

# 2022년 35회 ADsP 기출복원 (인쇄용)

tbcase73@gmail.com [계정 전환](#)



비공개

**R1. 사용자와 데이터베이스 사이에서 사용자 요구에 따라 정보처리 데이터베이스 관리를 하는 소프트웨어를 무엇이라고 하는가?** 1점

- ☐ DBMS
- ☐ Data Dictionary
- ☐ SQL
- ☐ ERD

**2. 데이터 사이언티스트의 필요 역량으로 적절하지 않은 것은?**

1점

- ☐ 하드 스킬과 소프트 스킬 능력
- ☐ 데이터 처리 기술
- ☐ 네트워크 최적화 능력
- ☐ 고객들에 대한 공감 능력

**3. 사생활 침해 방지 기술에 해당하는 것으로 개인 식별 정보를 삭제하거나 알아볼 수 없는 형태로 변환하는 포괄적 기술로 올바른 것은?** 1점

- ☐ 익명화
- ☐ 데이터 마스킹
- ☐ 가명
- ☐ 데이터값 삭제



4. 데이터 분석 알고리즘으로 부당한 피해를 보는 사람을 방지하기 위해서 생겨난 직업으로 데이터 분석 알고리즘으로 인해 피해를 입은 사람을 구제하는 전문가를 무엇이라 하는가? 1점

- ☐ 데이터 엔지니어
- ☐ 알고리즘리스트
- ☐ 데이터 사이언티스트
- ☐ 데이터 분석가

5. 빅데이터의 영향에 대해 옳바르지 않은 것은 무엇인가? 1점

- ☐ 산업 전체의 생산성이 향상되었다
- ☐ 맞춤형 서비스를 저렴한 비용으로 이용할 수 있게 되었다
- ☐ 사회 변화를 추정, 각종 재해 관련 정보를 추출할 수 있게 되었다
- ☐ 사물인터넷(IoT)의 발달로 인해 사람이 최대로 개입하게 되었다

6. 빅데이터의 기술 활용에 관련된 설명으로 거리가 먼 것은? 1점

- ☐ 기업은 원가절감, 제품차별화, 기업활동의 투명성 제공 등에 활용한다
- ☐ 미래 사회 도래에 대비한 법제도 및 거버넌스 시스템 정비 방향, 미래 성장 전략 등에 대한 정보 제공한다
- ☐ 정부의 이익을 위해 개인의 정보를 활용한다
- ☐ 적시에 필요한 정보를 얻어 다양한 형태로 기회 비용을 절약할 수 있다

7. 다음 중 빅데이터 위기요인과 통제방안에 대한 내용과 관련이 없는 것은? 1점

- ☐ 사생활 침해
- ☐ 데이터 오용
- ☐ 데이터 변화 관리
- ☐ 책임 원칙의 훼손



8. 다음 중 데이터베이스와의 통신을 위해 고안된 언어는 무엇인가?

1점

- ☐ Python
- ☐ Java
- ☐ R
- ☐ SQL

9. 문자, 기호, 음성, 화상, 영상 등 상호 관련된 다수의 콘텐츠를 정보처리 및 정보통신 기기에 의해 체계적으로 수집, 축적하여 다양한 용도와 방법으로 이용할 수 있도록 정리한 정보의 집합체는?

1점

내 답변

10. 다음 설명에 해당하는 빅데이터 활용 테크닉은 무엇인가?

1점

“최대의 시청률을 얻으려면 어떤 프로그램을 어떤 시간대에 방송해야 하는가와 같은 최적화의 메커니즘을 찾아가는 방법이다 ”

내 답변

1. 데이터 분석을 위한 조직 구조 중 분석 조직 인력들을 현업부서로 직접 배치하여 신속한 업무 수행이 가능한 구조는 무엇인가 ?

