

1. 다음 중 연속형 자료에 해당되는 것은?
- ① 흡연 여부
 - ② 섭씨 온도
 - ③ 혈액형
 - ④ 성별
2. 다음 중 산포도(흩어진 정도)의 측도로 사용할 수 없는 것은?
- ① 사분위수 범위
 - ② 변이계수(변동계수)
 - ③ 표준편차
 - ④ 중앙값
3. 다음 중 범주형 자료의 요약을 위한 방법으로 적절하지 않은 것은?
- ① 상자그림(box plot)
 - ② 분할표(contingency table)
 - ③ 파이차트(pie chart)
 - ④ 막대그래프(bar plot)
4. 숫자 아닌 자료들을 분석할 때 알 수 있는 것은?
- ① 자료값들의 평균(mean)
 - ② 자료값들의 중앙값(median)
 - ③ 자료값들의 최빈값(mode)
 - ④ 자료값들의 분산(variance)
5. 표본공간 S 에서 정의된 두 사건 A 와 B 에 대하여 옳지 않은 것은?
- ① $P(S) = 1$
 - ② $A \subset B$ 이면 $P(A) \leq P(B)$
 - ③ $P(A^c) = 1 - P(A)$, 단 A^c 는 A 의 여사건
 - ④ $A \cap B = \phi$ 이면 A 와 B 는 서로 독립이다.
6. 두 사건 A 와 B 가 서로 독립일 때 성립하지 않는 것은?
- ① $P(A \cap B) = P(A)P(B)$
 - ② $P(A|B) = P(A)$
 - ③ $P(B|A) = P(B)$
 - ④ $P(A \cup B) = P(A) + P(B)$

7. 두 사건 A 와 B 가 서로 독립이고, $P(A) = 0.3$, $P(B) = 0.2$ 이다. $P(A \cup B)$ 는 얼마인가?
- ① 0.41
 - ② 0.42
 - ③ 0.43
 - ④ 0.44
8. 확률변수 X 에 대해 $E(X)$ 는 기댓값(expectation), $Var(X)$ 는 분산(variance), $SD(X)$ 는 표준편차(standard deviation)을 나타낸다. 다음 중 맞지 않는 것은? (단, a 와 b 는 실수이다.)
- ① $E(aX+b) = aE(X) + b$
 - ② $Var(aX+b) = a^2 Var(X)$
 - ③ $SD(aX+b) = aSD(X)$
 - ④ $E(X^2) \geq \{E(X)\}^2$
9. 다음은 대학생 2,000명의 음주와 흡연 습관을 조사한 결과표이다. 흡연자들 중 음주를 하는 학생들의 비율은?
- | | 음주 | 비음주 | 합계 |
|------|-----|-------|-------|
| 흡연 | 300 | 200 | 500 |
| 비흡연자 | 600 | 900 | 1,500 |
| 합계 | 900 | 1,100 | 2,000 |
- ① 0.40
 - ② 0.45
 - ③ 0.60
 - ④ 0.75
10. 다음은 1개의 동전을 연속하여 3회 던져서 앞면의 수의 횟수 X 의 확률분포함수를 나타낸 표이다. 빈칸 () 에 들어갈 값은?
- | X | 0 | 1 | 2 | 3 |
|--------|---------------|---------------|---------------|-----|
| $P(x)$ | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | () |
- ① $\frac{1}{8}$
 - ② $\frac{2}{8}$
 - ③ $\frac{3}{8}$
 - ④ $\frac{4}{8}$

11. X 와 Y 의 결합확률질량함수가 다음 표와 같을 때, X 의 기대값 $E(X)$ 의 값은?

	$X=-1$	$X=0$	$X=1$
$Y=-1$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$Y=0$	$\frac{1}{8}$	0	$\frac{1}{8}$
$Y=1$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

- ① 0
- ② $\frac{1}{8}$
- ③ $\frac{2}{8}$
- ④ $\frac{3}{8}$

12. 확률변수 X 와 Y 의 결합확률질량함수가 아래와 같다. 틀린 것은?

