

# ADsP 44회 합격을 위한 출제 키워드 강의

+ + +



# ADsP

1과목

## 데이터 이해

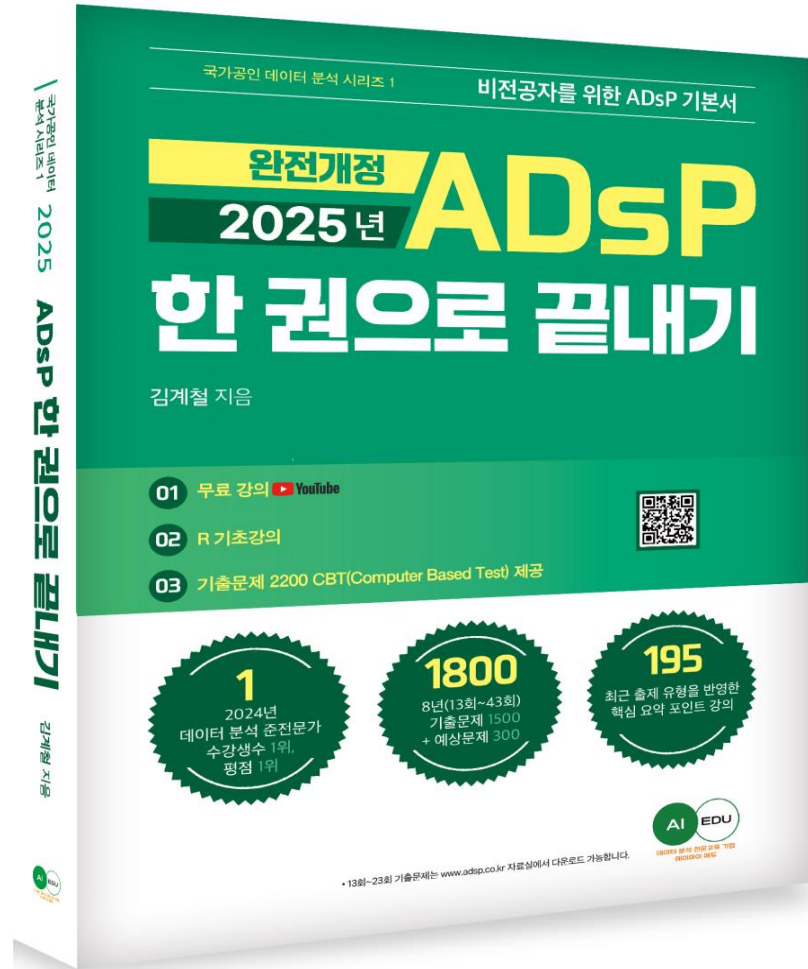
14%

17%

11%

9%

18%



## 출제 예상

### 1 과목 데이터 이해

#### 01장 데이터의 이해

##### 01 데이터와 정보

- ① 정성적 및 정량적 데이터 정의
- ② 암묵지와 형식지 상호작용(공통화→표출화→연결화→내면화)
- ③ DIKW 정의 및 사례 ★

##### 02 데이터 베이스 정의와 특징

- ④ DB vs DBMS(객체지향 DBMS)
- ⑤ 데이터베이스 vs ODS vs 데이터 레이크 vs 데이터웨어하우스 vs 데이터 마트 개념 구분 ★
- ⑥ DB 설계 프로세스, ETL 기능

##### 03 데이터베이스 활용

- ⑦ 기업 내부데이터 베이스 솔루션 7+1 정의 ★
- ⑧ 사회기반 데이터베이스 솔루션

## 출제 예상

### 1 과목 데이터 이해

#### 02장 데이터의 가치와 미래

##### 01 빅데이터의 이해

- ① 빅데이터의 특징 3V
- ② 빅데이터의 본질적인 변화
- ③ 빅데이터 플랫폼 역할

##### 02 빅데이터의 가치와 영향

- ④ 빅데이터의 가치 산정이 어려운 이유
- ⑤ 빅데이터의 영향

##### 03 데이터베이스 활용

- ⑦ 빅데이터 활용 테크닉 ★

## 출제 예상

### 1 과목 데이터 이해

#### 02장 데이터의 가치와 미래

#### 04 위기요인과 통제방안

- ① 빅데이터 시대의 위기요인과 통제방안 ★
- ② 개인정보 비식별화 기술 ★
- ③ 개인정보보호법

#### 05 미래의 빅데이터

- ④ 빅데이터 활용 3요소

## 출제 예상

### 1 과목 데이터 이해

#### 03장 가치 창조를 위한 데이터 사이언스와 전략 인사이트

##### 01 빅데이터 분석과 전략 인사이트

① 전략적 통찰이 없는 분석의 함정 (일차원적 분석 vs 전략 도출 위한 가치 기반 분석) ★

##### 02 전략 인사이트 도출을 위한 필요 역량

② 데이터 사이언스(vs 통계학 vs 데이터마이닝) vs 데이터 사이언티스트(하드 vs 소프트)★

##### 03 빅데이터 그리고 데이터 사이언스의 미래

③ 가치 패러다임의 변화

### 학습퀴즈

01. DIKW는 피라미드 계층적 구성요소 속에서 역할을 수행하며 인간이 사회 활동으로 추구하는 가치 창출을 위한 일련의 프로세스를 말한다.

DIKW의 계층적 구성 요소에 해당되지 않는 것은?

- ① 지혜
- ② 정보
- ③ 데이터
- ④ 아이디어

기출문제 : 13회 1과목 1번

02. 다음 데이터의 유형은 정성적 데이터와 정량적 데이터로 분류가 된다.

아래 보기에서 성격이 다른 것은?

- ① 풍향
- ② 습도
- ③ 기상특보
- ④ 시간당 강수량

기출문제 : 15회 1과목 2번

### 학습퀴즈

03. DIKW 피라미드 계층구조에서 지식(Knowledge)에 해당되는 것은?

- ① A 마트 100원, B 마트는 200원 연필을 판매한다.
- ② A 마트의 연필가격이 더 싸다.
- ③ 상대적으로 저렴한 A마트에서 연필을 사야 겠다.
- ④ A 마트의 다른 상품들도 B마트보다 쌀 것이라고 판단한다.

기출문제 : 15회 1과목 4번

### 학습퀴즈

04. 다음 보기는 암묵지와 형식지의 상호작용에 관한 설명이다.  
알맞게 연결된 것은?

1단계: 암묵적 지식 노하우를 다른 사람에게 알려주는 것 → ( )

2단계: 암묵적 지식 노하우를 책이나 교본 등 형식지로 만드는 것 → ( )

3단계: 책이나 교본(형식지)에 자신이 알고 있는 새로운 지식(형식지)를 추가하는 것 → ( )

4단계: 만들어진 책이나 교본(형식지)을 보고 다른 직원들이 암묵적 지식(노하우)을 습득 → ( )

① 공통화 → 표출화 → 연결화 → 내면화

② 표출화 → 공통화 → 연결화 → 내면화

③ 연결화 → 내면화 → 공통화 → 표출화

④ 내면화 → 공통화 → 연결화 → 표출화

기출문제 : 17회 1과목 7번

### 학습퀴즈

05. 다음 중 아래에서 설명하는 정보나 자료의 집합을 무엇이라 하는가?

다양한 전달 매체에 의한 데이터, 지식 등 인식이 가능한 자료

- ① 암묵지
- ② 형식지
- ③ 데이터베이스
- ④ 콘텐츠

기출문제 : 41회 1과목 2번

### 학습퀴즈

06. 데이터에 대한 설명으로 옳지 않은 것은?