1점

- ☐ 분산 조직 구조
- ☐ 집중형 조직 구조
- ☐ 기능 중심 조직 구조
- ☐ 혼합형 조직 구조



2. 데이터 거버넌스 체계 단계 중 메타데이터와 데이터 사전(Data Dictionary)의 관리 원칙 수립과 관련된 단계는 무엇인가? 1점

- ☐ 데이터 표준화
- ☐ 데이터 관리체계
- ☐ 데이터 저장소관리
- ☐ 표준화 활동

3. 분석 마스터 플랜의 과제 우선순위 결정과 관련된 내용으로 적절하지 않은 것은? 1점

- ☐ 난이도 판단기준은 데이터 획득/저장/가공 비용 및 분석 적용 비용, 분석 수준 등이 있다
- ☐ 시급성의 판단기준은 전략적 중요도가 핵심이다
- ☐ Value(가치)는 투자비용 요소이다
- ☐ ROI 관점에서의 분석 과제 우선순위 평가 기준은 시급성과 난이도가 있다

4. 다음 중 데이터 거버넌스의 구성요소가 아닌 것은? 1점

- ☐ 원칙(Principle)
- ☐ 조직(Organization)
- ☐ 분석 방법(Method)
- ☐ 프로세스(Process)

5. 분석 과제 도출 방법 중 상향식 접근 방식의 절차로 알맞은 것은? 1점

- ☐ 프로세스 분류 -> 프로세스 흐름분석 -> 분석요건 식별 -> 분석요건 정의
- ☐ 프로세스 흐름 분석 -> 분석요건 식별 -> 분석 요건 정의 -> 프로세스 분류
- ☐ 프로세스 흐름 분석 -> 분석요건 식별 -> 프로세스 분류 -> 분석 요건 정의
- ☐ 프로세스 분류 -> 분석요건 식별 -> 분석 요건 정의 -> 프로세스 흐름 분석



6. 다음 중 분석과제의 우선순위 선정 시 난이도와 시급성을 모두 고려하였을 때 우선적으로 추진해야하는 분석 과제는 무엇인가? 1점

- ☐ 난이도 - 어려움, 시급성 - 미래
- ☐ 난이도 - 쉬움, 시급성 - 미래
- ☐ 난이도 - 어려움, 시급성 - 현재
- ☐ 난이도 - 쉬움, 시급성 - 현재

7. 다음 중 기업의 분석 도입의 수준을 파악하기 위한 분석 준비도와 관계가 적은 항목은 무엇인가? 1점

- ☐ 분석 인력 및 조직
- ☐ 분석 기법
- ☐ 목표와 정책
- ☐ 분석 데이터

8. 다음 중 빅데이터 분석 방법론의 분석 기획 단계에서 프로젝트 위험 계획 수립 시 위험에 대한 대응 방법의 종류에 포함되지 않는 것은? 1점

- ☐ 회피(avoid)
- ☐ 전이(transfer)
- ☐ 완화(mitigate)
- ☐ 관리(management)

9. 문제가 주어지고 해답을 찾기 위해 각 과정이 체계적이고 단계화 되어 수행하는 분석 과제 도출 방식은 무엇인가? 1점

내 답변



10. 다음이 설명하는 분석 조직 구조는 무엇인가?

1점

- 조직내 별도 독립적인 분석 전담 조직 구성, 분석 전담 조직에서 회사의 모든 분석 업무를 담당
- 전사분석 과제의 전략적 중요도에 따라 우선 순위를 정해 추진
- 일부 협업 부서와 분석 업무가 중복 또는 이원화될 가능성이 있음