	$X=-1$	$X=0$	$X=1$
$Y=0$	0	$\frac{1}{3}$	0
$Y=1$	$\frac{1}{3}$	0	$\frac{1}{3}$

- ① $E(X) = 0$
- ② $E(XY) = 0$
- ③ $Cov(X, Y) = 0$
- ④ X 와 Y 가 서로 독립이다.

13. 확률변수 X 와 Y 의 상관계수가 0에 가까울 때의 정확한 해석은?

- ① X 와 Y 가 아무 관계가 없다.
- ② X 와 Y 사이에 곡선 관계가 존재한다.
- ③ X 와 Y 사이에 직선 관계가 약하다.
- ④ X 와 Y 사이에 곡선 관계가 약하다.

14. 확률변수 X 와 Y 의 분산을 각각 $Var(X)$, $Var(Y)$ 라 하고 X 와 Y 의 공분산을 $Cov(X, Y)$ 라 할 때, 다음 설명 중 옳지 않은 것은?

- ① X 와 Y 가 독립이면 $Cov(X, Y)$ 는 항상 0이다.
- ② $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$
- ③ $Cov(X, Y)$ 는 음수가 될 수도 있다.
- ④ $Var(X - Y)$ 는 음수가 될 수도 있다.

15. 확률변수 X, Y 와 상수 a, b, c, d 에 대하여 상관계수 $Corr(aX + b, cY + d)$ 와 같은 값은? (단, a 와 c 는 양수이다.)

- ① $Corr(X, Y)$
- ② $acCorr(X, Y)$
- ③ $acCorr(X, Y) + b + d$
- ④ $acCorr(X, Y) + bd$

16. 확률변수 X 와 Y 의 분산이 각각 $Var(X) = 3$, $Var(Y) = 2$ 이고 공분산이 $Cov(X, Y) = -\sqrt{3}$ 이다. $Var(2X - \sqrt{3}Y)$ 는 얼마인가?

- ① 10
- ② 20
- ③ 30
- ④ 40

17. A 산업단지의 공장의 크기가 다음과 같은 특성을 가지고 있다. 가장 작은 공장의 크기는 15에이커이고, 전체의 10%는 25에이커보다 작고, 전체의 25%는 40에이커보다 작으며 중앙값은 55에이커, 75%는 100에이커보다 작고, 90%는 125에이커보다 작다. 가장 큰 공장은 200에이커이다. 사분위수 범위(Interquartile Range, IQR)를 구하시오.

- ① 185
- ② 100
- ③ 60
- ④ 55

18. A 지역 휘발유의 리터 당 가격이 평균 1,800원이고 표준편차가 50원이다. A 지역 어떤 주유소에 리터 당 휘발유 가격이 1,900원일 때, 표준화 값인 z -값를 구하시오.

- ① 1.5
- ② 2
- ③ 2.5
- ④ 3

19. 창고에 100개의 제품이 있는데, 그 중 20개가 불량품이고, 80개가 양품이다. 정확한 불량률이 0.2라는 것을 모르기에, 불량률을 추정하기 위해 표본을 복원추출로 10개 뽑는다. 뽑힌 10개의 표본 중 불량품의 개수를 확률변수 X 라 하면 X 의 분산은?

- ① 1.4
- ② 1.6
- ③ 1.8
- ④ 2.0

20. A 은행에서는 오전 10시부터 11시 사이에 평균 60명의 손님이 방문한다고 한다. 10분간 도착한 손님의 수를 확률변수 X 라고 할 때, X 의 분산은?

- ① 1
- ② 6
- ③ 10
- ④ 60

21. 확률변수 X 의 평균이 2, 분산이 9이다. 이때 $2X$ 의 평균과 분산은?

- ① 평균: 2, 분산 9
- ② 평균: 4, 분산 18
- ③ 평균: 4, 분산 36
- ④ 평균: 8, 분산 36

22. 평균이 10, 분산이 25인 모집단으로부터 확률표본 100개를 추출하였을 때, 표본평균 $\bar{X}$ 의 평균과 분산은?

