

ADsP 30회 기출문제 답안

【객관식 정답】

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

【주관식 정답】

01	정보
02	ERP
03	a : 마스터 데이터, b : 메타데이터, c : 데이터사전
04	시급성
05	4개
06	양상불 기법
07	25%
08	포아송분포(Poisson Distribution)
09	0.8
10	머신러닝 or 기계학습

01. 빅데이터가 만들어내는 본질적인 변화는 사전처리에서 사후처리, 표본조사에서 전수조사, 질에서 양, 인과관계에서 상관 관계이다.
02. 구글, 애플 등의 기업에서는 정형화된 데이터 이외에 비정형 데이터를 수집하여 서비스에 활용하고 있다.
03. 미래 사회의 특성과 빅데이터의 역할은 불확실성 - 통찰력, 리스크 - 대응력, 스마트 - 경쟁력, 융합 - 창조력의 관계를 맺고 있다.
04. 빅데이터를 활용해 기업의 혁신, 경쟁력제고, 생산성이 향상되었으며, 정부입장에서는 환경탐색, 상황분석, 미래대응에 활용할 수 있게 되었다. 데이터 수집 및 저장은 일반적인 데이터를 활용해서도 가능하다.
05. 데이터 분석가는 데이터를 기반으로 의사결정을 하기 때문에 천재적 직관력은 필요한 능력이 아니다.
06. 데이터베이스의 일반적인 특징은 통합된 데이터(Integrated Data), 저장된 데이터(Stored Data), 공유 데이터(Shared Data), 변화되는 데이터(Changable Data)이다.
07. 빅데이터 가치 산정이 어려운 이유는 새로운 가치 창출, 데이터 활용 방식, 분석 기술의 발전이다.
08. 일차원적 분석 어플리케이션 중 에너지는 트레이딩, 공급/수요 예측이 있다.

09. 데이터 거버넌스 체계 중 데이터 관리 체계는 데이터 정확성 및 활용의 효율성을 위하여 표준데이터를 포함한 메타 데이터와 데이터 사전의 관리 원칙을 수립한다. 빅데이터의 경우 데이터 양의 급증으로 데이터의 생명 주기 관리 방안을 수립하지 않으면 데이터 가용성 및 관리비용 증대 문제에 직면하게 될 수 있다.
10. 데이터 거버넌스의 구성요소는 원칙(Principle), 조직(Organization), 프로세스(Process)는 유기적으로 조합하고 효과적으로 관리하여, 데이터를 비즈니스 목적에 부합하도록 하고 최적의 정보 서비스를 제공할 수 있도록 한다.
11. 과제 발굴 단계에서는 세부적인 구현 및 솔루션에 초점을 맞추는게 아니라, 문제를 해결함으로써 발생하는 가치에 중점을 두는 것이 중요하다.
12. 과제 중심적인 접근 방식의 특징에는 Speed & Test, Quick-Win, Problem Solving이 해당하며 Accuracy & Deploy는 장기적인 마스터 플랜 방식에 해당하는 내용이다.
13. 상황식 접근 방식은 프로토타이핑 등을 활용하여 반복적으로 결과를 개선해 나간다. Design Thinking의 Ideate 단계에는 diverge 단계에 해당한다.
14. 분석과제 우선순위 고려요소는 전략적 중요도, 비즈니스 성과/ROI, 실행 용이성이 있다.
15. 분석 방안 구체화에는 의사결정 요소 모형화, 분석 체계 도출, 분석 필요 데이터 정의, 분석 ROI 평가, 분석 활용 시나리오 정의, 분석 정의서 작성, 전사관점 분석적용이 있다.
16. 분석 준비도는 분석업무파악, 인력 및 조직, 분석기법, 분석 데이터, 분석문화, IT 인프라 총 6가지로 구성되어 있다.
17. 민코우스키거리는 맨하탄 거리와 유클리디안 거리를 한번에 표현한 공식으로 L1 거리(맨하탄거리), L2 거리(유클리디안 거리)라 불리고 있다.
18. 마할라노비스 거리는 통계적 개념이 포함된 거리이며 변수들의 산포를 고려하여 이를 표준화한 거리이다. 두 벡터 사이의 거리는 산포를 의미하는 표본 공분산으로 나눠주어야 하며, 그룹에 대한 사전 지식 없이는 표본 공분산 S를 계산할 수 없으므로 사용하기 곤란하다.
19. 각 모형의 상호 연관성이 높을수록 정확도는 떨어진다.
20. k평균군집은 한번 군집이 형성되더라도 다른 군집으로 이동이 가능하다.
21. Start AIC보다 작은 값들 중에 가장 작은 변수를 제외한 나머지 설명변수로 다음 Step을 이어 나간다.
22. income은 p-value 값이 0.71152로 나타나 default를 설명하는데 통계적으로 유의하지 않게 나타나고 있다.
23. 로지스틱 회귀분석의 모형 검정방법은 카이제곱 검정으로 이루어지며, 선형회귀분석은 F-검정, T-검정으로 이루어진다.
24. 주성분 분석은 분산을 최대화하는 차원을 찾는 방법으로 공분산행렬에서 고윳값을 통해 찾을 수 있다.