- ① 바이트가 데이터의 최소 단위이며, 이진수(0과1)로 이루어진다.
- ② 데이터는 예측, 추정, 추론 등에 활용된다.
- ③ 데이터는 가공하기 전의 순수한 수치나 기호로 구성된다.
- ④ 수치형 데이터는 텍스트 데이터보다 저장 공간이 더 적게 필요하다.

기출문제 : 40회 1과목 5번

### 학습퀴즈

07. 데이터베이스의 특징 중 적절하지 않은 것은?

- ① 응용프로그램 종속성
- ② 관리 용이
- ③ 무결성 유지
- ④ 중복성 최소화

기출문제 : 40회 1과목 3번

### 학습퀴즈

#### 08. 데이터웨어하우스의 고유 특성이 아닌 것은?

- ① 데이터웨어하우스는 기업 내의 의사결정 지원 애플리케이션을 위한 정보를 제공하는 하나의 통합된 데이터 저장 공간을 말한다.
- ② ETL은 주기적으로 내부 및 외부 데이터베이스로부터 정보를 추출하고 정해진 규약에 따라 정보를 변환한 후에 데이터웨어하우스에 정보를 적재한다.
- ③ 데이터웨어하우스에서 관리하는 데이터들은 시간적 흐름에 따라 변화하는 값을 유지한다.
- ④ 전사적 차원에서 접근하기보다는 재무, 생산, 운영과 같이 특정 조직의 특정 업무 분야에 초점을 두고 있다.

기출문제 : 20회 1과목 4번

### 학습퀴즈

09. 기업 내부 데이터베이스 활용과 관련 없는 것은?

- ① CRM
- ② ERP
- ③ ITS
- ④ KMS

기출문제 : 20회 1과목 2번

10. 고객별 구매이력 데이터 베이스를 분석하여 고객에 대한 이해를 돕고 이를 바탕으로 각종 마케팅 전략에 활용되는 데이터베이스 솔루션은?

- ① ITS
- ② SCM
- ③ CRM
- ④ NEIS

기출문제 : 29회 1과목 3번

## 학습퀴즈

11. 빅데이터가 만들어내는 본질적인 변화가 아닌 것은?

- 가) 사전처리 → 사후처리      나) 양 → 질  
다) 표본조사 → 전수조사      라) 상관관계 → 인과관계

- ① 가, 나
- ② 나, 다
- ③ 가, 다
- ④ **나, 라**

기출문제 : 40회 1과목 8번

## 학습퀴즈

### 12. 빅데이터 활용에 대한 설명으로 틀린 것은?

- ① 고객의 구매 패턴을 분석하여 맞춤형 마케팅을 제공한다.
- ② 신용카드 거래 데이터를 분석하여 이상 거래를 탐지한다.
- ③ 센서 데이터를 활용하여 기계의 이상 징후를 사전에 감지한다.
- ④ 택배 차량 배치에 요인 분석을 사용한다.

기출문제 : 43회 1과목 8번

## 학습퀴즈

13. 빅데이터 활용 기법에 관한 설명으로 옳지 않은 것은?

- ① 군집분석은 개인 신용평가에 활용될 수 있다.
- ② 기계학습은 사용자의 기호를 학습하여 추천 서비스를 제공할 때 활용된다.
- ③ 한국어는 언어 특성상 감정을 정확히 분석하는데 어려움이 있을 수 있다.
- ④ 소셜 네트워크 분석을 통해 개인 맞춤형 광고를 추천하는 것은 가능하다.

기출문제 : 40회 1과목 7번

## 학습퀴즈

14. 빅데이터의 '사생활 침해' 해결 방법에 대한 설명으로 옳은 것은?

- ① **제공자 동의에서 사용자 책임으로 전환한다.**
- ② 알고리즘에 대한 접근권을 보장한다.
- ③ 기존의 책임원칙을 보강하고 강화한다.
- ④ 예측 알고리즘의 활용을 높인다.

기출문제 : 42회 1과목 8번

## 학습퀴즈

15. 전략적 가치기반 분석을 위해 고려해야 할 요소가 아닌 것은?

- ① 핵심적인 비즈니스 이슈에 답을 주는 것이 가치가 있다.
- ② 반드시 복잡한 모델만이 가치기반에서 좋은 결과를 가져오는 것은 아니다.
- ③ 전략적 분석과 통찰력은 빅데이터 분석에서 핵심적 역할을 한다.
- ④ 기존의 성과를 유지하는 것이 가장 중요한 목표이다.

기출문제 : 40회 1과목 1번

### 학습퀴즈

16. 다음 중 데이터 사이언티스트에게 요구되는 소프트 스킬로 적절한 것을 모두 고르시오.

- 가. 이론적 지식
- 나. 분석 기술에 대한 숙련도
- 다. 커뮤니케이션 기술
- 라. 시각화를 활용한 설득력
- 마. 창의적 사고

- ① 가, 나, 다
- ② 가, 다, 라
- ③ 나, 다, 라
- ④ **다, 라, 마**

기출문제 : 41회 1과목 9번

### 학습퀴즈

17. 빅데이터를 빠르고 정확하게 처리해 개인과 기기와 사물들이 맺고 있는 연결을 효과적이고 효율적으로 관리해주기 위한 가치 패러다임을 무엇이라 하는가?

- ① 디지털화(Digitalization)
- ② 연결(Connection)
- ③ 에이전시(agency)
- ④ 알고리즘(Algorithm)

기출문제 : 42회 1과목 10번

## 기출문제

[03회] 다음 중 Cassandra와 관련이 있는 데이터 저장방식은?

- ① RDBMS    ② NoSQL    ③ HDFS    ④ SQL

**해설** MongoDB, Cassandra, Redis NoSQL 저장 방식이다.

[05회] 테이블의 행과 열로 구조화되어 있지는 않으나 스키마 및 메타데이터 특성을 가지고 있으며 주로 XML, HTML, JSON 등의 데이터 유형을 무엇이라 하는가?

- ① Structured Data  
② Semi-Structured Data  
③ Unstructured Data  
④ Unname Structured Data

**해설**

- 반정형 데이터는 구조에 따라 저장된 데이터지만 정형 데이터와 달리 데이터 내용 안에 구조에 대한 설명이 함께 존재한다.
- 그렇기 때문에 데이터 내용에 대한 설명인 구조를 파악하는 파싱 과정이 필요하고, 파일 형태로 저장된다.

## 기출문제

[08회] 다음 중 정량적 데이터와 정성적 데이터에 대한 설명으로 옳지 않은 것은?

- ① 정량적 데이터는 양적 데이터이다.
- ② 정성적 데이터는 질적 데이터이다.
- ③ 정량적 데이터 중 계수 데이터는 범주형 데이터로 변환 가능하다.
- ④ 정성적 데이터 중 변수 데이터는 연속형 데이터로 변환 가능하다.

**해설** 정량적 데이터는 수치로 나타낼 수 있는 양적 데이터이며, 정성적 데이터는 질적 정보를 담고 있다. 정량적 데이터 중 계수 데이터는 범주형 데이터로 변환할 수 있지만, 정성적 데이터는 연속형 데이터로 변환되지 않는다.

## 기출문제

[04회] 데이터 분석을 위해 추출할 때 변환하여 분석할 필요가 없는 원시 데이터셋의 저장소를 무엇이라 하는가?

