

- 구간화, 군집화 등의 기법 적용

- ① 집계 ② 일반화
- ③ 평활화 ④ 정규화

2과목 빅데이터 탐색

11. 특정한 데이터값의 변화에 따라 지도의 면적이 왜곡되는 지도로 가장 알맞은 것은?

- ① 코로플레스맵 ② 산점도
- ③ 버블플롯맵 ④ 카토그램

12. 웹상에서 자료를 주고받을 때 사용하는 데이터 포맷으로 자바 스크립트 구문 형식의 언어 독립형 데이터 포맷으로 가장 알맞은 것은?

- ① XML ② JSON
- ③ HTML ④ AJAX

13. 마스터 슬레이브 구조를 가지는 분산형 파일 시스템으로 가장 알맞은 것은?

- ① HDFS ② 맵리듀스
- ③ Avro ④ Hive

14. 데이터 중 최대 데이터 값(Max)가 10이고 최소 데이터값(Min)이 4일 경우 범위(Range)로 가장 알맞은 것은?

- ① 4 ② 6
- ③ 8 ④ 10

15. 독립변수가 2개 이고 종속변수가 1개일 경우

집단 간 종속변수의 평균차이를 분석하는 방법으로 가장 알맞은 것은?

- ① 일원분산 분석 ② 이원분산 분석
- ③ 다변량 분산 분석 ④ 공분산 분석

16. 20개의 전구에서 무작위로 5개의 전구를 추출하는 표본 추출 기법으로 가장 알맞은 것은

- ① Simple Random Sampling
- ② Systematic Sampling
- ③ Stratified Random Sampling
- ④ Cluster Sampling

17. 다음 중 연속확률 분포가 아닌 것은?

- ① 이항분포 ② 정규분포
- ③ Z-분포 ④ F-분포

18. 표본의 개수가 커지면 모집단의 분포와 상관 없이 표본 분포는 정규 분포에 근사한다는 법칙으로 가장 알맞은 것은?

- ① Population
- ② Law Large Number
- ③ Central Limit Theorem
- ④ Standard Deviation

19. 점추정 조건으로 가장 알맞지 않은 것은?

- ① Unbiasedness
- ② Efficiency
- ③ Consistency
- ④ Confidence Level

20. 현재까지 주장되어 온 것이거나 기존과 비교

하여 변화 혹은 차이가 없음을 나타내는 가설로 가장 알맞은 것은?

- ① Level of Significance
- ② p-value
- ③ Alternative Hypothesis
- ④ Null Hypothesis

3과목 빅데이터 모델링

21. 다음은 A마트 판매내역의 일부이다. 커피와 탄산의 지지도는 얼마인가?

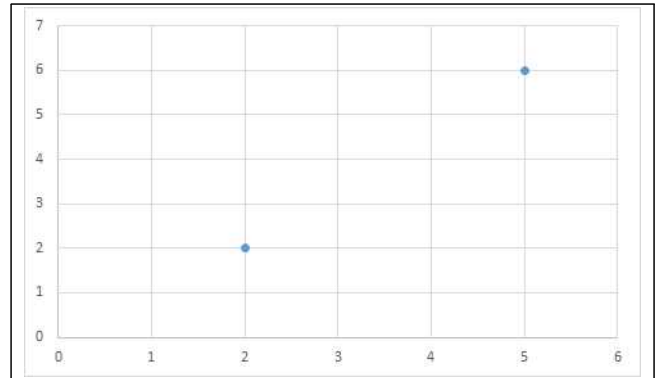
품목	판매 수량
커피	300
탄산	400
커피와 탄산 동시 구매	100
전체 판매량	1000

- ① 0.1 ② 0.8
- ③ 0.3 ④ 0.4

22. 군집간의 거리 측정 방법 중에서 군집 내의 오차 제곱합에 기초한 거리 측정 방법은 무엇인가?

- ① 최장연결법 ② 중심연결법
- ③ 와드연결법 ④ 최단연결법

23. 다음 그림에서 두 점 사이의 맨하튼 거리는 얼마인가?



- ① 5 ② 3
- ③ 4 ④ 7

24. 독립성 검정을 위한 변수의 범주가 아래와 같이 주어졌을 경우 자유도는 얼마인가?

범주 1) 대학교 1, 2, 3, 4학년
범주 2) 선호도 상, 중, 하

- ① 6 ② 7
- ③ 8 ④ 12

25. 주성분 분석에서 주성분의 수로 결정할 수 있는 누적 기여율은 얼마인가?

