

• 더 멋진 내일 Tomorrow 을 위한 내일 My Career •

# 내일은



Jade Jeong 지음

## 데이터분석 준전문가 (ADsP)

[부록] 제41회 기출복원문제



< 무료특강 바로가기  
제41회 기출복원문제 해설 저자 직강 무료특강을 제공합니다.



# 제41회 기출복원문제

제한시간: 90분  
문항당 2점  
시행일: 2024. 05. 11

## 1과목 데이터의 이해(10문제)

### 01 다음 보기에서 설명하는 용어로 옳은 것은?

- 데이터 사이언티스트가 한 일로 인해 부당하게 피해가 발생하는 것을 막기 위해 필요한 전문 인력
- 컴퓨터, 수학, 통계학뿐 아니라 비즈니스 전반에 대한 이해와 알고리즘 코딩 해석을 통해 빅데이터 알고리즘에 의해 부당하게 피해를 입은 사람을 구제하는 전문직 종사자

- ① 데이터 공학자
- ② 빅데이터 분석가
- ③ 알고리즘미스트
- ④ 데이터 사이언티스트

### 02 다음 중 빅데이터가 만들어 낸 변화의 양상으로 옳지 않은 것은?

- ① 사전처리에서 사후처리방식으로 변화했다.
- ② 양보다 질로 전환되고 있다.
- ③ 인과관계에서 상관관계 분석을 활용한다.
- ④ 표본조사에서 전수조사가 가능해졌다.

### 03 다음 중 빅데이터 시대의 위기 요인과 통제방안의 연결이 바르게 연결되지 않은 것은?

- ① 사생활 침해 - 개인정보 '제공자의 동의'를 통해 해결해야 한다.
- ② 책임원칙 훼손 - 특정인의 '성향'에 따라 처벌하는 것이 아닌 '행동 결과'를 보고 처벌해야 한다.
- ③ 데이터 오용 - 알고리즘에 대한 접근권을 허용한다.
- ④ 사생활 침해 - 사용 주체가 적극적인 보호 장치를 마련할 수 있도록 한다.

### 04 다음 보기에서 설명하는 기업 내의 데이터베이스 솔루션은?

- 고객별 구매 이력 데이터베이스를 분석하여 고객에 대한 이해를 돕고 이를 바탕으로 각종 마케팅 전략을 펼치는 것이다.
- 일반적으로 자재구매 데이터, 생산·재고 데이터, 유통·판매 데이터, 고객 데이터로 구성된다.

- ① SCM
- ② CRM
- ③ BAI
- ④ ERP

**05** 다음 중 빅데이터의 가치 창출 방식이 기업, 정부, 개인에 미치는 영향으로 적절하지 않은 것은?

- ① 정부: 전체로서는 GDP가 올라가는 효과
- ② 개인: 적시에 필요한 정보를 얻어 기회비용 절약
- ③ 기업: 강한 경쟁력을 확보
- ④ 정부: 미래 의제를 도출

**06** 다음 중 보기에서 설명하는 용어로 옳은 것은?

- 최적화가 필요한 문제의 해결책을 자연선택, 돌연변이 등과 같은 메커니즘을 통해 점진적으로 진화(Evolution)시켜 나가는 방법
- 최대의 시청률을 얻으려면 어떤 프로그램을 어떤 시간대에 방송해야 하는지 여부를 예로 들 수 있다.

- ① 유전 알고리즘
- ② 기계 학습
- ③ 연관 규칙 학습
- ④ 회귀 분석

**07** 다음 중 빅데이터의 가치 패러다임 변화 순서로 옳게 정렬된 것은?

- ① Digitalization→Agency→Connection
- ② Agency→Connection→Digitalization
- ③ Agency→Digitalization→Connection
- ④ Digitalization→Connection→Agency