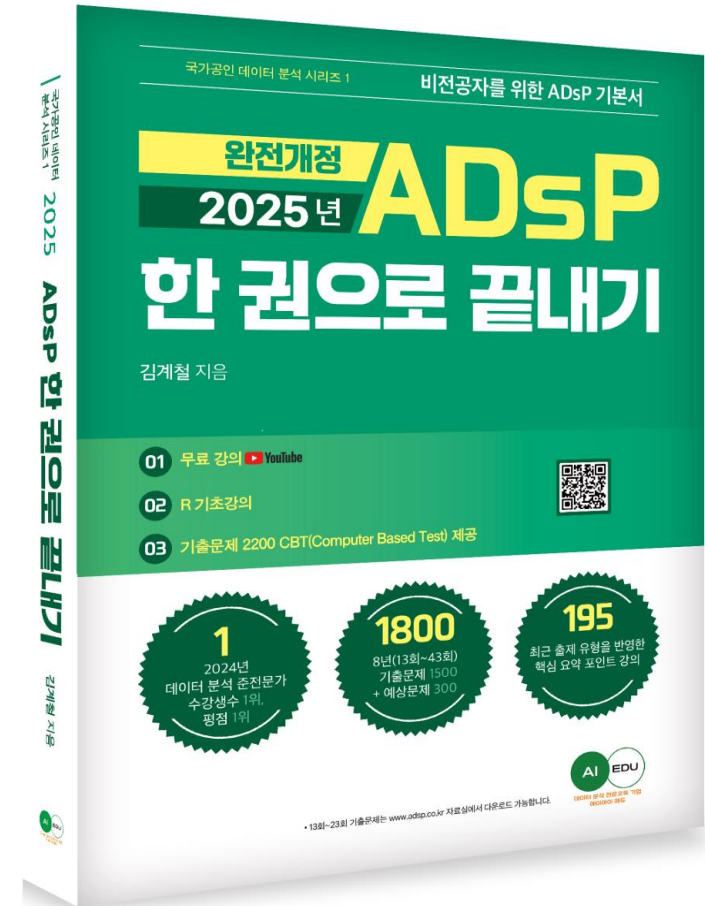
- ① 운영 데이터베이스
- ② 데이터웨어하우스
- ③ 하둡
- ④ 데이터레이크

**해설** 데이터레이크는 정형·비정형 데이터 종류와 모델에 상관없이 모든 유형의 데이터를 저장할 수 있는 중앙 집중식 저장소이다.

# ADsP

2과목

## 데이터 분석기획



### 출제 예상

2과목 데이터 분석 기획

4장 데이터 분석 기획의 이해

01 분석 기획 방향성 도출

- ① 분석 주제 유형 ★
- ② 과제단위와 마스터 플랜 단위
- ③ 분석 기획 시 고려사항- 정형, 반정형 구분/ 데이터 저장방식(RDB vs NoSQL)

02 분석 방법론

- ③ 분석 방법론의 구성요소
- ④ 폭포수(Waterfall) 모델 vs 나선형(Spiral) 모델 vs 프로토타입(Prototype) 모델
- ⑤ KDD vs CRISP-DM vs 빅데이터 분석 방법론(단계, 태스크 스텝) ★

### 출제 예상

2과목 데이터 분석 기획

4장 데이터 분석 기획의 이해

03 분석과제 발굴

- ⑥ 하향식과 상향식, 디자인 사고 ★
- ⑦ 하향식·상향식 접근 방식 프로세스 ★
- ⑧ 하향식 접근방식의 혁신적 관점
- ⑨ 지도학습과 비지도학습 차이
- ⑩ 프로토타이핑 프로세스 필요성
- ⑪ 분석과제 정의서

04 분석 프로젝트 관리 방안

- ⑫ 분석 프로젝트 5가지 주요 특성 ★
- ⑬ 분석 프로젝트 주요 10개 관리 항목

### 출제 예상

#### 2과목 데이터 분석 기획

#### 5장 분석 마스터 플랜

#### 01 마스터 플랜 수립

- ① 마스터 플랜 수립 우선순위 및 적용범위 고려요소
- ② 시급성과 난이도 구분★
- ③ 포트폴리오 사분면 분석을 통한 과제 우선순위를 선정

#### 02 분석 거버넌스 체계 수립

- ④ 분석 거버넌스 체계 구성요소
- ⑤ 분석준비도와 성숙도 ★
- ⑥ 데이터 거버넌스의 구성 요소
- ⑦ 데이터 거버넌스의 체계 요소 ★
- ⑧ 데이터 분석 업무 주체에 따른 유형
- ⑨ 분석 과제 관리 프로세스

## 학습퀴즈

01. 목표 시점별 분석 기획 방안 중 당면한 분석 주제를 해결하는 방법이 아닌 것은?

- ① Quick & Win
- ② Problem Solving
- ③ Accuracy & Deploy
- ④ Speed & Test

기출문제 : 40회 2과목 7번

### 학습퀴즈

02. 분석기획 유형 중 분석의 대상 및 방식에 대한 설명 중 가장 옳지 않은 것은?

- ① 분석대상을 모르고 분석방식을 알면 통찰력을 활용할 수 있다.
- ② 분석방식만 알고 있으면 솔루션의 이용이 가능하다.
- ③ 분석대상과 분석방식을 모두 알고 있는 경우 최적화 또는 예측이 가능하다.
- ④ 분석대상과 분석방식을 모두 모를 경우 탐색을 통한 발견이 필요하다.

기출문제 : 41회 2과목 1번

## 학습퀴즈

03. 분석기획 단계에서 프로젝트 범위 정의서 SOW(Statement of Work)가 작성이 되는 태스크(Task)로 옳은 것은?

- ① 비즈니스 이해 및 범위 설정
- ② 프로젝트 정의 및 계획 수립
- ③ 데이터 스토어 설계
- ④ 분석용 데이터 준비

기출문제 : 40회 2과목 3번

### 학습퀴즈

#### 04. KDD 분석 절차 중 아래 항목에 해당하는 단계는?

- 분석데이터 세트에 포함 되어 있는 이상값, 결측치를 식별하고 필요 시 제거한다.
- 프로세스를 수행하는 과정에서 추가적인 데이터 세트를 필요한 경우 선택한다.

① 데이터 세트 선택(Data Selection)

② 데이터 전처리(Data Preprocessing)

③ 데이터 변환(Data Transformation)

④ 데이터 마이닝(Data Mining)

기출문제 : 41회 2과목 3번

### 학습퀴즈

05. CRISP-DM의 업무이해 단계(Phase)의 주요 내용으로 옳은 것은?

- ① 업무 상황 파악, 데이터 마이닝 목표설정, 프로젝트 계획 수립
- ② 초기 데이터 수집, 데이터 기술 분석, 데이터 탐색, 데이터 품질 확인
- ③ 분석용 데이터셋 선택, 데이터 정제, 데이터 통합, 데이터 포매팅
- ④ 분석 결과 평가, 모델링 과정 평가, 모델 적용성 평가

기출문제 : 42회 2과목 2번

## 학습퀴즈

### 06. 아래에서 설명하는 문제탐색 기법으로 알맞은 것은?

유사. 동종 사례 벤치마킹을 통한 분석기회 발굴은 제공되는 산업별, 업무 서비스별 분석 테마 후보그룹(pool)을 통해 'Quick & Easy' 방식으로 필요한 분석 기회가 무엇인지에 대한 아이디어를 얻고 기업에 적용할 분석 테마 목록을 워크숍 형태의 브레인스토밍 (Brain Storming)을 통해 빠르게 도출하는 방법이다.

