

## ADsP 32회 기출문제 답안

### 【객관식 정답】

01	③	11	②	21	③	31	③
02	②	12	④	22	③	32	④
03	①	13	④	23	③	33	②
04	③	14	①	24	④	34	④
05	①	15	④	25	②	35	④
06	①	16	①	26	②	36	④
07	②	17	④	27	④	37	②
08	②	18	③	28	④	38	③
09	④	19	①	29	②	39	①
10	②	20	④	30	③	40	①

### 【주관식 정답】

01	사물인터넷 (IoT, Internet of Things)
02	SCM(Supply Chain Management)
03	프레이밍 효과(Framing effect)
04	활용
05	분해 시계열
06	AR 모형
07	과대적합(Overfitting)
08	인공 신경망
09	이익도표
10	데이터 마트

01. SQL 명령어 중 DML(Data Manipulation Language)은 데이터 조작어로 SELECT, INSERT, UPDATE, DELETE가 있다. CREATE는 데이터 정의어인 DDL(Data Definition Language)에 포함된다.
02. 고객이 제품 1개에 대해 여러 명이 구매할 수 있다. 그러므로 N:1이 정답이다.
03. ERP는 회사의 모든 정보 뿐만 아니라, 공급 사슬관리, 고객의 주문정보까지 포함하여 통합적으로 관리하는 시스템이다. 경영, 인사, 재무, 생산 등 기업의 전반적 시스템을 하나로 통합함으로써 효율성을 극대화하는 경영 전략이다.
04. SVM은 분류분석의 기법 중 하나로 딥러닝과 관련 없는 분석 기법이다.
05. 판별분석, 회귀분석, 분류분석은 지도학습의 방법이며, 군집분석은 비지도학습의 방법이다.
06. 빅데이터 활용의 기본 3요소는 데이터, 기술, 인력이다.
07. 컨버전스에서 디버전스로의 변화, 생산에서 서비스로의 변화, 생산에서 시장창조로의 변화가 인문학 열풍을 가져오게 한 외부환경 요소이다.
08. 알고리즘미스트는 알고리즘에 의해 부당하게 피해입은 사람을 구제한다.
09. 조직 및 해당 산업에 폭넓게 영향을 미치는 사회·경제적 요인을 STEEP으로 요약되는 Social(사회), Technological(기술), Economic(경제), Environmental(환경), Political(정치) 영역으로 폭넓게 나눈다.

10. SoW(Statement of Works)는 프로젝트 관리분야에서 서비스를 제공하기 위한 활동, 산출물, 작업 시간 등을 포함하는 기술서이다.
11. 데이터 표준화는 데이터 표준 용어 설정, 명명 규칙 수립, 메타 데이터 구축, 데이터 사전 구축 등의 업무로 구성된다.
12. 분석 ROI 요소에서 투자비용 요소는 크기(Volume), 다양성(Variety), 속도(Velocity)이다.
13. 데이터 거버넌스는 전사 차원의 모든 데이터에 대하여 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리체계를 수립하고 운영을 위한 프레임워크(Framework) 및 저장소(Repository)를 구축하는 것을 말한다.
14. 빅데이터 분석 방법의 5단계는 분석기획, 데이터 준비, 데이터 분석, 시스템 구현, 평가 및 전개 순으로 이루어진다.
15. 분석 기획 선별 방식 중 틈다운은 기업의 비즈니스 모델을 분석하여 경쟁력 강화를 위한 핵심 분석기회를 식별하며, 보텀업 경로 접근 방식은 특정 대상 프로세스를 선정한 후 주제별로 분석기회를 식별한다. 마지막으로 분석 유즈 케이스 벤치마킹 접근 방식은 제공되는 산업별, 업무 서비스별 분석 테마 후보 풀을 벤치마킹을 통한 분석 기회를 식별한다.
16. 분석 수준 진단의 대상은 분석 업무, 분석 인력/조직, 분석 기법, 분석 데이터, 분석 문화, 분석 인프라가 있다.
17. 유의수준은 귀무가설의 기각 여부를 결정하는데 사용하는 기준이 되는 확률로 귀무가설이 옳은데도 이를 기각하는 확률을 크기라고도 한다.
18. 리스트는 여러 자료형의 원소들이 포함될 수 있다.
19. Precision은 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율이다.
20.  $A \rightarrow B$ 의 신뢰도를 구하면  $\frac{P(A \cap B)}{P(A)} = \frac{300/1000}{600/1000} = 0.5$  이다.
21. 잡은 무작위적 변동이며 일반적인 원인이 알려져 있지 않다.
22. 민감도는  $\frac{TP}{TP + FN} = \frac{40}{40 + 60} = 0.4$  이다.
23. 향상도는  $\frac{\frac{300}{1000}}{\frac{600}{1000} \times \frac{600}{1000}} = \frac{5}{6} = 0.833...0$ 이므로 83%이다.
24. 연관분석은 상품의 구매, 서비스 등 일련의 거래 또는 사건들 간의 규칙을 발견하기 위해 적용하며, 카탈로그 배열, 교차 판매 등의 마케팅에 활용된다.
25. 특이도=  $TN / (TN + FP)$ 이므로 여기서  $TN + FP$ 은  $N$ 이다. 즉, 민감도=  $TN \div N$  이다.
26. 혼합분포군집은 모형기반의 군집 방법이며, 데이터가  $k$ 개의 모수적 모형의 가중합으로 표현되는 모집단 모형으로부터 나왔다는 가정하에서 모수와 함께 가중치를 자료로부터 추정하는 방법을 사용한다.