**08** 다음 중 데이터 사이언티스트에게 요구되는 소프트 스킬(Soft Skill)에 해당하는 것을 모두 고른 것은?

- ㉠ 빅데이터에 대한 이론적 지식
- ㉡ 설득력 있는 전달
- ㉢ 분석 기술에 대한 숙련
- ㉣ 통찰력 있는 분석
- ㉤ 다분야 간 협력

- ① ㉡, ㉣
- ② ㉡, ㉣, ㉤
- ③ ㉠, ㉢, ㉤
- ④ ㉠, ㉡, ㉤

**09** 다음 중 DIKW 피라미드의 계층 구조의 요소가 아닌 것은?

- ① 기술
- ② 정보
- ③ 지혜
- ④ 데이터

**10** 데이터 사이언티스트가 갖춰야 할 인문학적 사고의 특성과 역할 중 6가지 핵심 질문에 해당하지 않는 것은?

- ① 경고
- ② 리포팅
- ③ 모델링
- ④ 결론 도출

## 2과목 데이터분석 기획(10문제)

**11** 다음 중 분석 거버넌스 체계의 구성요소로 옳지 않은 것은?

- ① 시스템(System)
- ② 조직(Organization)
- ③ 데이터(Data)
- ④ 통합(Integration)

**12** 분석수준 진단 결과 중 준비도는 높지만, 분석 업무 및 기법들이 부족한 기업의 유형으로 옳은 것은?

- ① 정착형
- ② 준비형
- ③ 확산형
- ④ 도입형

**13** 다음 중 비즈니스 모델기반 문제탐색 영역으로 옳지 않은 것은?

- ① 업무 영역
- ② 제품 영역
- ③ 경쟁자 영역
- ④ 규제와 감사 영역

**14** 다음 설명에 해당하는 분석 프로젝트 영역별 주요 관리 항목으로 옳은 것은?

프로젝트 목적성에 맞는 외부 소싱을 적절하게 운영할 수 있으며, 특히 PoC 형태의 프로젝트는 인프라뿐만 아니라 클라우드 등의 다양한 방안을 검토할 필요성이 있다.

- ① 조달(Procurement)
- ② 원가(Cost)
- ③ 통합(Integration)
- ④ 의사소통(Communication)

**15** 분석 기획 유형 중 주제에 따른 분류에 대한 설명으로 옳지 않은 것은?

- ① 최적화(Optimization) - 해결해야 할 문제가 무엇인지를 알고, 이미 분석의 방법도 인지하는 경우
- ② 솔루션(Solution) - 분석의 대상이 무엇인지 알지만, 분석의 방법을 모르는 경우 사용
- ③ 통찰(Insight) - 분석의 대상 자체를 새롭게 도출함
- ④ 발견(Discovery) - 분석의 대상과 방법을 모르는 경우 사용

**16** 다음 중 분석과제 발굴 방법 중 상향식 접근 방식에 대한 설명으로 옳지 않은 것은?

- ① 주어진 데이터를 기반으로 문제를 재정의하고 해결 방안을 탐색하는 방식이다.
- ② 해결 방안 탐색(Solution Search) 단계에서는 정의된 데이터 분석 문제를 해결하기 위해 분석 기법 및 역량에 따라 다양한 방안으로 탐색한다.
- ③ 문제 해결에 필요한 모든 데이터가 존재하지 않을 경우, 이에 대한 수집·대체 방안을 수립해야 한다. 대체 불가능한 데이터가 존재하는지 사전에 확인하여 프로젝트가 중단되는 위험을 방지할 수 있다.
- ④ 필요 데이터 존재 여부 자체가 불확실한 경우에도 사용할 수 있다.

**17** 다음 중 분석 기획 단계에서 프로젝트 위험 계획 수립 시 대응 방안으로 옳지 않은 것은?

- ① 발생 가능한 모든 위험(Risk)을 발굴하여 사전에 대응 방안을 수립함으로써 프로젝트 진행의 완전성을 높인다.
- ② 예상되는 위험에 대한 대응은 회피(Avoid), 전이(Transfer), 완화(Mitigate), 수용(Accept)으로 구분한다.
- ③ 모델의 생명주기를 설정하고 주기적으로 평가하여 유지·보수하거나 재구축 방안을 마련한다.
- ④ 예상되는 위험의 관리 계획서를 작성한다.

