

제35회 데이터 분석 준전문가 자격검정 시험 복원 문제

•검정일시 : 2022. 10. 29(토) / 10:00~11:30

•수험번호 :

•성 명 :

01. 다음 중 데이터베이스와의 통신을 위해 고안된 언어로 가장 적절한 것은?

- ① Java
- ② R
- ③ Python
- ④ SQL

02. 다음 중 데이터 사이언티스트의 필요 역량으로 가장 부적절한 것은?

- ① 설득력 있는 스토리텔링
- ② 통찰력 있는 분석
- ③ 네트워크 최적화
- ④ 다분야 간 협력을 위한 커뮤니케이션

03. 다음 중 빅데이터 위기 요인 중 사생활 침해를 막기 위한 방지 기술로 적절한 것은 무엇인가?

- ① 익명화
- ② 일반화
- ③ 정규화
- ④ 표준화

04. 다음 중 빅데이터가 발생시키는 문제를 중간자 입장에서 중재하고 해결하는 역할을 하는 직업은 무엇인가?

- ① 데이터 관리자
- ② 알고리즘미스트
- ③ 정보보안 전문가
- ④ 에널리스트

05. 다음 중 빅데이터 및 데이터 사이언스 등의 기술이 가져올 변화로 가장 적절하지 않은 것은?

- ① 해당 기술은 비용절감, 고객 서비스 향상, 내부 의사결정 지원 등에서 엄청난 가치를 발견할 것이다.
- ② 급변하는 환경에서 예측하지 못했던 전환이나 위기에 빨리 적응할 수 있게 할 것이다.
- ③ 사물인터넷의 적용으로 사람의 개입이 최소화 되어 실시간으로 데이터를 수집할 것이다.
- ④ 디지털화된 정보와 대상들이 서로 연결되기 때문에 연결이 얼마나 원활할 지가 중요해질 것이다.

06. 다음 중 빅데이터 기술의 활용에 대한 설명으로 가장 적절하지 않은 것은?

- ① 기업 활용 사례로서 구글 검색 기능, 월마트 매출 향상 등이 있다.
- ② 정부 활용 사례로서 실시간 교통 정보 제공, 기후 정보 제공, 각종 지원 활동 예측 등이 있다.
- ③ 정부는 이익을 목적으로 개인의 정보를 활용할 수 있는 방안을 모색한다.
- ④ 가수는 팬들의 음악 청취 기록을 분석해 공연의 음악 순서 방안을 모색한다.

07. 빅데이터 시대 위기 요인으로 가장 부적절한 것은?

- ① 데이터 오용
- ② 사생활 침해
- ③ 데이터 분석 예측
- ④ 책임원칙 훼손

08. 다음 중 사용자와 데이터베이스 사이에서 사용자의 요구에 따라 정보를 처리해주고 데이터베이스를 관리해 주는 소프트웨어는 무엇인가?

- ① SQL
- ② ERD
- ③ Data Dictionary
- ④ DBMS

09. 다음 중 아래의 데이터 거버넌스 체계가 설명하는 항목은?

아래

메타데이터 관리, 데이터 사전관리, 데이터 생명주기 관리

- ① 데이터 표준화
- ② 데이터 관리 체계
- ③ 데이터 저장소 관리
- ④ 표준화 활동

10. 다음 중 데이터 거버넌스의 구성 요소가 아닌 것은?

- ① 원칙(principle)
- ② 조직(organization)
- ③ 데이터 매니지먼트(Data Management)
- ④ 절차(process)

11. 분석 마스터 플랜 수립에서 과제 우선순위 결정과 관련된 내용으로 부적절한 것은?

- ① 가치는 투자비용 요소이다.
- ② 전략적 중요도, ROI, 실행 용이성은 분석 과제 우선순위 결정에 고려할 사항이다.
- ③ 시급성과 전략적 필요성은 전략적 중요도의 평가 요소이다.
- ④ 적용 기술의 안전성 검증은 기술 용이성의 평가 요소이다.

12. 기업의 데이터 분석 도입 수준을 명확하게 파악하기 위한 방법으로 분석 준비도(readiness)를 진단 할 수 있다. 다음 중 분석 준비도를 측정하기 위한 요소가 아닌 것은?