- ① 시장니즈의 분석
- ② 역량의 재해석
- ③ 외부참조 모델
- ④ 경쟁자 분석

기출문제 : 41회 2과목 5번

## 학습퀴즈

07. 비즈니스 모델의 고객 영역에 존재하는 현재 고객을 확장하여 전체 시장을 대상으로 사회적, 문화적, 구조적 트렌드 변화에 기반한 분석 기회를 도출하는 거시적 관점의 영역을 무엇이라 하는가?

- ① 환경(Environmental)
- ② 사회(Social)
- ③ 기술(Technological)
- ④ 정치(Political)

기출문제 : 43회 2과목 6번

## 학습퀴즈

08. 상향식 접근 방식에 대한 설명으로 옳지 않은 것은?

- ① 문제정의를 명확히 할 수 있을 때 사용한다.
- ② 데이터 기반의 문제 재정의 및 해결방안 탐색을 지속하며 개선하는 방식을 의미한다.
- ③ 비지도학습(Supervised Learning) 방법에 따라 수행된다.
- ④ 다량의 데이터 분석을 통해서 “왜” 그러한 일이 발생하는지 역으로 추적하면서 문제를 도출하는 접근 방식이다.

기출문제 : 42회 2과목 5번

## 학습퀴즈

09. 시급성과 난이도를 고려했을 때 우선적으로 수행하는 업무는?

- ① 시급성 : 높다, 난이도 : 낮다
- ② 시급성 : 낮다, 난이도 : 낮다
- ③ 시급성 : 높다, 난이도 : 높다
- ④ 시급성 : 낮다, 난이도 : 낮다

기출문제 : 43회 2과목 3번

### 학습퀴즈

10. 분석 조직 인력들을 협업부서로 직접 배치하여 신속한 업무 수행 가능한 구조를 무엇이라 하는가?

- ① 집중구조    ② 기능중심구조    ③ 분산구조    ④ 확산구조

기출문제 : 35회 12번

11. 다음 중 데이터 거버넌스의 구성요소가 아닌 것은?

- ① 원칙    ② 조직    ③ 분석방법    ④ 절차

기출문제 : 35회 18번

### 학습퀴즈

12. 분석 준비도와 분석 성숙도를 고려한 도입형, 확산형, 준비형, 정착형 중에 확산형에 대한 옳은 설명을 고르시오.

- ① 낮은 준비도, 높은 성숙도
- ② 높은 준비도, 낮은 성숙도
- ③ 낮은 준비도, 낮은 성숙도
- ④ 높은 준비도, 높은 성숙도

기출문제 : 43회 2과목 10번

## 기출문제

[06회] 다음 중 기업의 분석 수준 진단에 대한 서술로 옳지 않은 것은?

- ① 준비형 : 데이터 분석을 위한 낮은 준비도와 낮은 숙성도
- ② 정착형 : 조직 및 인력, 분석 업무, 분석 기법을 내부에 오픈
- ③ 확산형 : 6가지 분석 구성요소를 모두 갖춰 지속적 확산이 가능
- ④ 도입형 : 업무 기법은 충분하나, 조직 인력이 부실

**해설** 도입형은 도입되어 있는 분석 업무 및 분석 기법은 부족하지만, 조직 및 인력에 대한 준비도가 높은 수준을 의미한다.

## 기출문제

[06회] 데이터 분석 조직에 관한 설명으로 옳지 않은 것은?

- ① 기능형은 전사의 핵심 업무를 분석하지 못한다.
- ② 집중구조는 별도의 분석 조직이 존재하므로 협업 조직과의 업무 중복 가능성이 없다.
- ⑤ 분산구조는 전담 조직 인력을 협업부서에 배치하므로 신속한 업무에 적합하다.
- ④ 기능형은 별도로 분석조직이 없다.

**해설** 집중구조는 각 부서별로 분석을 진행하며 전사 분석 업무를 별도의 분석 전담 조직에서 담당하기 때문에 분석 업무가 이중화, 이원화 될 수 있다.

## 기출문제

[06회] 분석준비도의 진단 영역으로 옳지 않은 것은?

- ① 분석 성과 평가
- ② 분석 업무
- ③ 분석 인력
- ④ 분석 기법

**해설** 분석준비도는 빅데이터 분석 업무를 도입하기 위한 조직의 분석 업무의 수준을 파악하는 지표로 분석 업무 및 파악, 인력 및 조직, 분석 기법, 분석 데이터, 분석 문화, IT 인프라가 있다.

## 기출문제

[06회] 다음 중 데이터 거버넌스의 구성요소에 해당하지 않는 것은?

- ① 원칙
- ② IT 인프라
- ③ 프로세스
- ④ 조직

**해설** 데이터 거버넌스는 전사 차원의 데이터에 대한 정책과 지침, 운영조직 및 책임 등 표준화된 관리 체계를 수립하고 운영하는 기반을 마련하는 것으로 원칙, 조직, 프로세스로 구성되어 있다.

## 기출문제

[08회] 다음 보기에서 설명하고 있는 내용으로 가장 적절한 것은?

수집한 데이터를 저장, 처리하고 분석할 수 있도록 포괄적으로 지원한다.

- ① 빅데이터 마이닝
- ② 빅데이터 플랫폼
- ③ 빅데이터 처리기술
- ④ 빅데이터 탐색기술

**해설** 데이터의 수집, 저장, 처리, 분석이 수행되는 계층은 플랫폼 계층이다.

## 기출문제

[03회] 개인정보처리자의 의무사항과 관련이 없는 것은?

- ① 개인정보를 수집할 때에는 그 목적에 필요한 범위에서 최소한의 개인정보만을 수집해야 한다.
- ② 반드시 익명으로 처리해야 한다.
- ③ 개인의 사생활 침해가 최소화되는 방법으로 처리한다.
- ④ 개인정보처리자는 특별한 경우를 제외하고 고유식별정보를 처리할 수 없다.

**해설** 익명으로 처리가 가능한 경우 익명으로 처리

## 기출문제

[07회] 다음 중 데이터 3법과 관련이 없는 것은?

- ① 개인정보보호법
- ② 정보통신망 이용 촉진 및 정보보호 등에 관한 법률
- ③ 신용정보의 이용 및 보호에 관한 법률
- ④ 공공 데이터 제공 및 이용 활성화에 관한 법률

**해설** 데이터 3법은 개인정보보호법, 정보통신망법, 신용정보 법이다.

## 기출문제

[07회] 분석 기획 단계에서 분석 과제의 우선순위를 정할 때 고려할 요소가 아닌 것은?

- ① 전략적 중요도
- ② 비즈니스 성과
- ③ 분석 데이터 적용 수준
- ④ 실행 용이성

**해설** 분석 마스터 플랜을 수립할 때는 전략적 중요도 비즈니스 성과, 실행 용이성을 고려하여 우선순위를 정한다. 우선순위가 설정되면 업무 내재화 적용 수준, 분석 데이터 적용 수준, 기술 적용 수준을 고려하여 적용 범위 및 방식을 정한 뒤 분석 로드맵을 수립한다.

## 기출문제

[02회] 다음 중 비즈니스 모델 분석의 상향식 접근방식의 순서는?