내 답변

1. 다음 중 회귀분석의 결정계수에 관한 설명으로 적절하지 않은 것은 무엇인가? 1점

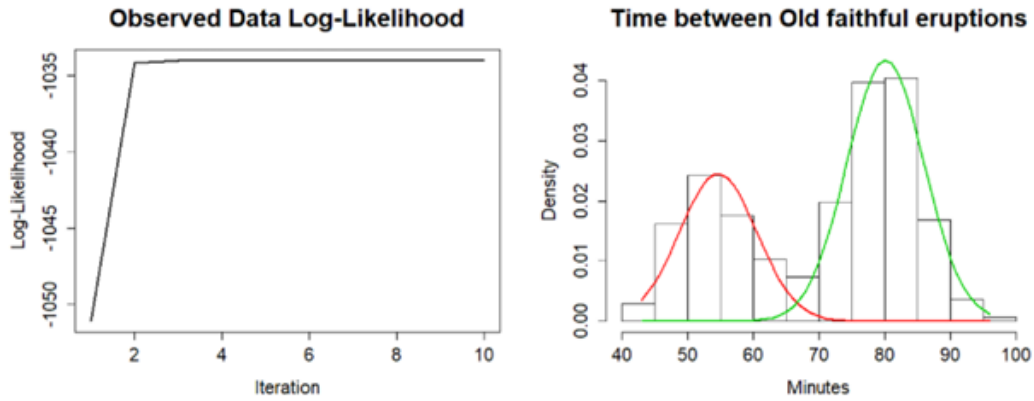
- ☐ 결정계수는 회귀제곱합(SSR) / 총제곱합(SST) 로 구할 수 있다
- ☐ 종속변수와 독립변수 사이의 표본 상관계수와 값이 같다
- ☐ 결정계수가 커질수록 회귀방정식의 설명력이 높다고 할 수 있다
- ☐ 일반적으로 결정계수는 0 ~ 1의 값을 갖는다

2. 다음 중 목표변수가 연속형인 회귀나무의 분류 기준값을 선택하는 기준으로 구성된 것은 무엇인가? 1점

- ☐ 지니 지수(gini), 엔트로피 지수(entropy)
- ☐ 카이 제곱 통계량, 분산 감소량
- ☐ F 통계량, 분산 감소량
- ☐ 엔트로피 지수(entropy), 카이 제곱 통계량



3. EM알고리즘을 사용한 혼합분포 모형의 결과 해석에 대한 설명으로 적절한 것은 무엇인가? 1점



- ☐ 반복횟수 2회 만에 로그가능도 함수가 최대가 됨을 알 수 있다
- ☐ 로그 가능도 함수의 최대값은 -1050이다
- ☐ 결과적으로 3개의 정규분포가 혼합된 것을 알 수 있다
- ☐ 모수 추정을 위해 8회 이상의 반복이 필요함을 알 수 있다

4. Lasso 회귀 모형의 정의로 옳지 않은 것은?

1점

- ☐ 모형에 포함된 회귀계수의 절댓값이 클수록 penalty를 부여하는 방식이다.
- ☐ 람다값(lambda)으로 penalty의 정도를 조정한다.
- ☐ 자동적으로 변수 선택을 하는 효과가 있다.
- ☐ L2 norm을 사용하여 penalty를 부여한다.

5. SOM에 대한 설명으로 옳지 않은 것은?

1점

- ☐ SOM은 비지도 학습이다
- ☐ SOM은 차원축소와 군집화를 동시에 수행하는 기법이다
- ☐ 입력층과 출력층이 부분연결 되어 있다
- ☐ 출력 뉴런들은 승자 뉴런이 되기 위해 경쟁하고 오직 승자만이 학습한다.

6. 다음 오분류표를 사용하여 특이도(Specificity)를 구한 결과는 무엇인가?

1점

오분류표		예측값		합계
		TRUE	FALSE	
실제값	TRUE	300	300	600
	FALSE	450	150	600
합계		750	450	1200

- ☐ 0.375
- ☐ 0.75
- ☐ 0.25
- ☐ 0.5

7. 다음 중 입력신호를 받아 출력신호로 연결하기 위한 활성화 함수로 로지스틱 회귀 모델에서도 사용하는 함수는 무엇인가?

1점

- ☐ sigmoid
- ☐ ReLU
- ☐ tanh
- ☐ log

8. 로지스틱 회귀에 대한 특징으로 적절한 것은?