**18** 다음 중 분석조직에 대한 설명으로 옳지 않은 것은?

- ① 기능 구조는 전사적 핵심 분석이 어려우며 과거에 국한된 분석을 수행한다는 한계가 있다.
- ② 집중 구조는 전략적 중요도에 따라 분석 조직이 우선순위를 정해서 진행할 수 있다.
- ③ 분산 구조는 전사 차원의 우선순위를 수행한다.
- ④ 집중 구조는 업무 과다와 이원화 가능성이 존재할 수 있으므로 각 부서의 분석 업무와 역할 분담이 명확해야 한다.

**19** 다음 중 분석 로드맵 설정 단계인 ‘데이터 분석 체계 도입→데이터 분석 유효성 검증→데이터 분석 확산 및 고도화’에서 유효성 검증을 위한 추진과제로 알맞은 것은?

- ① 분석 알고리즘 및 아키텍처 설계
- ② 분석기회 발굴 및 분석과제 정의
- ③ 업무 프로세스 내재화를 위한 변화 관리
- ④ 빅데이터 분석-활용 시스템 구축

**20** KDD 분석방법론의 프로세스 중 다음에서 서술하는 단계로 옳은 것은?

• 데이터베이스 또는 원시 데이터에서 선택 혹은 추가 생성  
• 데이터 마이닝에 필요한 목표 데이터(Target Data) 구성

- ① 데이터 전처리(Preprocessing)
- ② 데이터 변환(Transformation)
- ③ 데이터 세트 선택(Selection)
- ④ 데이터 마이닝(Data Mining)

### 3과목 데이터분석(30문제)

21 다음 내용에서 설명하는 척도로 옳은 것은?

- 측정 대상이 갖고 있는 속성의 양을 측정하는 척도이다.
- 절대적 원점이 없다.

- ① 순서 척도
- ② 명목척도
- ③ 비율척도
- ④ 구간척도

22 다음 중 산포와 위치의 측도에 대한 설명으로 옳지 않은 것은?

- ① 분산은 평균으로부터 얼마나 떨어져 있는지를 나타내는 값이다.
- ② 첨도는 데이터 분포의 뾰족한 정도를 설명하는 통계량이다.
- ③ 왜도가 1보다 작거나 1보다 클 경우 데이터는 상당히 치우쳐 있다.
- ④ 사분위수는 데이터 세트 내에서 특정값의 위치를 나타낸다.

23 다음은 sleep 데이터를 활용한 summary 함수 결과이다. 이에 대한 해석으로 옳지 않은 것은?

```
> summary(sleepstudy)
```

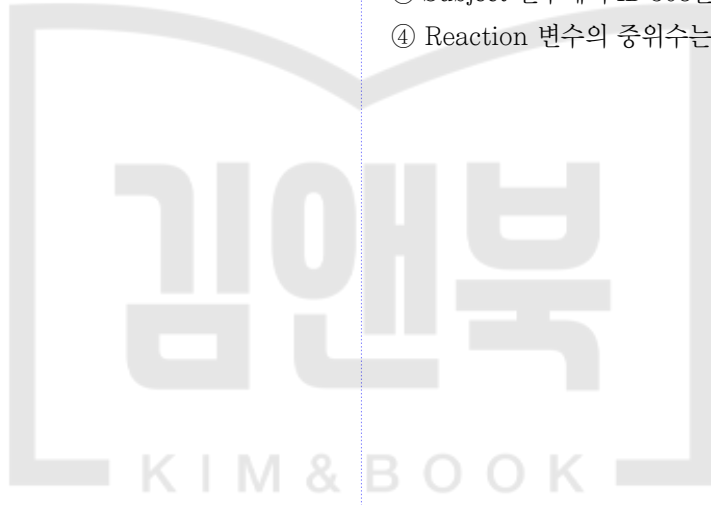
Reaction		Days		Subject	
Min.	:194.3	Min.	:0.0	308	:10
1st Qu.	:255.4	1st Qu.	:2.0	309	:10
Median	:288.7	Median	:4.5	310	:10
Mean	:298.5	Mean	:4.5	330	:10
3rd Qu.	:336.8	3rd Qu.	:7.0	331	:10
Max.	:466.4	Max.	:9.0	332	:10

(Other):120

- ① Reaction 변수의 평균은 298.5이다.
- ② Days 변수의 최대값은 10.0이다.
- ③ Subject 변수에서 ID 308은 10번 관측되었다.
- ④ Reaction 변수의 중위수는 288.7이다.