- ① 분석 목표 및 전략
- ② 분석 기법
- ③ 분석 데이터
- ④ 분석 인력 및 조직

13. 다음 중 난이도와 시급성을 고려하였을 때 우선적으로 추진해야 하는 분석 과제로 적절한 것은?

- ① 난이도 : 쉬움(Easy), 시급성 : 현재
- ② 난이도 : 어려움(Difficult), 시급성 : 미래
- ③ 난이도 : 쉬움(Easy), 시급성 : 미래
- ④ 난이도 : 어려움(Difficult), 시급성 : 현재

14. 아래에서 설명하는 데이터 분석 조직 구조로 가장 적절한 것은?

아래

분석 조직 인력을 현업 부서에 배치하여 분석 업무를 수행하는 형태로서, 전사 차원에서 분석 과제의 우선 순위를 선정하여 수행할 수 있고, 분석 결과를 신속하게 실무에 적용할 수 있다.

- ① 집중 구조
- ② 기능 구조
- ③ 분산 구조
- ④ 확산 구조

15. 분석 과제 발굴의 상향식 접근법에서 프로세스 분석을 통한 절차로 가장 적절한 것은?

- ① 분석 요건 정의 → 분석 요건 식별 → 프로세스 분류 → 프로세스 흐름 분석
- ② 분석 요건 식별 → 프로세스 흐름 분석 → 프로세스 분류 → 분석 요건 정의
- ③ 프로세스 흐름 분석 → 프로세스 분류 → 분석 요건 정의 → 분석 요건 식별
- ④ 프로세스 분류 → 프로세스 흐름 분석 → 분석 요건 식별 → 분석 요건 정의

16. 다음 중 빅데이터 분석 방법론의 분석 기획 단계에서 프로젝트 위험 대응 계획을 수립할 때 예상되는 위험에 대한 대응 방법의 구분으로 부적절한 것은?

- ① 회피(Avoid)
- ② 관리(Manage)
- ③ 완화(Mitigate)
- ④ 수용(Accept)

17. 다음 중 lasso 회귀모형에 대한 설명으로 부적절한 것은?

- ① 모형에 포함된 회귀계수들의 절대값의 크기가 클수록 penalty를 부여하는 방식이다.
- ② 자동적으로 변수선택을 하는 효과가 있다.
- ③ Lpenalty의 정도를 조정하는 모수가 있다.
- ④ L2 penalty를 사용한다.

18. 아래 데이터 셋(data set) A, B간의 유사성을 맨하탄 거리로 계산하면?

물 품	A	B
키	180	175
몸무게	65	70

- ① 0
- ② 10
- ③ $\sqrt{10}$
- ④ $\sqrt{50}$

19. 혼합분포군집 모형의 특징으로 적절하지 않은 것은?

- ① 확률분포를 도입하여 군집을 수행하는 모형 기반 군집 방법이다.
- ② 군집을 몇 개의 모수로 표현할 수 있다.
- ③ 모수 추정에서 데이터가 커지면 수행하는 데 시간이 걸릴 수 있다.
- ④ 군집의 크기가 작을수록 추정의 정도가 쉽다.

20. 다음 중 분류(Classification) 모델링에 대한 설명으로 가장 적절한 것은?