1점

- ☐ 모형 검정에는 F 검정이 사용된다
- ☐ 종속변수(=반응변수)가 범주형인 경우 적용되는 회귀분석 모형이다
- ☐ softmax 함수를 사용하여 종속변수를 전체 실수 범위로 확장하여 분석한다
- ☐ 모형 탐색 방법에는 최소자승법(최소제곱법)이 있다





9. 선형회귀 모델의 통계적 유의성 검증을 위해 사용하는 것은?

1점

- ☐ 회귀계수의 t 통계량
- ☐ 결정 계수
- ☐ 잔차 통계량
- ☐ F 통계량

10. 의사결정 나무에 대한 설명 중 적절하지 않은 것은?

1점

- ☐ 비지도 학습으로 상향식 접근 방법을 이용한다.
- ☐ 구조가 단순하여 해석이 용이하다.
- ☐ 목표변수가 이산형인 경우 분류나무, 목표변수가 연속형인 경우 회귀나무가 있다.
- ☐ 부모마디보다 자식마디의 순수도가 증가하도록 분류나무를 형성해 나간다

11. 다음 중 군집의 수를 미리 지정하지 않으며 탐색적 기법에 적합한 군집 방법은 무엇인가? 1점

- ☐ 계층적 군집
- ☐ 비계층적 군집
- ☐ K-means 군집
- ☐ 혼합분포 군집

12. 확률변수 x의 기댓값은 무엇인가?

1점

X	1	2	3
f(X)	0.5	0.3	0.2

- ☐ 0.5
- ☐ 1.7
- ☐ 6
- ☐ 2



13. 다음 두 좌표(A, B) 간의 맨해튼 거리(Manhattan distance)를 구하시오.

1점

	A	B
키	175	180
몸무게	70	65

- ☐ 10
- ☐ 50
- ☐  $\sqrt{10}$
- ☐  $\sqrt{50}$

14. 혼합분포 군집의 특징으로 적절하 않은 것은 무엇인가?

1점

- ☐ 군집을 몇 개의 모수로 표현할 수 있으며, 확률분포를 도입하여 군집 수행한다
- ☐ EM 알고리즘을 이용한 모수 추정에서 데이터가 커지면 수렴에 시간이 더 많이 걸릴 수 있다.
- ☐ 군집의 크기가 작을 수록 추정이 쉽고, 정밀한 추정이 가능하다
- ☐ 복잡한 형태를 가진 분포의 경우 여러 분포를 확률적으로 선형 결합한 혼합분포로 설명할 수 있다

15. 다음 시계열 분석에 대한 설명 중 옳지 않은 것은 무엇인가?

1점

- ☐ 데이터가 추세를 가지면 변환(자연로그)를 사용하여 정상시계열로 만든다
- ☐ 정상 시계열인 경우 평균값 주변에서의 변동은 대체로 일정한 폭을 갖는다
- ☐ 시계열 데이터는 대부분 비정상 시계열이기 때문에 정상 시계열로 만든 후 분석을 수행한다
- ☐ 시계열 그래프를 보면서 이상치와 정상성 여부를 확인할 수 있다



16. 다음 중 군집분석에 대한 설명으로 옳지 않은 것은 무엇인가?

1점

- ☐ 유사성을 이용하여 몇 개의 집단으로 그룹화하는 분석이다
- ☐ 집단별 특성이 유사할 경우 안정성이 높다
- ☐ 군집 분석은 이상치 자료에 민감한 특성이 있다
- ☐ 안정성 검토 방법으로 지도학습과 동일한 교차타당성(Cross Validation)을 사용한다