27. 카이제곱 검정은 모수적 검정 방법이다.
28. 마할라노비스 거리는 통계적 개념이 포함된 거리이며 변수들의 산포를 고려하여 표준화한 거리이다. 두 벡터 사이의 거리를 산포를 의미하는 표본 공분산으로 나눠주어야 하며, 그룹에 대한 사전 지식 없이는 표본 공분산을 계산 할 수 없으므로 사용하기 곤란하다.
29. 시계열 자료는 시간의 흐름에 따라 관찰된 값을 의미한다.
30. 군집분석은 각 객체의 유사성을 측정하여 유사성이 높은 대상 집단을 분류하고, 군집에 속한 객체들의 유사성과 서로 다른 군집에 속한 객체간의 상이성을 규명하는 분석 방법이다.
31. 상자수염그림 안에 그려진 선은 중앙값을 의미한다.
32. 비표본오차는 표본오차를 제외한 모든 오차로서 조사 과정에서 발생하는 모든 부주의나 실수, 알 수 없는 원인 등 모든 오차를 의미하며 조사대상이 증가하면 오차가 커진다.
33. LOOCV는 전체 관측치(n) 중 단 하나의 관측값만을 Validation set으로 사용하고 나머지 n-1개 관측값은 Train set으로 사용하므로  $k=n$ 인 경우이다.
34. 제1주성분은 변동을 최대로 설명해주는 방향으로의 변수들의 선형결합식이다.
35. k-평균군집은 먼저 원하는 군집의 개수와 초기 값들을 정해 seed 중심으로 군집을 형성하고 각 데이터를 거리가 가까운 seed가 있는 군집으로 분류한 후 각 군집내의 자료들의 평균을 계산하며 모든 개체가 군집으로 할당될 때까지 과정을 반복한다.
36. 사분위수를 이용하여  $(Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR) = (4 - 1.5 \cdot (12 - 4), 12 + 1.5 \cdot (12 - 4)) = (-8, 24)$ 로 이상값을 판단하는 하한선과 상한선을 구할 수 있다.
37. 사분위수 범위는 3사분위수에서 1사분위수를 뺀 값으로 전체 자료의 중간에 있는 절반의 자료들이 지나는 값의 범위를 말하며, 중간에 50%의 데이터들이 흩어진 정도를 의미한다.
38. 주성분 분석은 여러 변수들의 변량을 주성분이라는 서로 상관성이 높은 변수들의 선형 결합으로 만들어 기존의 상관성이 높은 변수들을 요약, 축소하는 기법이다.
39. 많은 모형에서 공통적으로 사용될 수 있는 변수는 요약변수이다.
40. 정상성의 기준은 모든 시점에서 일정한 평균을 가지며, 분산도 일정, 마지막으로 공분산은 시차에만 의존한다.

## 【 단답형 】

---

단답형 01. 사물인터넷(IoT, Internet of Things)

단답형 02. SCM(Supply Chain Management)

단답형 03. 프레임링 효과(Framing effect)

단답형 04. 활용

단답형 05. 분해 시계열은 시계열 데이터를 요인에 따라 분해하는 기법으로 추세, 순환, 계절, 불규칙요인이 있다.

단답형 06. AR모형은 p시점 전의 자료가 현재 자료에 영향을 주는 모형이다.

단답형 07. 너무 큰 나무모형은 자료를 과대적합하고 너무 작은 나무모형은 과소적합할 위험이 있다.

단답형 08. 인공신경망은 인간의 신경세포를 통한 학습방법에서 아이디어를 얻어 이를 디지털 네트워크 모형으로 구현한 것이다.

단답형 09. 이익도표는 분류모형의 성능의 평가하기 위한 척도로, 분류된 관측치에 대해 얼마나 예측이 잘 이루어졌는지를 나타내기 위해 임의로 나눈 각 등급별로 반응검출율, 반응률, 리프트 등의 정보를 산출하여 나타내는 도표이다.

단답형 10. 데이터 마트는 데이터 웨어하우스와 사용자 사이의 중간층에 위치한 것으로, 하나의 주제 또는 하나의 부서 중심의 데이터 웨어하우스라고 할 수 있다.