25. 평균 고유값 방법은 고유값들의 평균을 구한 후 고유값이 평균값 이상이 되는 주성분을 제거하는 것이 아니라 설정하는 것이다.
26. 잔차분석에서 잔차는 독립성, 등분산성, 정규성을 만족해야 한다.
27. 순환변동은 경제적이거나 자연적인 이유 없이 알려지지 않은 주기를 가지고 변화하는 자료를 의미한다.
28. 4번의 설명은 연속형 확률밀도함수에 대한 설명이다.
29. 지지도는 전체 거래 중 항목 A와 항목 B를 동시에 포함하는 거래의 비율이다.
30. 경쟁층은 입력벡터의 특성에 따라 벡터가 한 점으로 클러스터링 되는 층이다.
31. DFFITS는 절대값이 공식에 대입한 값보다 큰 값이 나타나야 높은 영향력으로 간주한다.
32. Durbin-Watson test는 회귀 모형 오차항의 자기상관이 있는지에 대한 검정이다. 히스토그램, Q-Q plot, Shapiro-Wilk 검정 등을 활용하여 데이터의 정규성을 확인한다.
33. 1종 오류는 귀무가설이 옳은데도 귀무가설을 기각하게 되는 오류이다.
34. 이상치는 이상치 자체의 의미가 있을 수 있어 평균으로부터 3*표준편차를 벗어나더라도 제거하면 안된다.
35. (가)는 $\alpha/2$ 인데, 여기서 α 는 0.1이므로 $\alpha/2=0.05$ 이다. (나)에는 표본수이므로 70이다.
36. 사회연결망 분석은 여러 개의 기법들로 구성되어 있고, 국내의 사회연결망 연구에서 많이 활용되고 있는 기법은 중심성, 밀도, 중심화 등이 있다. 그 중 밀도는 연결망 내에서 전체 구성원들이 서로 간에 얼마나 많은 관계를 맺고 있는가를 나타낸다.
37. Weight의 중앙값은 258.00이다.
38. 전체 관측치 수는 50개이다.
39. 안정적 시계열은 시간의 추이와 관계없이 평균, 분산이 불변하여, 변화했다고 해도 다시 평균으로 회귀하는 경향을 보인다.
40. 유의수준 0.05와 p-value를 비교했을 때, p-value 값이 현저하게 작게 나타나 귀무가설이 기각되어 '대학의 평균 교재 구입비용이 570달러와 같다'는 기각된다

【 단답형 】

단답형 01. 정보는 데이터의 가공, 처리와 데이터간 연관관계 속에서 의미가 도출된 것이다.

단답형 02. ERP는 기업내부 데이터베이스 중 기업 전체가 경영자원을 효과적으로 이용하기 위해서 기업의 모든 자원을 통합하여 관리하기 위한 기업 경영정보 시스템이다.

단답형 03. 데이터 거버넌스에서 마스터 데이터, 메타데이터, 데이터 사전은 중요한 관리대상이다.

단답형 04. 시급성은 전략적 중요도와 목표가치(KPI)이 핵심이다.

단답형 05. BIC 값이 가장 큰 지점을 찾았을 때, 4가 가장 크므로 최적의 군집의 수는 4개라고 할 수 있다.

단답형 06. 앙상블 기법은 주어진 자료로부터 여러 개의 예측모형들을 만든 후 예측모형들을 조합하여 하나의 최종 예측 모형을 만드는 방법이다.

단답형 07. 데이터를 크기순으로 나열했을 때, 3사분위수가 92이라고함은 데이터의 75%위치에 있는 값이 92이기 때문에 92보다 작은 값이 75%정도 있을 것이며, 나머지 25%는 92보다 큰 값을 가진다.

단답형 08. 포아송 분포(Poisson Distribution)는 단위 시간 안에 어떤 사건이 몇번 발생할 것인지를 표현하는 이산확률분포이다.

단답형 09. 민감도가 0.8이면 $TP=4FN$ 이다. 실제/예측 TRUE와 실제/예측 FALSE가 100으로 통일이라면 아래와 같이 된다. $TP+FP=100$, $FN+TN=100$, $TP+FN=100$, $FP+TN=100$ 이 중 $TP+FN=100$ 에서 $FN=20$ 이므로 $TP=80$ 이다. 그러면 $FP=20$, $TN=80$ 이므로 Precision을 구하면 $TP/TP+FP=80/100=0.8$ 이다.

단답형 10. 머신러닝은 학습과 개선을 위해 명시적으로 컴퓨터를 프로그래밍하는 대신, 컴퓨터가 데이터로 학습하고 경험을 통해 개선하도록 훈련하는 데 중점을 둔다.