전체 시스템이 고장 나지 않는다.

- ① ㉠
- ② ㉡, ㉢
- ③ ㉠, ㉡
- ④ ㉠, ㉡, ㉢

9. 하둡 분산파일시스템은 일반적으로 ( ㉠ )와 ( ㉡ )로 구성되어 있다. ㉠과 ㉡를 가장 바르게 나열한 것은? (3점)

- ① 마스터노드, 네임노드
- ② 베이직노드, 네임노드
- ③ 마스터노드, 슬레이브노드
- ④ 데이터노드, 슬레이브노드

10. HTML이나 XML과 같은 특정한 규칙에 근거하여 만들어진 데이터를 분해하고, 사용자가 정한 저장 규칙에 따라 분해한 데이터를 저장하는 것을 무엇이라 부르는가? (2점)

- ① 룰링(ruling)
- ② 맵핑(mapping)
- ③ 파싱(parsing)
- ④ 로케이팅(locating)

11. 데이터 시각화의 전문가 Edgar Tufte는 데이터 시각화의 8가지 원칙을 세웠는데 다음 중에서 이에 해당하는 것을 모두 고르시오. (4점)

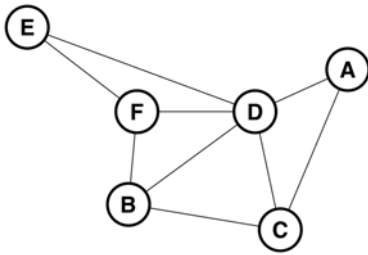
- ㉠ 데이터시각화에서 데이터 그 자체를 보여주는 것이 중요하다.
- ㉡ 데이터시각화에서 서로 다른 데이터를 손쉽게 비교할 수 있게 한다.
- ㉢ 데이터시각화에서 많은 양의 데이터가 일관성 있어야 한다.

- ① ㉠, ㉢
- ② ㉠, ㉡
- ③ ㉡, ㉢
- ④ ㉠, ㉡, ㉢

12. 다음 중 자바 기반 미디어 아트를 위한 시각화도구는? (2점)

- ① R
- ② 타블로
- ③ Excel
- ④ 프로세싱

※ (13~14) 다음 소셜네트워크 그래프에 대해 답하시오.



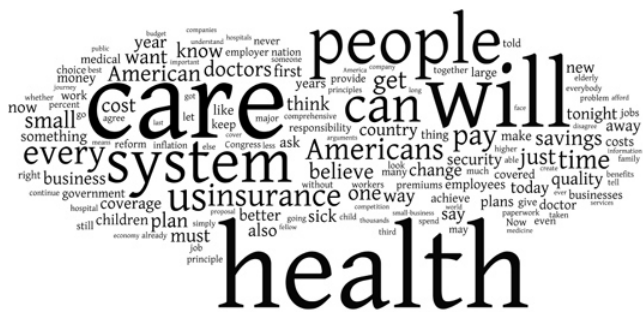
13. 노드 F의 디그리(degree)는? (2점)

- ① 1                      ② 3  
③ 5                      ④ 7

14. 소셜네트워크의 중심으로 가장 적당한 노드는? (2점)

- ① A                      ② C  
③ D                      ④ F

15. 다음과 같은 그래프를 무엇이라 부르는가? (3점)



- ① 태그 클라우드                      ② 버블 그래프  
③ 시간 그래프                        ④ 워드 트리

16. 브레이먼(Breiman)이 설명한 설명변수가 매우 많은 고차원 회귀 모형에서 변수선택방법을 적용할 때 발생하는 현상을 가장 바르게 기술된 것은? (3점)

- ① 추정결과가 안정적이거나 예측력이 현저히 저하된다.
- ② 추정결과가 안정적이고 예측력이 현저히 개선된다.
- ③ 추정결과가 불안정적이고 예측력이 현저히 저하된다.
- ④ 추정결과가 불안정적이지만 예측력이 현저히 개선된다.

17. 라쏘(lasso) 추정량에 대한 설명 중 가장 바르게 기술된 것은?  
(2점)

- ① 별점화 최소제곱추정량이고 성긴 성질을 가지지 않는다.
- ② 별점화 최소제곱추정량이고 성긴 성질을 가진다.
- ③ 일반 최소제곱추정량이고 성긴 성질을 가지지 않는다.
- ④ 일반 최소제곱추정량이고 성긴 성질을 가진다.

18. 배킹방법에 대한 설명 중 가장 바른 것은? (3점)

- ① 주어진 데이터에 대하여 여러 개의 붓스트랩 데이터를 생성하고 이를 이용한 여러 개의 모형을 만든 후 이들을 결합하는 방법
- ② 주어진 데이터에 대하여 한 개의 붓스트랩 데이터를 생성하고 이를 이용한 모형을 만드는 방법
- ③ 주어진 데이터에 대하여 여러 개의 붓스트랩 데이터를 생성하는 방법
- ④ 주어진 데이터로부터 다양한 변수를 조합하여 여러 개의 모형을 만든 후 이들을 결합하는 방법

19. 학습기에 사용되는 설명변수들 중 일부를 골라내 그것의 영향력을 시각화하는 방법은? (2점)

- ① 갭마인드 그래프
- ② 부분의존도 도표
- ③ 협력 스케일 그래프
- ④ 상관도표

20. 확률모형을 가정하지 않은 군집분석의 경우 유사성의 측도가 필요한데 그 유사성의 측도로 올바르지 **않는** 것은 무엇인가?  
(2점)

- ① 유클리디언 거리
- ② 앙상블 거리
- ③ cosine 유사성
- ④ Jaccard 계수

※ (21~22) 편의점의 거래내역을 보고 물음에 답하시오.

고객 번호	품목
1	빵, 라면
2	우유, 빵, 식기세척제
3	식기세척제, 세제
4	빵, 세제, 라면

21. “뽕을 구매하면 라면을 구매한다.”의 지지도는? (2점)

- ①  $\frac{1}{5}$                       ②  $\frac{1}{4}$   
 ③  $\frac{1}{3}$                       ④  $\frac{1}{2}$

22. “빵을 구매하면 라면을 구매한다.”의 신뢰도는? (3점)

- $$\begin{array}{ll} \textcircled{1} \quad \frac{1}{5} & \textcircled{2} \quad \frac{1}{4} \\ \textcircled{3} \quad \frac{2}{3} & \textcircled{4} \quad \frac{3}{4} \end{array}$$

23. 주어진 고객과 상품들에 대한 선호도가 비슷한 고객을 조사하고, 선호도가 비슷한 고객들이 좋아하는 상품 중 주어진 고객이 모르고 있는 상품을 추천하는 것은? (3점)

- ① 랜덤 포레스트
- ② 협력적 정화방법
- ③ 선형적 알고리즘
- ④ 계층적 추천법

24. 다음 중 가장 프라이버시라고 판단되는 3개를 고르시오. (3점)

㉠ 질병정보	㉡ 정치적 신념
㉢ 직장명	㉣ 범죄경력
㉤ 이름	

- ① 가, 나, 라  
② 나, 다, 마  
③ 가, 나, 다  
④ 나, 다, 라

25. 개인정보보호 기술 중 개인 식별자를 제거하여 개인정보를 숨기는 방법을 고르시오. (2점)

- ① 데이터 마이닝
- ② 범주화 기술
- ③ 등급 분류 기술
- ④ 데이터 마스킹