24 다음 중 Cook's Distance에 대한 설명으로 옳지 않은 것은?

- ① 회귀모형이 데이터를 얼마나 잘 설명하는지 나타낸다.
- ② 하나의 관측치가 (회귀)모형에 미치는 전반적인 영향을 나타내는 척도이다.
- ③ 1 이상일 경우 매우 큰 영향을 주는 관측값으로 간주한다.
- ④ x축은 관측값을 순서대로 나열, y축은 해당 지점의 쿡의 거리를 나타낸다.



**25** 다음 중 계층적 군집 분석에 대한 설명으로 옳지 않은 것은?

- ① 개별 관측치 간의 거리를 계산해서 가장 가까운 관측치부터 결합하면서 계층적 트리 구조를 형성하고 이를 통해 군집화를 수행하는 방법이다.
- ② 최장 연결법은 같은 군집에 속하는 관측치는 알려진 최대 거리보다 짧으며 군집들의 내부 응집성에 중점을 둔 방법이다.
- ③ 평균 연결법은 모든 항목에 대한 거리 평균을 구하면서 군집화를 하기 때문에 계산량이 불필요하게 많아질 수 있다.
- ④ 와드 연결법은 생성된 군집과 기존의 데이터들의 거리를 군집 내 분산이 최소가 되는 데이터로 계산하는 방법이다.

**26** 다음 중 연관 분석에 대한 설명으로 옳지 않은 것은?

- ① 데이터 내부에 존재하는 항목 간의 유용한 패턴을 찾아내는 분석 기법으로, 연관규칙의 모든 규칙이 유용한 것은 아니다.
- ② FP-Growth는 FP Tree구조를 이용하여 Apriori 알고리즘보다 훨씬 빠른 속도로 계산되고, 대용량 데이터 세트에서 메모리가 효율적이고 설계가 쉽다.
- ③ 거래 내용에 대한 데이터를 변환 없이 그 자체로 이용할 수 있는 간단한 자료구조를 갖는 분석 방법이다.
- ④ 상대적으로 거래량이 적은 품목은 당연히 포함된 거래 수가 적을 것이므로 규칙 발견 시 제외되기 쉽다.

**27** 다음 거래 데이터에서 연관규칙으로 ‘빵과 우유’ 간의 지지도, 신뢰도가 순서대로 나열된 것은?

- A: 책, 빵
- B: 책, 우유, 와인
- C: 책, 빵, 우유, 과자
- D: 책, 과자, 우유
- E: 책, 우유, 빵, 주스

- ①  $\frac{2}{5}, \frac{2}{3}$                       ②  $\frac{3}{5}, \frac{2}{3}$
- ③  $\frac{3}{5}, \frac{3}{5}$                       ④  $\frac{2}{5}, \frac{3}{5}$

**28** 의사결정나무 중 목표 변수가 연속형인 회귀나무에서 분류 기준값으로 적절한 것은?

- ① 카이제곱 통계량의 p값
- ② 엔트로피 지수
- ③ 지니 지수
- ④ F-통계량

**29** 다음 중 앙상블 모형의 배깅(Bagging)에 대한 설명으로 옳지 않은 것은?

- ① 다수의 부트스트랩 자료를 생성하고, 각 자료를 모델링한 후 결합해 최종 예측모형을 만든다.
- ② 높은 바이어스(Bias)나 과소 적합(Under-Fitting) 문제, 높은 분산(Variance)으로 인한 과대 적합(Over-fitting) 문제를 해결하는 데 도움을 준다.
- ③ 계산이 간단하다는 장점이 있다.
- ④ 여러 가지 모형들의 예측 및 분류 결과를 종합하여 최종적인 의사결정에 활용한다.