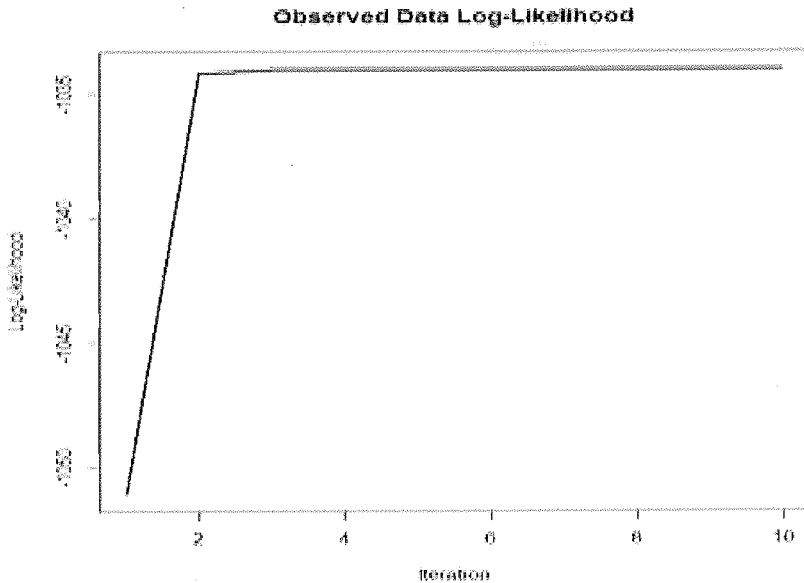
- ① 데이터의 이해를 더 쉽게하기 위해 데이터를 특정 기준으로 분류 및 범주화하고 등급화하는 방법을 말한다.
- ② 같이 팔리는 물건과 같이 유사 아이템을 분류하는 것을 의미한다.
- ③ 군집분석과 동일하게 레코드 자체가 먼저 분류 되어있지 않아도 적용할 수 있다.
- ④ 대표적인 분석 방법으로 장바구니 분석 기법이 존재한다.

21. 다음 중 아래의 표가 나타내는 확률질량함수를 가진 확률변수 x 의 기댓값 $E(x)$ 로 가장 적절한 것은?

x	1	2	3	4
$f(x)$	0.5	0.3	0.2	0

- ① 1
- ② 1.7
- ③ 2.5
- ④ 10

22. EM 알고리즘을 사용하여 혼합분포 모형을 추정하고자 한다. 아래와 같은 그래프가 도출되었을 때, 다음 중 가장 적절한 해석은?



- ① 반복횟수 2회만에 로그-가능도 함수가 최대가 되었다.
- ② 정규혼합분포가 2가지로 관찰되었다.
- ③ 모수의 추정을 위해 10회 이상의 반복횟수가 필요하다.
- ④ 로그-가능도 함수의 최소값이 -1040이다.

23. 확률변수 X의 확률은 아래와 같이 나타낼 수 있다. 다음 중 옳은 것은?

아래

$$P(X=1)=\frac{1}{3}, P(X=2)=\frac{1}{6}, P(X=3)=\frac{1}{2}$$

- ① X의 기댓값은 $13/6$ 이다.
- ② X가 1 혹은 2일 확률은 $1/2$ 보다 크다.
- ③ X가 4일 확률은 0보다 크다.
- ④ X가 1, 2, 3 중 하나의 값을 가질 확률은 1보다 작다.

24. 아래 오분류표를 이용하여 계산된 정밀도는 무엇인가?

		예측치		합계
		True	False	
실제값	True	30	70	100
	False	60	40	100
합계		90	110	200

- ① $3/10$
- ② $4/10$
- ③ $3/9$
- ④ $7/11$

25. Credit 데이터는 400명의 신용카드 고객에 대해 신용카드 대금(balance)과 소득(income), 학생여부(student=Y/N)를 포함한다. Balance를 종속변수로 하는 아래의 모형 적합 결과 중 가장 부적절한 것은?

아래

```
Call:
lm(formula = Balance ~ (Income + Student)^2, data = Credit)

Residuals:
    Min       1Q   Median       3Q      Max
-773.39 -325.70 -41.13  321.65  814.04

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    200.6232    33.6984   5.953 5.79e-09 ***
Income           6.2182     0.5921  10.502 < 2e-16 ***
StudentYes     476.6758    104.3512   4.568 6.59e-06 ***
Income:StudentYes -1.9992     1.7313  -1.155   0.249
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 391.6 on 396 degrees of freedom
Multiple R-squared:  0.2799, Adjusted R-squared:  0.2744
F-statistic: 51.3 on 3 and 396 DF, p-value: < 2.2e-16
```

- ① 위의 모형은 Balance를 설명하는데 통계적으로 유의하다.
- ② Income이 증가할수록 Balance가 증가하는 경향이 있다.
- ③ Income과 StudentYes의 교호작용은 유의하지 않다.
- ④ Income이 증가함에 따라 커지는 Balance의 증가분이 학생 여부에 따라 유의적인 차이가 있다.