- ① 평균: 10, 분산 25
- ② 평균: 10, 분산 0.25
- ③ 평균: 0.1, 분산 25
- ④ 평균: 0.1, 분산 0.25

23. 확률변수 Z 가 평균이 0이고 표준편차가 1인 표준정규 분포를 따른다고 한다. 이때 $P(Z > 0)$ 은?

- ① 0
- ② 0.5
- ③ 0.9
- ④ 1.0

24. 확률변수 X 가 아래의 확률밀도함수를 갖는다. X 의 중앙값은?

$$f(x) = \begin{cases} -\frac{3}{4}(x+2)x, & (-2 < x < 0) \\ 0 & (\text{그 외}) \end{cases}$$

- ① -2
- ② -1
- ③ 1
- ④ 2

25. 성공확률이 p , 시행횟수가 n 인 이항분포 $B(n, p)$ 를 따르는 확률변수 X 에 대한 설명으로 옳지 않은 것은?

- ① 확률변수 $X_1, \dots, X_n$ 이 서로 독립이고 동일한 성공확률 p 를 가지는 베르누이 확률변수일 때 $X = \sum_{i=1}^n X_i$ 이다.
- ② X 의 기댓값은 np , 분산은 $np(1-p)$ 이다.
- ③ X 가 가질수 있는 값은 1부터 n 까지의 자연수이다.
- ④ 확률변수 Y 가 이항분포 $B(m, p)$ 를 따르고 X 와 Y 가 독립일 때, $X+Y$ 는 이항분포 $B(n+m, p)$ 를 따른다(m 은 자연수).

26. 분포에 대한 아래의 설명 중 항상 옳은 것은?

- ① 표준정규분포는 t 분포에 비해 꼬리가 두껍다.
- ② 두 확률변수 Z_1, Z_2 가 표준정규분포를 따를 때, $Z_1^2 + Z_2^2$ 은 자유도가 2인 카이제곱 분포를 따른다.
- ③ t 분포를 따르는 확률변수를 제곱한 확률변수는 F 분포를 따른다.
- ④ t 분포는 모분산에 대한 추론 시 사용하는 분포이다.

27. 어느 회사에서 하나의 새로운 상품을 개발하였다. 기존 고객 명단에서 임의로 100명을 추출하여 신상품 구입 예약을 받은 결과 그 중 50명이 구입을 희망하였다. 이 표본자료에 기반하여 신상품 구입을 원하는 고객의 비율 p 를 추정하고자 한다. 이 추정량의 표준 오차의 추정치는?

- ① 0.01
- ② 0.05
- ③ 0.1
- ④ 0.2

28. 정규모집단의 평균 μ 의 구간추정에 대한 설명 중 옳은 것은?

- ① 같은 신뢰수준을 가지는 조건 하에서 신뢰구간의 길이가 길수록 더 좋은 구간추정량이다.
- ② 대한민국 성인 평균 키 μ 에 대한 95% 구간추정치가 (157, 172)이면 $P(157 < \mu < 172) = 0.95$ 이다.
- ③ 모든 조건이 동일할 때, 표본의 수가 커지면 신뢰구간의 길이가 짧아진다.
- ④ 모든 조건이 동일할 때, 신뢰수준(confidence level)이 95%에서 99%로 커지면 신뢰구간의 길이가 짧아진다.

29. 다음은 어떤 통계적 방법론에 대한 설명인가?

설명변수 X 가 주어졌을 때, 반응변수 Y 의 조건부 기댓값 $E(Y|X=x)$ 를 추론하고자 하는 통계적 방법론

- ① 분산분석
- ② 회귀분석
- ③ 군집분석
- ④ 표본조사 방법론

30. 분산분석(Analysis of Variance, ANOVA)에 대한 설명 중 옳지 않은 것은?

- ① 여러 집단의 분산을 비교하기 위한 분석 방법이다.
- ② 자료의 총변동을 분해하여 집단 간 차이가 있는지 파악한다.
- ③ 검정통계량의 분포는 귀무가설 하에서 F 분포를 따른다.
- ④ 귀무가설이 기각되었을 시 사후검정으로 다중비교를 실시한다.