**30** 다음 중 단계적 변수 선택 방법에 대한 설명으로 옳지 않은 것은?

- ① 전진 선택법은 설명계수를 기준으로 가장 설명을 잘하는 변수부터 추가하고 그렇지 않은 경우 추가를 중단한다.
- ② 단계적 선택법은 전진 선택법으로 중요도가 높은 변수를 추가하면서 중요도가 작은 변수를 후진 제거법으로 제거하는 혼합방식이다.
- ③ 후진 제거법은 모든 독립변수가 추가된 전체 모형에서 설명력이 작은 변수부터 제거하는 방법이다.
- ④ 후진 제거법은 더 이상 유의하지 않은 변수가 없을 때까지 설명변수를 제거하고, 이때의 모형을 선택한다.

**31** 다음 회귀분석 결과에 대한 설명으로 옳지 않은 것은?

```

> summary(model)

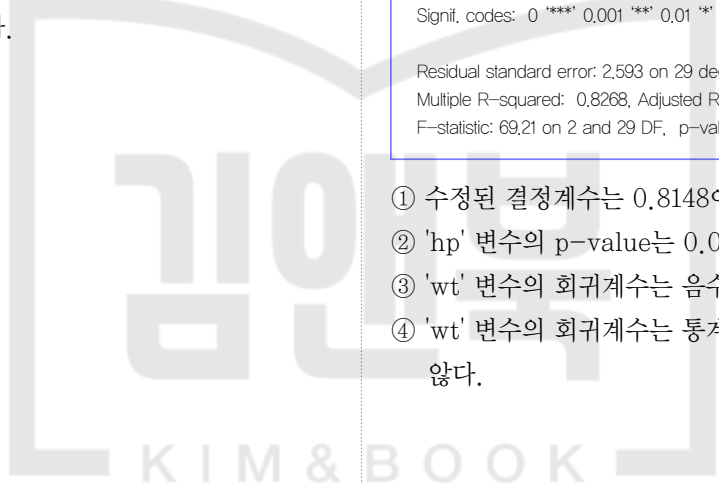
Call:
lm(formula = mpg ~ wt + hp, data = mtcars)

Residuals:
    Min       1Q   Median       3Q      Max
-3.941  -1.600   -0.182    1.050    5.854

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.22727    1.59879   23.285 < 2e-16 ***
wt           -3.87783    0.63273   -6.129 1.12e-06 ***
hp           -0.03177    0.00903   -3.519 0.00145 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.593 on 29 degrees of freedom
Multiple R-squared:  0.8268, Adjusted R-squared:  0.8148
F-statistic: 69.21 on 2 and 29 DF, p-value: 9.109e-12
    
```

- ① 수정된 결정계수는 0,8148이다.
- ② 'hp' 변수의 p-value는 0,01보다 작다.
- ③ 'wt' 변수의 회귀계수는 음수이다.
- ④ 'wt' 변수의 회귀계수는 통계적으로 유의하지 않다.





**32** 다음 중 다중 공선성에 대한 설명으로 옳지 않은 것은?

- ① 상태지수를 활용하여 1 이상이면 문제, 5보다 크면 심각한 문제가 있는 것으로 본다.
- ② 분산팽창요인(VIF)을 활용하여 해당 값이 1에 가까울수록 다중 공선성이 낮으며, 4보다 크면 다중 공선성이 존재하고, 10보다 크면 심각한 문제가 있는 것으로 본다.
- ③ 다중 선형회귀 분석에서 사용된 모형의 일부 독립변수가 다른 독립변수와 상관관계가 강하게 나타나 정확한 추정이 어려워지는 문제가 있다.
- ④ 하나의 독립변수가 다른 독립변수로 잘 예측될 수 있는 상황을 의미한다.