26. 다음 중 회귀분석에서 모형의 설명력을 확인하기 위해 사용되는 결정계수의 특성으로 부적절한 것은?

- ① 결정계수는 0에서 1의 값을 가진다.
- ② 높은 값을 가질수록 측정된 회귀식의 설명력이 높다.
- ③ 종속변수와 독립변수 사이의 표본상관 계수값과 같다.
- ④ 총 변동에서 추정된 회귀식에 의해 설명되는 변동의 비율로 나타낼 수 있다.

27. 다음 중 목표변수가 연속형인 회귀나무에서 분류 기준값의 선택 방법으로 가장 적절한 것은?

- ① 카이제곱 통계량, 지니지수
- ② 지니지수, F-통계량
- ③ F-통계량, 분산 감소량
- ④ 분산 감소량, 엔트로피 지수

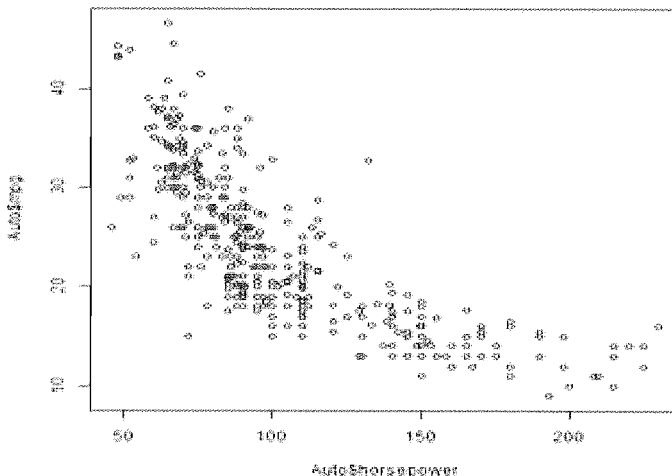
28. 다음 중 자기조직화지도(Self-Organizing Maps, SOM)에 대한 것으로 옳지 않은 것은?

- ① SOM 모델은 입력층과 경쟁층으로 구성되어 있다.
- ② 입력층의 뉴런은 경쟁층에 있는 뉴런들과 부분적으로(locally) 연결되어 있다.
- ③ 단 하나의 전방 패스를 사용함으로써 속도가 매우 빠르다.
- ④ 경쟁 학습으로 연결 강도를 반복적으로 재조정하여 학습한다.

29. 다음 중 선형회귀모형이 통계적으로 유의미한지 평가하는 통계량으로 가장 적절한 것은?

- ① F-statistics
- ② T-statistics
- ③ Chi-statistics
- ④ R-square

30. 아래 그래프는 392대의 자동차에 대한 연비(mpg)와 엔진 마력(horsepower)을 포함하고 있다. 다음 중 이에 대한 설명으로 가장 적절하지 않은 것은?



- ① mpg를 설명하기 위해 horsepower를 설명변수로 하는 단순선형회귀모형은 적절하다.
- ② horsepower가 증가할수록 mpg가 감소하는 경향이 있다.
- ③ mpg와 horsepower 간의 피어슨 상관계수는 두 변수의 관계를 잘 설명하지 못할 수도 있다.
- ④ mpg와 horsepower는 음의 상관관계를 가진다.

31. 다음 중 연관규칙의 측정 지표 중 품목 A, B에 대한 지지도를 구하기 위한 식으로 적절한 것은?

- ① $(A \text{ 또는 } B \text{가 포함된 거래 수}) / (\text{전체 거래 수})$
- ② $(A \text{와 } B \text{가 동시에 포함된 거래 수}) / (\text{전체 거래 수})$
- ③ $(A \text{와 } B \text{가 동시에 포함된 거래 수}) / (A \text{를 포함하는 거래 수})$
- ④ $(A \text{와 } B \text{가 동시에 포함된 거래 수}) / (A \text{ 또는 } B \text{가 포함된 거래 수})$

32. 다음 중 아래 데이터 마이닝 추진 단계를 순서대로 나열한 것은?

