

1번) 데이터 사이언티스트의 역할로 틀린 것?

-> 알고리즘에 의해 부당하게 피해 입은 사람을 구제

2번) 빅데이터가 기업/개인.정보에 미치는 영향으로 적절하지 않은 것?

-> 개인은 아직 데이터를 활용할 수 없다.

3번) 정보시스템으로 적절한 것?

-> ERP

4번) 데이터 베이스의 특징으로 틀린 것?

-> 통합된 데이터로 동일한 내용의 데이터가 중복되어 저장

5번) 데이터에 대한 설명으로 틀린 것?

-> 비정형 데이터는 데이터 내부에 메타 데이터를 갖고 있고 파일 형태로 저장

6번) 데이터 모델링에 대한 설명으로 적절한 것?

-> 기업의 원천성 데이터베이스는 데이터마트이다.

7번) 데이터-정보-지식-지혜 구조와 예시로 잘 연결 된 것?

-> (가) 데이터 / (나) 지혜 / (다) 정보 / (라) 지식

8번) 가치 기반 분석을 위해 고려해야하는 사항으로 틀린 것?

-> 인구통계학적 변화

9번) 분석 기획 단계에서 수행하는 주요 태스크로 틀린 것?

-> 필요 데이터 정의

10번) 분석 과제 발굴에 대해 틀린 것?

-> 분석해야 할 대상이 확실하면 상향식 접근 방식이다.

11번) 과제의 우선순위를 결정할 때 고려하는 요소로 아닌 것?

-> 기술 적용 수준

12번) Accurcay와 precision에 대해 틀린 것?

-> 정확도는 true로 예측된 것 중 실제 true인 비율, 정밀도는 실제 true인 경우에서 true로 예측된 비율

13번) 빈칸에 들어갈 과제 도출 방식?

-> ⓐ: 발견, ⓑ: 통찰, ⓒ: 발산

14번) 데이터 거버넌스 체계로 맞는 것?

-> 데이터 표준화

15번) 기업의 분석 성숙도 진단 대상으로 틀린 것?

-> 서비스 부문

16번) 모델링 목적에 따라 변수를 정의하고 필요한 데이터를 소프트웨어에 적용하기 위한 활동 수행 단계?

-> 가공

17번) 모델의 과대 또는 과소 적합에 대한 미세 조정 절차를 위해 사용되는 데이터?

-> Validation Data

18번) summary 결과로 틀린 것?

-> 중앙값은 (평균값으로 추정) 이다.

19번) 부스트랩에서 하나의 관측치가 선정되지 않을 확률?

->  $(1-1/d)^d$

20번) 데이터 탐색 과정에 대해 틀린 것?

-> 변수별 결측치 존재를 확인하고 결측치가 있는 변수는 삭제한다.

21번) t.test 결과로 틀린 것?

-> 유의수준 0.05일 때, 귀무가설이 기각되지 않는다.

22번) 연관분석에 대한 설명으로 틀린 것?

-> 시차 연관 분석의 결과는 원인과 결과의 형태로 해석되지 않는다.

23번) 통계적 가설검정에 대해 틀린 것?

-> p-value가 작을수록 해당 검정통계량의 관측값은 귀무가설을 지지하는 것으로 해석가능

24번) 다차원 척도법에 대해 맞는 것?

-> 대상들을 동일한 상대적 거리를 가진 실수 공간의 점들로 배치시켜 자료들의 상대적 관계를 이해하는 시각화 방법

25번) 회귀분석에 대해 맞는 것?

-> 관측된 연속형 변수들에 대해 두 변수 사이의 모형을 추정한 뒤 변수 간 관계를 파악

26번) 신경망 모형에 대한 설명으로 틀린 것은?

-> 은닉층의 뉴런 수와 개수는 신경망 모형에서 자동으로 설정되며....

27번) 모분산의 추론에 대해 틀린 것?

->

28번) 신경망 모형에서 입력된 데이터를 다음 층으로 출력하는 형태를 결정하는 함수?

-> 활성화 함수

29번) 혼합분포모형에 대한 최대가능도 추정을 위해 사용되는 알고리즘?

-> EM 알고리즘

30번) 주성분분석의 설명으로 틀린 것?

-> 원변수의 선형결합 중 가장 분산이 작은 것을 PC1으로 설정한다.

31번) 텍스트 마이닝 설명으로 틀린 것?

-> .... 비구조화된 단계를 코퍼스라고 한다.

32번) 최단 연결법으로 계층적 군집화 할 때, 첫 번째단계에서 형성되는 군집과 a와의 거리?

-> 3.2

33번) 적합한 ARIMA 모델 결정 방법?

-> ACF 값은 급격히 감소하고 PACF의 절단점이 3인 경우 ARMA(2,0)모델로 정의

34번) MA 모형에 대해 적절한 것?

-> 정상성을 만족하기 위한 조건이 필요함

35번) 단순선형모형의 분산분석표 해석으로 틀린 것?

-> 오차항의 분산의 불편추정량값은 0.1

36번) 다중 공선성에 대해 틀린 것?

-> 다중공선성에서 높은 상관관계에 있는 설명변수에 대한 계수는 표본의 크기에 관계 없이 항상 일정

37번) 사회연결망분석에서 연결망을 표현하는 분석 방법으로 틀린 것?  
→ k-means

38번) 주성분 분석결과에서 첫 번째 주성분 식으로 적절한 것?  
→  $PC1 = -0.54 * Murder - 0.58 * Assault$  ,,,

39번) 회귀분석에서 확인해야할 사항으로 틀린 것?  
→ 모형의 설명력은 -1에서 1사이의 값을 갖는 결정계수를 확인

40번) 사회연결망분석의 중심성을 측정하는 방법으로 틀린 것?  
→ 링크 중심성

단 1) 정보

단 2) 연관규칙분석

단 3) ISP(Information Strategy Planning)

단 4) 프로토 타입 모델

단 5) SOM(Self-Organizing Map)

단 6) 4개

단 7)  $a/(a+b)$

단 8) 부스팅

단 9) 잔차

단 10) 1.2