**33** 다음 중 선형회귀 분석의 기본가정으로 옳지 않은 것은?

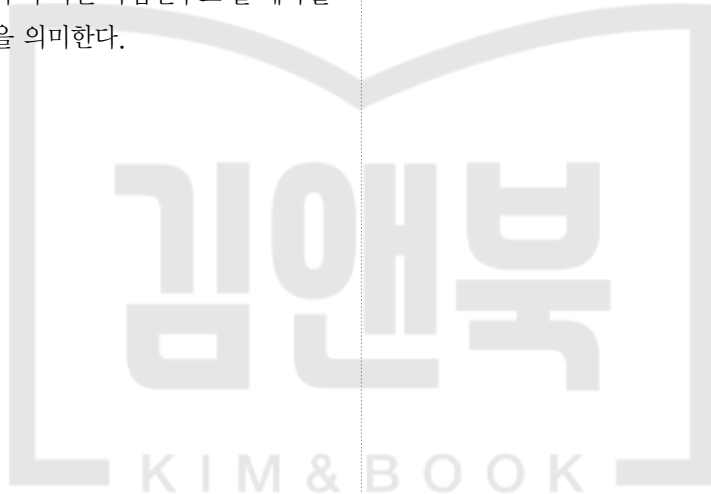
- ① 선형성
- ② 비상관성
- ③ 정상성
- ④ 연관성

**34** 다음 중 시계열 분석에 대한 설명으로 옳지 않은 것은?

- ① 자기회귀 누적 이동평균 모형은 비정상 시계열 모형이기 때문에 차분이나 변환을 통해 정상화 차수를 설명한다.
- ② 시계열 분석에서 시점에 관계 없이 시계열의 특성이 일정한 것을 정상성이라고 한다.
- ③ 시간의 흐름에 따라서 관측된 데이터를 분석하여 미래를 예측하기 위한 분석 기법이다.
- ④ 자기회귀 모형(AR 모형: Auto Regressive Model)은 고전적인 시계열 분해기법으로 추세-주기를 측정하기 위해 사용한다.

**35** 다음 중 신경망 모형에 대한 설명으로 옳은 것은?

- ① 항상 선형 관계만을 모델링할 수 있다.
- ② 활성화 함수는 비선형 함수를 사용할 수 있다.
- ③ 은닉층의 수와 뉴런 수를 늘리면 항상 신경망의 성능이 향상된다.
- ④ 학습률(Learning Rate)은 학습 과정에서 변경할 수 없다.



**36** 다음 보기에서 설명하는 군집분석 모델의 평가지표로 옳은 것은?

- 클러스터 내 최대 거리에 대한 클러스터 간의 최소 거리의 비율이다.
- 군집 간 거리는 멀수록, 군집 내 분산은 작을수록 좋은 군집화이며, 값이 클수록 좋다.

- ① 실루엣 계수
- ② 중심 연결법
- ③ Elbow Method
- ④ Dunn Index

**37** 다음 중 상관계수에 대한 설명으로 옳지 않은 것은?

- ① 피어슨 상관계수는 +1과 -1 사이의 값을 가지며, +1은 완벽한 양의 선형 상관관계, 0은 선형 상관관계 없음, -1은 완벽한 음의 선형 상관관계를 의미한다.
- ② 두 변수 사이에 연관성을 수치적으로 객관화하여 두 변수 사이의 방향성과 강도를 표현한 것이다.
- ③ 스피어만 상관분석은 비율척도인 두 변수의 순위 사이의 통계적 의존성을 측정하는 비모수적 척도이다.
- ④ 스피어만 상관분석은 두 변수 간의 연관관계 여부를 알려주며, 자료의 이상값이 있거나 표본 크기가 작을 때 유용하다.

**38** 다음 중 범주형 종속변수에는 지니 지수를, 연속형 종속변수에는 분산감소량을 사용하는 의사결정나무 알고리즘은?

- ① C4.5와 C5.0
- ② CART
- ③ AID
- ④ CHAID

**39** 다음 중 통계적 가설검정에 대한 설명으로 옳지 않은 것은?

- ① p-값이 작을수록 귀무가설을 지지하는 것으로 해석할 수 있다.
- ② 가설검정이란 모집단에 대해 어떤 가설을 설정하고 표본 관찰을 통해 그 가설의 채택 여부를 결정하는 분석 방법이다.
- ③ 유의수준은 귀무가설의 기각 여부를 결정하는데 사용하는 기준이 되는 확률이다.
- ④ 기각역은 귀무가설이 기각되고 대립가설이 채택되는 검정통계량의 영역을 의미한다.