- ① 분석요건 식별→분석요건 정의→프로세스 분류→프로세스의 흐름 분석
- ② 프로세스 분류→프로세스의 흐름 분석→분석요건식별→분석요건 정의
- ③ 프로세스의 흐름 분석→분석요건식별→분석요건 정의→프로세스 분류
- ④ 분석요건 정의→프로세스 분류→프로세스의 흐름 분석→분석요건 식별

## 해설

상향식 접근 방식

하위단위 비즈니스 프로시저, 프로세스를 통해 비즈니스를 분석하여 분석요건을 식별하고 정의한다.

## 기출문제

[05회,7회] 빅데이터 분석 방법 프로세스에서 WBS 작성되는 태스크(Task)로 맞는 것은?

- ① 비즈니스 이해범위 및 범위설정
- ② 프로젝트 정의 및 계획수립
- ③ 프로젝트 위험계획 수립
- ④ 필요데이터 준비

**해설** 프로젝트 정의 및 계획수립 태스크 주요 내용

- 프로젝트의 목표 및 KPI(핵심성과지표), 목표수준 등을 구체화하여 상세 프로젝트 정의서를 작성
- 프로젝트 수행계획서를 작성하는 단계로서 프로젝트의 목적 및 배경, 기대효과, 수행방법, 일정 및 추진조직, WBS를 작성한다.

## 기출문제

[07회] 빅데이터 분석 기획에서 하지 않는 것은?

- ① 비즈니스 이해 및 범위 설정
- ② 프로젝트 정의 및 계획 수립
- ③ 프로젝트 위험 계획 수립

④ 분석 데이터 준비 및 알고리즘 후보군 탐색

**해설** 빅데이터 분석 방법론의 분석 기획 단계에서는 비즈니스 이해 및 범위 설정, 프로젝트 정의 및 계획 수립, 프로젝트 위험 계획 수립을 진행한다.

## 기출문제

[05회] 개인정보 비식별화 기술에 대한 설명 중 적절하지 않은 것은?

- ① 가명처리 : 개인 식별 정보에 중요한 데이터를 다른 데이터로 변경
- ② 총계처리 : 데이터 총합으로 처리해 개별 데이터로 보이지 않게 처리
- ③ 범주화 : 주요 식별 정보 일부 삭제
- ④ 데이터 마스킹 : 주요 식별 정보를 ○○, \*\* 처리

**해설** 범주화는 개인식별정보에 대한 수치데이터를 임의의 기준의 범위로 설정하는 기법

[03회] 다음 중 재현자료(Synthetic Data)에 대한 설명 중 옳은 것은?

- ① 통계적 원본 보존
- ② 재현된 부분의 원자료가 남아 있지 않아 개인정보 노출 위험이 높지 않다.
- ③ 재현데이터는 비민감정보의 경우 정보 노출의 위험성이 없다.
- ④ 샘플수를 무한대로 증가할 수 없다.

**해설** 재현자료는 원자료와 다르지만 원자료와 동일한 분포를 따르도록 통계적으로 생성한 자료이다.

## 기출문제

[07회] 데이터의 일부를 공백 처리하거나 노이즈를 추가하는 비식별화 기술로 알맞은 것은?

- ① 가명처리
- ② 데이터 삭제
- ③ 데이터 범주화
- ④ 데이터 마스킹

**해설** 데이터 마스킹은 전체 또는 부분을 대체 값으로 변환하는 기법으로 임의 잡음 추가, 공백, 대체 방법이 있다.

## 기출문제

[08회] 다음 중 보기에서 설명하고 있는 비식별화 기법과 세부 기술로 옳은 것은?

사용자에 대한 정보를 뒤섞어 정보의 손실 없이  
특정 개인에 대한 추측을 할 수 없도록 한다.

## ① 총계처리 - 재배열

- ② 데이터 마스킹 - 잡음 추가
- ③ 가명처리 - 휴리스틱 익명화
- ④ 데이터 범주화 - 랜덤 라운딩

## 해설

설명에서 '사용자에 대한 정보를 뒤섞어 정보의 손실 없이 특정 개인에 대한 추측을 할 수 없도록 한다'는 총계처리 방식의 재배열을 나타낸다. 재배열은 데이터를 집계하여 개인 식별이 불가능하게 만드는 방법 중 하나이다.

# ADsP

3과목

## 데이터 분석

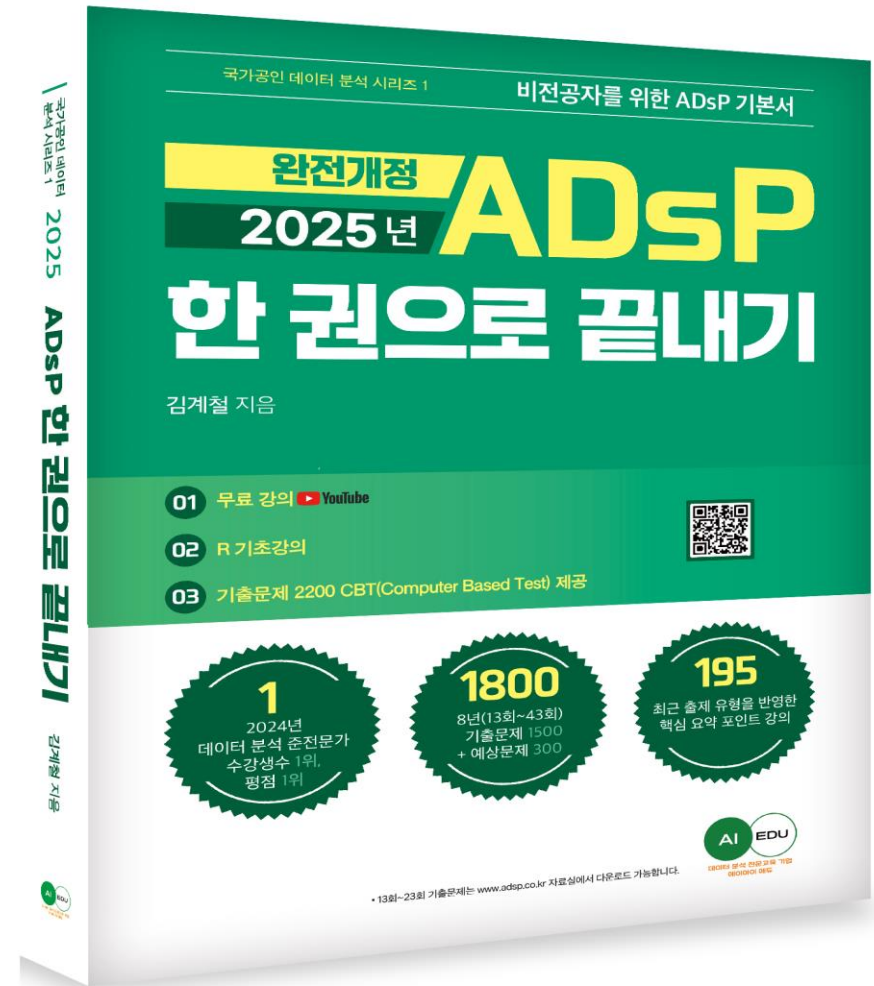
14%

17%

11%

9%

3%



## 출제 예상

### 3과목 데이터 분석

#### 6장 R 기초와 데이터 매트

##### 01 R기초

- ① 데이터 형태(벡터 vs 행렬 vs 데이터 프레임 vs 리스트)
- ② 상자그림(boxplot)과 히스토그램 해석 ★
- ③ 데이터 비대칭 분포(중앙값과 평균 크기 비교)