아래

- 가. 목적 정의
- 나. 데이터 준비
- 다. 데이터 가공
- 라. 데이터 마이닝 기법 적용
- 마. 검증

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

33. 아래에서 설명하는 활성화 함수로 가장 적절한 것은?

아래

입력층이 직접 출력층에 연결되는 단층 신경망에서 이 활성화 함수를 사용하면 로지스틱 회귀모형의 작동 원리가 유사해진다.

- ① 계단 함수
- ② \tanh 함수
- ③ ReLU 함수
- ④ 시그모이드 함수

34. 다음 중 로지스틱 회귀모형에 대한 설명으로 가장 적절한 것은?

- ① 일반적으로 반응변수가 범주형인 경우에 적용되는 모형이다.
- ② 시계열 예측에서 가장 많이 활용되는 모형 중 하나이다.
- ③ 반응변수가 비율 척도일 때, 많이 활용되는 모형 중 하나이다.
- ④ 로지스틱 회귀모형은 오즈의 관점에서 해석할 수 없다.

35. 다음 중 군집분석에 대한 설명으로 가장 적절하지 않은 것은?

- ① 분할적 군집은 모든 데이터를 단일 군집에 속한다고 정의하고 시작하는 방법으로 상위 군집에서 잘못된 결정을 하면 하위 군집에 파급되는 정도가 크다는 단점이 있다.
- ② k-평균법은 중심으로부터 거리를 기반으로 군집화하기 때문에 구형으로 뭉쳐져 있는 볼록(Convex)한 데이터 세트에서는 비교적 잘 작동되나 오목한(non-convex) 형태의 군집 모델은 특성을 구별해내는 데 성능이 떨어진다.
- ③ k-medoid 모델은 실제 데이터에 있는 값을 중심으로 하기 때문에 이상값이나 잡음(noise) 처리에 매우 우수하나, k-평균법에 비해 계산량이 많다는 단점이 있다.
- ④ 밀도 기반 클러스터링(DBSCAN) 모델은 밀도 있게 연결된 데이터 집합을 동일한 군집으로 판단하는 방법이지만 k-평균법 모델처럼 오목한 형태의 데이터 세트에서는 군집 특성을 잘 찾아내지 못한다.

36. 다음 중 예측모형의 과적합을 방지하기 위해 활용되는 자료 추출 방법으로 가장 적절하지 않은 것은?

- ① 홀드아웃 방법
- ② 교차검증
- ③ 부스트랩
- ④ 의사결정나무

37. 아래 오분류표를 이용하여 계산된 특이도는 무엇인가?

		예측치		합계
		True	False	
실제값	True	200	400	600
	False	300	100	400
합계		500	500	1000

- ① 0.20
- ② 0.25
- ③ 0.75
- ④ 0.80

38. 다음 중 군집의 개수를 미리 정하지 않아도 되어 탐색적 분석에 사용하는 군집 모형으로 적절한 것은?

- ① k-평균군집 모형
- ② SOM 모형
- ③ 계층적군집 모형
- ④ 혼합분포군집 모형

39. 다음 중 연관분석의 설명으로 가장 적절한 것은?

- ① 품목 수와 상관없이 분석에 필요한 계산은 일정하다.
- ② 세분화된 품목에 대해 연관 규칙을 찾으려 할 때 적절한 방법이다.
- ③ 상대적으로 거래량이 적은 품목에 대해서 적용하기 좋은 방법이다.
- ④ 조건 반응(if-then)으로 표현되는 연관분석의 결과를 이해하기 쉽다.

40. 다음 중 시계열 데이터의 정상성(stationary)에 대한 설명으로 가장 적절하지 않은 것은?

- ① 비정상 시계열 자료는 정상성을 만족하도록 데이터를 정상 시계열로 만든 후에 시계열 분석을 수행한다.
- ② 정상 시계열은 어떤 일정한 값을 중심으로 일정한 변동 폭을 가진다.
- ③ 시계열 자료가 추세를 보이는 경우에는 차분(differencing)을 통해 비정상 시계열을 정상 시계열로 바꿀 수 있다.
- ④ 시계열 자료가 정상성을 만족하는지 판단하기 위해 시계열 자료 그림을 통해 자료의 이상점 등을 살핀다.