- ① 75% 이상
- ② 80% 이상
- ③ 85% 이상
- ④ 90% 이상

26. ARIMA(3, 4, 2)모형에서 ARMA 모형으로 정상화를 하려고 할 경우에 필요한 차분의 수는 얼마인가?

- ① 3 ② 4
- ③ 2 ④ 0

27. 시각적 이미지를 분석하는 데 사용되며 심층 신경망으로 합성곱 신경망이라고도 불리고, 기존 영상 처리의 필터 기능과 신경망을 결합하여 성능을 발휘하도록 만든 심층신경망은 무엇인가?

- ① DNN ② RNN
- ③ GAN ④ CNN

28. 다음 중에서 비모수 통계에 대한 설명으로 가장 옳바르지 않은 것은?

- ① 모수적 방법에 비해 통계량의 계산이 간편하고 직관적으로 이해하기 쉽다.
- ② 모집단의 분포에 무관하게 사용할 수 있다.
- ③ 이상값으로 인한 영향이 적다.
- ④ 모수 통계로 검정이 가능한 데이터에 비모수 통계를 이용하면 모수 통계에 비해 효율성이 향상된다.

29. 다음 중 비모수 통계 검정 방법이 아닌 것은 무엇인가?

- ① 부호 검정
- ② 런 검정
- ③ 크루스칼-왈리스 검정
- ④ 피어슨 상관 분석

30. 웹 로그를 통해 사용자의 행위 패턴을 분석하여 의미 있는 정보 추출하는 웹 마이닝 유형은 무엇인가?

- ① 웹 구조 마이닝 ② 웹 내용 마이닝
- ③ 웹 사용 마이닝 ④ 오피니언 마이닝

4과목 빅데이터 결과 해석

31. 다음 중 데이터 시각화 유형과 기법이 잘못 짝지어진 것은?

- ① 분포 시각화 - 파이 차트, 도넛 차트
- ② 관계 시각화 - 산점도, 버블차트
- ③ 공간 시각화 - 등치선도, 도트맵
- ④ 비교 시각화 - 막대그래프, 점그래프

32. 다음 중 두 변수 간의 관계를 시각적으로 나타내는 표현방법은 어떤 기법인가?

- ① 히트맵(Heat Map)
- ② 히스토그램(Histogram)
- ③ 산점도(Scatter plot)
- ④ 평행 좌표 그래프(Parallel Coordinate Plots)

33. 다음이 설명하는 데이터 시각화 유형은 무엇인가?

- 다변량 데이터 사이에 존재하는 변수 사이의 연관성, 분포와 패턴을 찾는 시각화 방법
- 산점도, 버블 차트, 히스토그램 등이 존재

- ① 비교 시각화 ② 관계 시각화
- ③ 시간 시각화 ④ 공간 시각화

34. 다음 중 비즈니스 기여도 평가 지표로 가장 옳바르지 않은 것은?

- ① 투자 대비 효과(ROI)
- ② 목표복구 시간(RTO)
- ③ 순 현재가치(NPV)
- ④ 내부 수익률(IRR)

35. 다음 중 데이터 시각화 분석을 위한 데이터 유형으로 가장 옳지 않은 것은?

- ① 의견, 경험 ② 관계, 연결
- ③ 추세, 패턴 ④ 범주, 비율

36. 다음 중 모델에서 외적인 요소로 데이터 분석을 통해 얻어지는 값이 아닌 사용자가 직접 설정해주는 값은 무엇인가?

- ① 파라미터(Parameter)
- ② 하이퍼 파라미터(Hyper Parameter)
- ③ 모집단(Population)
- ④ 공분산(Covariance)

- 활용사례로는 표본 평균, 다변량 정규분포, 다중 상관계수가 존재

- ① 카이제곱 검정
- ② F-검정
- ③ T-검정
- ④ Z-검정

37. 다음 중 중요 정보를 하나의 그래픽으로 표현해 보는 사람들이 쉽게 정보를 이해할 수 있게 하는 시각화 방법은 무엇인가?

- ① 인포그래픽(Infographic)
- ② 타이포그래피(Typography)
- ③ 지오그래픽(Geographic)
- ④ 모션그래픽(Motion graphic)

38. 다음 중 빅데이터 모형의 개선방안을 수립하기 위한 절차로 가장 옳지 않은 것은 무엇인가?

- ① 예측 오차 계산
- ② 예측모형 개선방향 결정
- ③ 예측모형 분석 인력 분배
- ④ 예측모형 점검 여부 결정

39. 실제는 입력되지 않았지만 입력되었다고 잘못 판단된 변수를 일컫는 말은 무엇인가?

- ① 예측 변수
- ② 정규 변수
- ③ 파생 변수
- ④ 노이즈 변수

40. 다음이 설명하는 검정(Test)은 무엇인가?

- 두 표본의 분산에 대한 차이가 통계적으로 유의한가를 판별하는 검정 기법
- 두 모집단 분산 간의 비율에 대한 검정