##### 02 데이터 매트

- ④ reshape, plyr 패키지

##### 03 결측값 처리와 이상값 검색

- ⑤ 단순대치법(완전분석법, 평균, 중앙값, 최빈값, 단순확률, 회귀)★
- ⑥ summary( )해석 ★
- ⑦ 단변량 이상치 검색 (ESD)

## 출제 예상

### 3과목 데이터 분석

#### 7장 통계분석

##### 01 통계학 개론

- ① 확률적 표본 추출 ★
- ② 자료의 척도 ★
- ③ 조건부 확률(독립 vs 종속)
- ④ 베이지안 정리
- ⑤ 이산확률과 연속확률 기댓값 ★
- ⑥ 이산확률분포(이항,포아송,초기하) 연속확률분포(z분포,t분포,카이제곱,F분포) ★
- ⑦ 중심극한정리 개념
- ⑧ 구간추정과 신뢰수준 의미 ★
- ⑨ 제1종 오류와 제2종 오류 ★
- ⑩ 유의확률과 유의수준 ★
- ⑪ 모수와 비모수 특징

## 출제 예상

3과목 데이터 분석

7장 통계분석

02 기초 통계분석 ★★

- ① 회귀분석의 가정
- ② 잔차분석
- ③ 다중회귀분석 해석
- ④ 최적회귀방정식(전진 vs 후진)
- ⑤ 다중공선성
- ⑥ 다중회귀 vs 다항회귀
- ⑦ 회귀분석의 분산분석표

## 출제 예상

3과목 데이터 분석

7장 통계분석

03 다변량분석

- ① 상관계수 정의 및 유의성 검정
- ② 피어슨 vs 스피어만 ★
- ③ 다차원척도
- ④ 주성분 분석 개념 및 결과해석 ★

## 출제 예상

3과목 데이터 분석

7장 통계분석

04 시계열 예측

- ① 정상성 ★
- ② 모형 식별(AR,MA,ARIMA) ★
- ③ 이동평균법 vs 지수평활법
- ④ 분해시계열 ★
- ⑤ 비정상 시계열 전환

## 출제 예상

### 3과목 데이터 분석

### 8장 정형 데이터마이닝

#### 01 데이터마이닝

- ① 데이터 마이닝 프로세스
- ② 데이터 마이닝 기능

#### 02 모형평가

- ③ 오분류표를 활용한 평가지표 ★
- ④ ROC Curve ★
- ⑤ 이익도표와 향상도 곡선
- ⑥ 교차검증(홀드아웃 vs k-fold vs 붓스트랩) ★

## 출제 예상

### 3과목 데이터 분석

### 8장 정형 데이터마이닝

#### 03 분류분석

- ① 로지스틱 회귀분석 vs 일반회귀분석
- ② 인공신경망 역전파 알고리즘
- ③ 활성화 함수 역할
- ④ 의사결정나무의 불순도 측정 지표
- ⑤ 분류나무 vs 회귀나무
- ⑥ CHAID vs Cart
- ⑦ 의사결정나무 특징
- ⑧ 앙상블 모형(배깅 vs 부스팅 vs 랜덤포레스트 vs 스택킹)

## 출제 예상

3과목 데이터 분석

8장 정형 데이터마이닝

04 군집분석

- ① 계층 군집 vs 비계층 군집
- ② k-means 프로세스 및 단점
- ③ EM 알고리즘
- ④ SOM vs ANN
- ⑤ 밀도기반군집
- ⑥ 실루엣 계수, 엘보우 기법, 덴드로 그래프

## 출제 예상

3과목 데이터 분석

8장 정형 데이터마이닝

05 연관분석

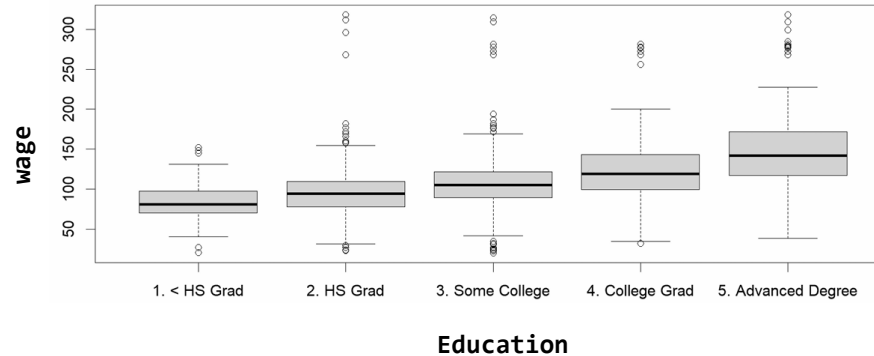
- ① 지지도 vs 신뢰도 vs 향상도
- ② 연관규칙의 장점과 단점
- ③ apriori 알고리즘
- ④ 순차패턴분석

## 학습퀴즈

### 01. 아래 분석 결과의 해석으로 옳지 않은 것은?

summary(wage)

Education		wage
1. < HS Grad	: 268	Min. : 20.09
2. HS Grad	: 971	1st Qu. : 85.38
3. Some College	: 650	Median : 104.92
4. College Grad	: 685	Mean : 111.70
5. Advanced Degree	: 426	3rd Qu. : 128.68
		Max. : 318.34



- ① 'wage' 변수는 오른쪽 꼬리 분포로 추정할 수 있다.
- ② 'HS Grad' 수는 971명이다.
- ③ 'wage'의 상위 25%는 128.68보다 큰 'wage'를 받는다.
- ④ 'education'의 최소값은 20.09이다.

기출문제 : 40회 3과목 5번

## 학습퀴즈

02. 결측값 처리 방법 중 완전 사례 분석 (Complete Case Analysis)에 대한 설명으로 옳지 않은 것은 무엇인가?

- ① 결측값이 있는 자료를 모두 제거하는 방법이다.
- ② 결측값이 적은 경우 효율적으로 사용할 수 있다.
- ③ 결측값이 많은 경우 데이터 손실로 이어질 수 있다.
- ④ 결측값을 변수의 평균값으로 대체하는 방법이다.

기출문제 : 41회 3과목 1번

## 학습퀴즈

03. 5개의 오렌지 나무(tree)의 연령(age)과 둘레(circumference)를 측정된 자료이다. 다음 중 자료에 대한 설명으로 가장 옳지 않은 것은?

```
> head(Orange)
```

Tree	age	circumference
1 1	118	30
2 1	484	58
3 1	664	87
4 1	1004	115
5 1	1231	120
6 1	1372	142

```
> summary(Orange)
```

Tree	age	circumference
3 : 7 Min.	: 118.0	Min. : 30.0
1 : 7 1st Qu.	: 484.0	1st Qu. : 65.5
5 : 7 Median	: 1004.0	Median : 115.0
2 : 7 Mean	: 922.1	Mean : 115.9
4 : 7 3rd Qu.	: 1372.0	3rd Qu. : 161.5
	Max. : 1582.0	Max. : 214.0

① 관측값은 6개이다.

② 나이의 평균값은 922.1이다.

③ tree는 명목척도이다.