**40** 다음 중 모수검정과 비모수검정에 대한 설명으로 옳지 않은 것은?

- ① 모수검정은 보통 데이터의 순위나 순서를 사용하여 검정한다.
- ② 데이터가 가정된 분포를 따르지 않을 때 비모수검정이 더 적합할 수 있다.
- ③ 모수검정은 데이터가 어떤 분포를 따르지 않아도 사용할 수 있다.
- ④ 만-휘트니(Mann-Whitney) U 검정은 비모수검정의 예이다.

**41** 다음 중 결측값에 대한 설명으로 옳지 않은 것은?

- ① 결측값은 입력이 빠진 값으로 R에서는 NA, NAN(0으로 나누는 것이 불가능한 값) 등으로 표현된다.
- ② 완전 무작위 결측이란 어떤 변수에서 발생한 결측값이 다른 변수들과 아무런 상관이 없는 경우이다.
- ③ 무작위 결측이란 어떤 변수의 누락 데이터가 특정 변수와 관련되어 일어나지만, 그 변수의 결과는 관계가 없는 경우이다.
- ④ 처리 절차는 '부호화-대체-식별' 순으로 진행된다.

**42** 다음 중 결측값 처리에 활용되는 완전 분석법에 대한 설명으로 옳지 않은 것은?

- ① 누락된 데이터가 발생한 데이터를 제외하고 분석을 수행한다.
- ② 부분적으로 무시되는 자료로 인해 통계적 추론의 타당성 문제가 발생한다.
- ③ 분석이 간단해진다는 장점이 있다.
- ④ 통계량의 표준오차가 과소 추정될 수 있다.

**43** 다음 trees 데이터 세트에는 31개 나무의 둘레(Girth)와 높이(Height) 정보가 포함되어 있다. 이 두 변수 간의 관계를 알아보기 위해 단순 선형 회귀 분석을 수행한 결과로 옳지 않은 것은?

```
> summary(model)

Call:
lm(formula = Height ~ Girth, data = trees)

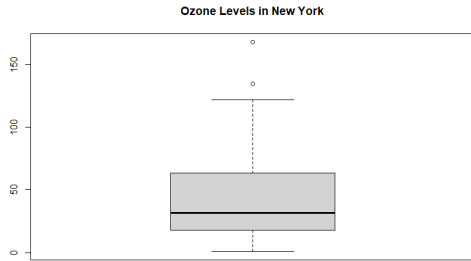
Residuals:
    Min       1Q   Median       3Q      Max
-12.5816  -2.7686   0.3163   2.4728   9.9456

Coefficients:
            Estimate Std. Error value Pr(>|t|)
(Intercept)  62.0313    4.3833   14.152 1.49e-14 ***
          Girth   1.0544    0.3222    3.272 0.00276 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.538 on 29 degrees of freedom
Multiple R-squared:  0.2697, Adjusted R-squared:  0.2445
F-statistic: 10.71 on 1 and 29 DF, p-value: 0.002758
```

- ① Girth가 Height의 전체 변동성의 약 26.97%를 설명한다.
- ② Girth의 표준오차는 0.3222이다.
- ③ Girth의 p-value는 0.00276로 유의수준 0.05에서 유의미하다.
- ④ 모델의 자유도는 30이다.

**44** 다음은 airquality 데이터 세트에서 Ozone 변수에 대한 상자 그림이다. 이에 대한 설명으로 옳지 않은 것은?



- ① 최솟값, 제1사분위수, 중앙값, 제3사분위수, 최댓값을 나타낸다.
- ② 150 이상이면 이상값으로 볼 수 있다.
- ③ 이상치가 없다면 수염이 없을 수 있다.
- ④ 데이터의 대부분이 1사분위수(Q1)인 약 18부터 3사분위수(Q3)인 약 63 사이에 분포되어 있다.