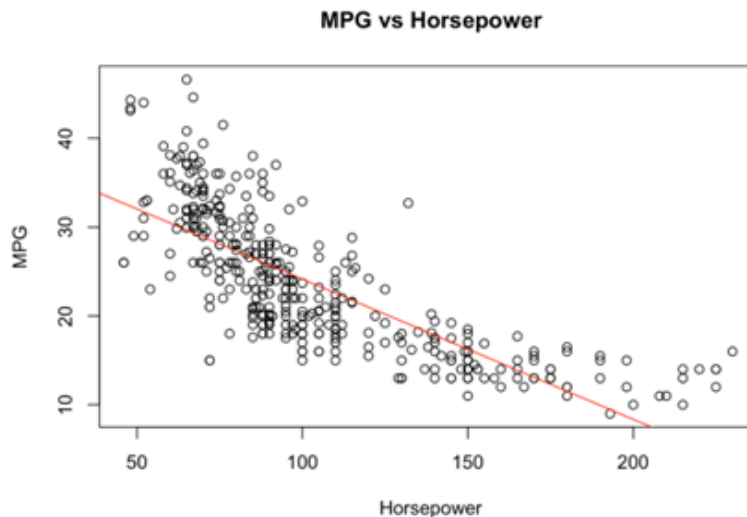
17. 다음 연관규칙 관련 식 중 A --> B 일 때의 지지도(Support)에 대한 올바른 식은 무엇인가?

1점

- ☐ A와 B가 동시에 포함된 거래 수 / B가 포함된 거래 수
- ☐ A와 B가 동시에 포함된 거래 수 / A가 포함된 거래 수
- ☐ A와 B가 동시에 포함된 거래 수 / 전체 거래 수
- ☐ 품목B를 구매한 고객 대비 품목 A를 구매한 후 품목 B를 구매하는 고객에 대한 확률

18. 아래 산점도는 차량 392대의 연비(mpg)와 마력(horsepower)에 관한 그래프이다. 이와 관한 설명으로 가장 적절하지 않은 것은 무엇인가?

1점



- ☐ 연비와 마력은 음의 상관관계이다
- ☐ 연비-마력의 상관관계는 피어슨 상관계수로 분석이 가능하지 않다
- ☐ 연비와 마력 간의 영향력으로 단순 선형회귀모형 추정이 가능하다
- ☐ 마력이 증가할 때 연비가 감소하는 경향이 있다

19. 다음 이산형 확률분포의 확률변수  $x$ 에 대한 설명 중 적절한 것은 무엇인가? 1점

$x$	1	2	3
$f(x)$	$1/6$	$1/2$	$1/3$

- ☐ 확률변수  $x$ 의 확률의 합은 1보다 작거나 클 수 있다
- ☐ 확률변수  $x$ 가 0이거나 4일 확률은 0이 아니다
- ☐ 확률변수  $x$ 에 대한 기댓값은  $13/6$  이다
- ☐ 확률변수  $x$ 가 1이거나 2일 확률은  $5/6$  이다

20. 다음 회귀 분석 결과에 대한 해석 중 옳지 않은 설명은 무엇인가?

1점

```
Call:
glm(formula = default ~ ., family = "binomial", data = Default)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4691  -0.1418  -0.0557  -0.0203   3.7383

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.087e+01  4.923e-01 -22.080  < 2e-16 ***
studentYes  -6.468e-01  2.363e-01  -2.738  0.00619 **
balance      5.737e-03  2.319e-04  24.738  < 2e-16 ***
income       3.033e-06  8.203e-06   0.370  0.71152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1571.5  on 9996  degrees of freedom
AIC: 1579.5

Number of Fisher Scoring iterations: 8
```

- ☐ 로지스틱 회귀 모델을 사용한 분석 결과이다
- ☐ income은 default를 설명하는데 통계적으로 유의미한 변수이다.
- ☐ studentYes의 값이 Yes 일 때, 채무불이행(default) 될 확률이 낮다.
- ☐ balance는 default를 설명하는데 통계적으로 유의미한 변수이다



21. 다음 중 데이터 마이닝 프로세스 순서로 올바르게 나열한 것은 무엇인가?