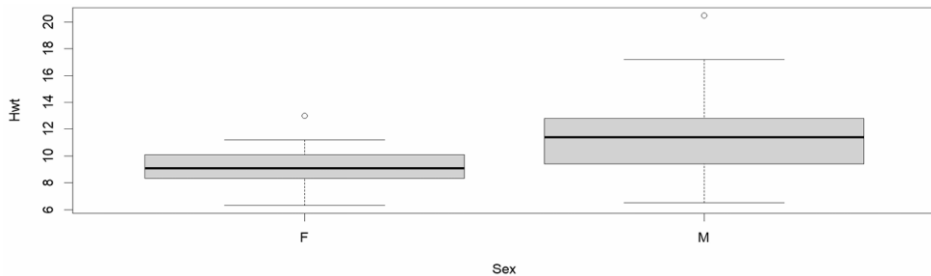
④ 나무 둘레의 50%는 115.0 이상이다.

기출문제 : 41회 3과목 8번

## 학습퀴즈

### 04. 데이터셋 결과에 대한 해석으로 틀린 것은?

```
> summary(cats)
Sex      Bwt      Hwt
F : 47  Min.   : 2.000  Min.   : 6.30
M : 97  1st Qu.: 2.300  1st Qu.: 8.95
      Median : 2.700  Median : 10.10
      Mean   : 2.724  Mean   : 10.63
      3rd Qu.: 3.025  3rd Qu.: 12.12
      Max.   : 3.900  Max.   : 20.50
```



- ① 전체 데이터 수는 144개이다.
- ② 문자 또는 문자열 변수가 존재한다.
- ③ 이상값이 존재하고, 결측치가 1개 있다.
- ④ Hwt의 평균값은 10.63이다.

기출문제 : 43회 3과목 13번

## 학습퀴즈

05. 확률에 대한 설명으로 가장 적합하지 않은 것은?

- ① 모집단 내에서 특정 사건이 발생할 확률은 0과 1 사이의 값이다.
- ② 모든 가능한 사건의 집합을 고려하면 그 확률의 합은 항상 1이다.
- ③ 두 사건 A와 B가 독립일 때, 결합 확률은 각 사건의 확률의 합으로 계산한다.
- ④ 표본공간의 단일 원소일 때 이를 근원사건(Elementary Event)이라 한다.

기출문제 : 40회 3과목 4번

## • 두 사건의 독립성과 종속성일 때, 교집합(결합확률)의 확률

### ① 독립 사건(Independent Events)

- 한 사건의 발생 여부가 다른 사건의 발생에 영향을 미치지 않음
- 독립 사건의 교집합 확률

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A|B) = P(A)$$

### ② 종속 사건(Dependent Events)

- 한 사건의 발생 여부가 다른 사건의 발생에 영향을 미침
- 종속 사건의 교집합 확률

$$P(A \cap B) = P(A) \cdot P(B|A)$$

$$P(A|B) \neq P(A)$$

## 4. 베이즈 정리

- 베이즈 정리는 어떤 사건이 발생한 후에, 그 사건이 특정 조건에 의해 발생했을 가능성 (사후확률)을 계산 하는 방법이다.
- **공식**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- $P(A|B)$  : 사후 확률(B가 발생한 후 A가 일어날 확률)
- $P(A)$  : 사전 확률(A가 일어날 초기 확률)
- $P(B|A)$  : 우도(A가 일어났을 때 B가 관측될 확률)
- $P(B)$  : 증거 확률(B가 관측될 확률)

## 학습퀴즈

6. 점추정과 구간추정에 대한 설명으로 틀린 것은?

- ① 점추정은 모집단의 모수를 하나의 값으로 추정하는 방법이다.
- ② 구간추정은 모집단의 모수가 특정 구간에 포함될 확률을 제시한다.
- ③ 신뢰구간이 항상 모수를 포함한다.
- ④ 점추정과 구간추정 모두 모집단의 모수를 추정하는 방법이다.

기출문제 : 43회 3과목 11번

### 학습퀴즈

#### 07. Chickweight 데이터 셋을 활용한 회귀분석의 결과 해석으로 옳은 것은?

```
Call :  
lm(formula = weight ~ Time, data = Chick)  
Residuals :  
  Min       1Q   Median       3Q      Max  
-14.3202 -11.3081  -0.3444  11.1162  17.5346  
Coefficients :  
              Estimate Std. Error  t value Pr(>|t|)  
(Intercept)  24.4654    6.7279    3.636   0.00456 **  
Time         7.9879    0.5236   15.255  2.97e-08 ***  
--  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
Residual standard error: 12.29 on 10 degrees of freedom  
Multiple R-squared: 0.9588, Adjusted R-squared: 0.9547  
F-statistic: 232.7 on 1 and 10 DF, p-value: 2.974e-08
```

- ① 결정계수가 0.00456로 유의수준 0.05보다 크므로 회귀모형이 유의하다.
- ② 독립변수는 weight, 종속변수는 time이다.
- ③ Time이 1단위 증가할 때, weigh가 7.99만큼 증가한다.
- ④ 유의수준 5%하에서 Time 변수는 통계적으로 유의하지 않다.

기출문제 : 40회 3과목 28번

### 학습퀴즈

08. 선형 회귀 모형에서 오차항이 만족해야 하는 가정으로 옳은 것은 무엇인가?

- ① 등분산성, 정규성, 선형성
- ② 등분산성, 정규성, 독립성
- ③ 표준성, 신뢰성, 정확성
- ④ 독립성, 정확성, 신뢰성

기출문제 : 41회 3과목 9번

### 학습퀴즈

#### 09. 회귀방정식 탐색 방법 중 잘못된 것은?

- ① 후진 제거법(Backward Elimination)은 상수항만 남을 때까지 제거한다.
- ② 전진 선택법 (Forward Selection)은 상수항에서 시작하여 유의미한 변수를 하나씩 추가한다.
- ③ 단계별 방법(Stepwise Selection)은 전진 선택법과 후진 선택법을 결합하여 변수를 추가하거나 제거한다.
- ④ 최적의 회귀방정식을 찾으면 변수 간 관계를 가장 잘 설명할 수 있는 모델을 구성할 수 있다.

기출문제 : 42회 3과목 7번

### 학습퀴즈

10. 다중공선성에 대한 설명으로 옳지 않은 것은?

- ① VIF(Variance Inflation Factor)로 확인되며, 10을 넘기면 다중공선성의 가능성이 있다.
- ② 다중공선성을 제거하려면 독립변수를 더 추가해야 한다.
- ③ 다중공선성은 독립변수들 사이의 선형관계에서 발생한다.
- ④ 다중공선성이 있는 경우에도 통계적으로 유의미한 결과가 나올 수 있다.

기출문제 : 43회 3과목 2번

### 학습퀴즈

11. 주성분 개수 선택 방법에 대한 설명으로 가장 적합하지 않은 것은?

- ① 전체 변이 공헌도를 기준으로 누적 기여율이 70~90%에 도달하는 주성분의 개수를 선택한다.
- ② 고유값이 평균값 이상인 주성분을 선택하는 방법을 적용할 수 있다.
- ③ 스크리 플랏(Scree Plot)을 활용하여 변화가 완만해지는 지점의 주성분 개수를 선택할 수 있다.
- ④ 전체 변이 공헌도는 다른 방법보다 항상 더 유효한 기준이다.

기출문제 : 40회 3과목 2번

### 학습퀴즈

12. 다음 중 시계열 분석에 활용되는 ARIMA(p,d,q)모형에 대한 설명으로 옳은 것은?