**45** 다음 중 의사결정나무의 분리기준에 대한 설명으로 옳지 않은 것은?

- ① 지니 지수는 값이 낮을수록 이질적(Diversity)이며, 순수도(Purity)가 낮다고 볼 수 있다.
- ② 엔트로피 지수는 값이 클수록 순수도(Purity)가 낮다.
- ③ 카이제곱 통계량의 불순도 계산 공식은 
$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$
이다.
- ④ 분산의 감소량은 회귀나무에서 사용되며 예측 오차를 최소화하는 기준으로 분산의 감소량을 최대화하는 기준의 최적 분리를 진행한다.

**46** 다음 중 의사결정나무의 분석 과정에 대한 설명으로 옳지 않은 것은?

- ① 의사결정나무 형성 - 분리 규칙을 찾아 나무를 성장시키는 과정으로 적절한 정지 규칙을 만족하면 중단한다.
- ② 정지규칙(Stopping Rule) - 더 이상 분리가 일어나지 않고 현재의 마디가 끝마디가 되도록 하는 규칙으로, 끝마디의 자료 개수는 분석이 끝난 후 지정한다.
- ③ 가지치기 - 분류 오류를 크게 할 위험이 높거나 부적절한 추론 규칙을 갖고 있는 가지를 제거한다.
- ④ 해석 및 예측 - 이익도표, 위험도표를 이용하여 모형을 평가한다.

**47** 다음 중 다차원 척도법에 대한 설명으로 옳은 것은?

- ① 주어진 데이터의 원래 공간상에서의 위치를 보존한다.
- ② 데이터의 차원을 줄이는 방법이다.
- ③ 다차원 척도법은 데이터 간의 관계를 거리 행렬로 나타내거나 유사도 행렬로도 나타낼 수 있다.
- ④ 주로 범주형 데이터에 적용된다.

**48** 다음 중 혼동행렬을 통한 분류모형의 평가 지표 중 하나인 특이도 계산식으로 적절한 것은?

- ①  $\frac{TP}{TP+FN}$
- ②  $\frac{TP+TN}{TP+TN+FN+FP}$
- ③  $\frac{TN}{TN+FP}$
- ④  $\frac{FP}{TN+FP}$

**49** 다음 중 로지스틱 회귀 분석에 대한 설명으로 옳지 않은 것은?

- ① 오즈에 로그를 씌우는 변환인 로짓변환은 범위가  $[-1,1]$ 일 때 출력값의 범위를  $(-\infty, \infty)$ 로 조정한다.
- ② 종속변수가 범주형인 경우 적용되는 회귀 분석 모형이다.
- ③ 새로운 설명변수의 값이 주어질 때 반응변수의 각 범주에 속할 확률이 얼마인지 추정하여 추정 확률을 기준치에 따라 분류하는 목적으로 사용된다.
- ④ 오즈비는 각 모수에 대해 비선형식이며 승산으로 로짓변환을 통해 선형 함수로 치환이 가능하다.

**50** 다음 중 시계열분석의 분석절차로 옳은 것은?

- ① 잔차에 대한 모델 적합→추세와 계절성 제거→시간 그래프 작성→잔차 예측→미래 예측
- ② 시간 그래프 작성→추세와 계절성 제거→잔차 예측→잔차에 대한 모델 적합→미래 예측
- ③ 추세와 계절성 제거→시간 그래프 작성→잔차 예측→잔차에 대한 모델 적합→미래 예측
- ④ 잔차에 대한 모델 적합→시간 그래프 작성→추세와 계절성 제거→잔차 예측→미래 예측

**정답**

1	2	3	4	5	6	7	8	9	10
③	②	①	②	①	①	④	②	①	④
11	12	13	14	15	16	17	18	19	20
④	④	③	①	③	②	③	④	①	③
21	22	23	24	25	26	27	28	29	30
④	④	②	①	④	②	①	④	③	①
31	32	33	34	35	36	37	38	39	40
④	①	④	④	②	④	③	②	①	③
41	42	43	44	45	46	47	48	49	50
④	④	④	③	①	②	③	③	①	②