1점

(가) 목적 정의

(나) 데이터 준비

(다) 데이터 가공

(라) 데이터 마이닝 기법의 적용

(마) 검증

☐ (가)-(마)-(다)-(나)-(라)

☐ (가)-(나)-(다)-(라)-(마)

☐ (가)-(나)-(라)-(마)-(다)

☐ (나)-(가)-(다)-(라)-(마)

22. 다음 중 과대적합 방지를 위한 방법이 아닌 것은 무엇인가?

1점

☐ 배깅(bagging)

☐ 홀드 아웃( Hold Out)

☐ 의사결정 나무

☐ Lasso, Ridge 모델

23. 다음 중 분류 모형에 대한 설명으로 적절한 것은 무엇인가?

1점

☐ 레코드 자체가 가진 다른 레코드와의 유사성에 의해 그룹화되고 이질성에 의해 세분화된다

☐ 카탈로그 배열, 교차판매 등의 마케팅 계획에 사용되는 데이터마이닝 기법이다

☐ 새롭게 나타난 현상을 검토하여 기존의 분류, 정의된 집합에 배정하는 것으로 현상 이해를 위해 데이터를 범주, 등급 등으로 나눈다

☐ 데이터가 가진 특징 및 의미를 단순하게 설명하는 것이다



24. 다음 중 연관분석에 대한 특징으로 적절한 것은 무엇인가?

1점

- ☐ 조건반응(if ~ then)으로 표현되는 연관 분석의 결과를 이해하기 쉽다
- ☐ 강력한 목적성 분석 기법에 해당한다
- ☐ 세분화된 품목을 가지고 연관규칙을 찾아야만 의미 있는 분석 결과가 도출된다
- ☐ 분석 품목 수가 증가하더라도 분석 계산이 많이 증가하지는 않는다

25. 다음이 설명하는 이산형 확률분포는 무엇인가?

1점

“단위 시간이나 단위 공간에서 어떤 사건이 몇 번 발생할 것인지를 표현하는 분포로 특정 기간 동안 사건 발생의 확률을 구할 때 사용된다”

내 답변

26.  $P(A)=0.3$ ,  $P(B)=0.4$  일 때, 사건 A와 사건 B가 독립사건일 경우  $P(B|A)$ 는?

1점

내 답변

27. 다음 빈 칸에 들어갈 알맞은 용어는 무엇인가?

1점

(        ) 두 군집 사이의 거리를 군집에서 하나씩 관측 값을 뽑았을 때 나타날 수 있는 거리의 최솟값을 측정하는 계층적 군집의 거리 기반 측정 방법이다. 사슬 모양으로 생길 수 있으며 고립된 군집을 찾는데 중점을 두는 방식이다.

내 답변

28. 다음이 설명하는 데이터마이닝의 모형평가 방법은 무엇인가?

1점

원천 데이터를 랜덤하게 두 분류로 분리하여 교차검정을 실시하는 방법으로 하나는 모형 학습 및 구축을 위한 훈련용 자료로 다른 하나는 성과평가를 위한 검증용 자료로 사용하는 방법이다.

내 답변



29. 설명변수 선택 방법 중에서 독립변수 후보를 모두 포함한 모형에서 출발해 1점  
제곱합의 기준으로 가장 적은 영향을 주는 변수부터 하나씩 제거하면서 더 이상  
유의하지 않은 변수가 없을 때까지 설명변수를 제거하는 모형은 무엇인가?

내 답변

30. 차원 축소 기법 중, 객체들 사이의 유사성, 비유사성을 2차원 혹은 3차원 공 1점  
간상에 점으로 표현하여 개체 사이의 군집을 시각적으로 표현하는 기법은 무엇  
인가?

내 답변

제출

양식 지우기

Google Forms를 통해 비밀번호를 제출하지 마세요.

이 콘텐츠는 Google이 만들거나 승인하지 않았습니다. [악용사례 신고](#) - [서비스 약관](#) - [개인정보처리방침](#)

Google 설문지