- ① ARIMA(p,d,q) 모형의 차수 q는 차분과 관련이 있다.
- ② ARIMA(p,d,q) 모형의 차수 p는 MA모형과 관련이 있다.
- ③ ARIMA(p,d,q) 모형을 사용할 때는 정상성 여부를 확인할 필요는 없다.
- ④ 차수p가 0인 모형을 IMA(d,q) 모형이라고 한다.

기출문제 : 40회 3과목 19번

### 학습퀴즈

13. 시계열 데이터에서 정상성(Stationarity)에 대한 설명 중 옳은 것은?

- ① 정상성을 갖는 시계열 데이터는 시간에 따라 평균이 증가한다.
- ② 해당시계열 확률분포의 모수가 시점에 의존하지 않는다.
- ③ 정상성을 띄면 데이터의 이상치가 없음을 의미한다.
- ④ 공분산은 시점에 의존한다.

기출문제 : 40회 3과목 24번

## 학습퀴즈

14. 시계열 분석에 대한 설명 중 옳지 않은 것은?

① AR의 PACF는 절단면 이후 급격히 감소하지 않고 ACF는 시간이 지남에 따라

점차적으로 감소한다.

② MA의 PACF는 시간이 지남에 따라 감소하고 ACF는 절단면 이후 급격히 감소한다.

③ ARMA의 ACF와 PACF 모두 점차적으로 감소하며 절단 현상이 없다.

④ AR 모델은 자기회귀 변수를 사용하고, PACF는 특정 시차 이후 급격히 감소하여 절단되는 현상을 보인다.

기출문제 : 41회 3과목 13번

## 학습퀴즈

15. 아래 오분류표(혼동행렬, Confusion Matrix)를 사용하여 재현율을 정의하면?

		예측값	
		TRUE	FALSE
실제값	TRUE	TP	FN
	FALSE	FP	TN

①  $TP/(TP+FN)$

②  $TP/(TP+FP)$

③  $TP/FP$

④  $FP/TN$

기출문제 : 41회 3과목 22번

## 학습퀴즈

16. 의사결정나무의 분리 기준으로 사용하는 엔트로피 지수의 계산식을 올바르게 나타낸 것은?

①  $1 - \sum_{i=1}^n p_i^2$

②  $1 - \max(p_i)$

③  $-\sum_{i=1}^e p_i \times \log_2(p_i)$

④  $\sum \frac{(O_i - E_i)^2}{E_i}$

기출문제 : 40회 3과목 3번

## 학습퀴즈

17. 앙상블 모형(Ensemble Model)에 대한 설명으로 올바른 것은?

- ① 앙상블 모형은 단일 모델을 여러 번 반복하여 동일한 데이터를 예측하는 방법이다.
- ② 대표개념으로 배깅(Bagging)과 부스팅(Boosting)이 있다.
- ③ 앙상블 모형은 항상 단일 모델보다 예측 성능이 떨어진다.
- ④ 랜덤 포레스트는 과적합을 방지하기 위해 트리의 깊이를 무제한으로 설정한다.

기출문제 : 40회 3과목 12번

## 학습퀴즈

18. 다음 중 계층적 군집분석(Hierarchical Clustering)에 대한 설명으로 옳은 것은 무엇인가?

- ① 계층적 군집분석은 군집의 수를 미리 정한 상태에서 개체를 포함해가는 방법이다.
- ② 계층적 군집분석에서 군집들은 중복되지 않는 부분집합으로 구성된다.
- ③ K-Means와 K-Medoids는 계층적 군집분석의 대표적인 알고리즘이다.
- ④ 계층적 군집분석은 분할적 군집분석(Partitional Clustering)에 속한다.

기출문제 : 41회 3과목 12번

## 학습퀴즈

19. 군집분석에 대한 설명 중 옳지 않은 것은?

- ① 계층적 군집분석은 덴드로그램의 형태로 표현된다.
- ② 비지도 학습이다.
- ③ 데이터에 명확한 기준이 없으면 적용하지 못한다.
- ④ 계층적 군집화에서는 일단 군집이 형성되면 다른 군집으로 이동할 수 없다는 제약이 있다.

기출문제 : 41회 3과목 27번

## 학습퀴즈

20. apriori 알고리즘에 대한 설명으로 옳은 것은?

- ① 연관 규칙을 도출하기 위해 데이터에서 모든 조합을 계산한다.
- ② 구조화된 트리를 사용하여 빈발 항목 집합을 효율적으로 추출한다.
- ③ 사전에 정의된 최소 지지도를 가정으로 한다.
- ④ 빈발 항목 집합을 찾고 유의미한 연관 규칙만을 탐색한다.

기출문제 : 42회 3과목 27번

## 학습퀴즈

21. 연관분석에 대한 설명 중 옳지 않은 것은?

- ① 연관분석 결과는 이해하기 쉽다.
- ② 거래 횟수(거래량)가 적은 품목의 규칙은 쉽게 발견할 수 있다.
- ③ 거래량이 감소했을 때 충분한 표본 크기를 확보하는 것이 중요하다.
- ④ 데이터를 변환하지 않아도 된다.

기출문제 : 43회 3과목 17번

## 기출문제

[02회] 입력값에 대한 라벨 데이터를 학습시켜 새로운 입력값에 대해 예측하는 방법으로 주로 분류, 회귀문제에 활용하는 분석 방법은?

- ① 지도학습                      ② 강화학습
- ③ 전이학습                      ④ 비지도학습

## 해설

- 지도학습은 말 그대로 정답이 있는 데이터를 활용해 데이터를 학습시키는 것. 입력 값(X data)이 주어지면 입력값에 대한 Label(Y data)을 주어 학습시키며, 대표적으로 분류, 회귀 문제가 있다.
- 강화학습은 지도학습이나 비지도학습으로 달성불가능한 과제를 강화학습으로 해결, 결과값 대신 보상으로부터 학습
- 강화학습을 쉽게 이해할 수 있는 사례
  - 굵긴 쥐를 상자에 넣는다.
  - 쥐는 돌아다니다가 우연히 상자 안에 있는 지렛대를 누른다.
  - 지렛대를 누르자 먹이가 나온다.
  - 이 과정을 반복하면서 쥐는 지렛대를 누르면 먹이를 먹을 수 있다는 것을 학습한다.

## 기출문제

[05회] 인공지능, 머신 러닝, 딥 러닝의 포함 관계 중 옳은 것은?

- ① 머신 러닝>인공지능>딥 러닝
- ② 머신 러닝>딥 러닝>인공지능
- ③ 인공지능>딥 러닝>머신 러닝
- ④ 인공지능>머신 러닝>딥 러닝

## 해설

- 인공지능이 가장 큰 개념으로 머신 러닝 개념을 포함한다.
- 인공지능(Artificial intelligence)은 인간의 학습 능력과 추론 능력, 지각 능력 등을 컴퓨터 프로그램으로 실현한 기술을 말하는데 그 연구 분야 중 하나가 바로 머신 러닝이다.
- 그리고 딥러닝은 인공신경망을 이용한 머신 러닝의 한 종류로 머신 러닝의 하위 개념으로 볼 수 있다.

## 기출문제

[08회] 다음 중 지도학습 모델 선정 시 고려요소로 옳지 않은 것은?

- ① 데이터
- ② 분석 목적
- ③ 자기상관성
- ④ 변수의 중요도

**해설** 지도학습 모델에서는 데이터, 분석 목적, 변수의 중요도 등이 고려되지만, 자기상관성은 지도학습 모델 선정 시 필수 요소가 아니다. 자기상관성은 주로 시계열 분석에서 고려된다.