31. 통계적 가설검정에서 귀무가설이 사실임에도 불구하고 귀무가설을 기각할 오류를 범할 확률의 최대 허용 한계를 의미하는 통계 용어는 무엇인가?

- ① 검정력
- ② 제 I종 오류
- ③ 유의수준
- ④ 유의확률

32. 어느 한 선거구에서 지역에 따라 A, B, C 세 정당 후보자에 대한 지지도가 다른지 알아보기 위해 각 도시에서 100명씩 조사한 결과 다음의 자료를 얻었다. 지역에 따라 정당 후보자 지지도가 다른지 검정하고자 하는 카이제곱검정 통계량의 자유도는?

	A	B	C	계
지역 1	35	30	35	100
지역 2	50	26	24	100
지역 3	40	33	27	100
지역 4	30	50	20	100

- ① 6
- ② 9
- ③ 12
- ④ 20

33. 한 개의 주사위와 한 개의 동전을 동시에 던질 때, 주사위의 짝수의 눈과 동전의 앞면이 나올 확률은?

- ① $\frac{1}{8}$
- ② $\frac{1}{6}$
- ③ $\frac{1}{4}$
- ④ $\frac{1}{2}$

34. 남학생과 여학생에 따라 영어, 수학, 국어, 세 과목에 대한 선호도가 다른지 조사한 결과가 아래 표와 같다. “남학생과 여학생의 선호도가 같다”는 귀무가설 하에서 남학생 영어 선호도의 추정된 기대값은?

	영어	수학	국어	표본 크기
남학생	100	80	70	250
여학생	110	60	80	250
합계	210	140	150	500

- ① 102
- ② 103
- ③ 104
- ④ 105

35. 로지스틱 회귀분석에 대한 설명 중 옳지 않은 것은?

- ① 설명변수와 반응변수가 모두 범주형변수일 때에만 적용 가능한 분석 방법이다.
- ② 로짓(logit) 연결함수를 사용해 예측값을 0과 1 사이로 제한한다.
- ③ 오즈비(odds ratio)를 이용해 모형 해석이 가능하다.
- ④ 금융회사에서 고객의 신용카드 발급 여부를 판단할 때 사용할 수 있는 분석 방법이다.

36. 강우량 정도에 따른 대기오염 제거의 정도를 알아보기 위해 데이터를 이용하여 추정된 회귀직선 $\hat{y}=110-5x$ 를 얻었다. 강우량이 $x=10$ 일 때, 대기오염 제거 정도의 추정값은?

- ① 10
- ② 30
- ③ 40
- ④ 60

37. 회귀분석을 통해 다음과 같은 분산분석표를 얻었다.
오차항의 모분산 σ^2 의 추정치는?

요인	제곱합	자유도	평균제곱	F 값
회귀	200	1	200	4.0
잔차	300	6	50	
계	500	7		

- ① 50
- ② 200
- ③ 300
- ④ 500

38. 어떤 제품을 만들 때 처리액의 농도가 제품의 강도에 영향을 미치는지 조사하기 위해 네 가지 농도에서 각 농도 당 5회를 반복하여 총 20회의 실험을 랜덤하게 처리한 후 강도를 측정했다. 이 자료에 일원배치법을 적용하여 작성한 분산분석표가 아래와 같다. (ㄱ)의 F값은?

요인	제곱합	자유도	평균제곱	F 값	p-값
처리	1800	3		(ㄱ)	
잔차	1600	16			
합계					

- ① 6
- ② 7
- ③ 8
- ④ 9

39. 독립변수 x 와 종속변수 y 사이에 직선 관계를 가정하고 단순선형회귀 모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, 2, \dots, n$ (ϵ_i 들은 서로 독립이고, ϵ_i 들이 각각 정규분포 $N(0, \sigma^2)$ 을 따름)을 적용시켰다. β_0, β_1 의 최소제곱추정량을 각각 b_0, b_1 이라 하고, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$,

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $\hat{y}_i = b_0 + b_1 x_i$ 라 하자. 아래의 내용 중 틀린 것은?

- ① $b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$
- ② $b_0 = \bar{y} - b_1 \bar{x}$
- ③ $\frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 은 σ^2 의 불편(비편향)(unbiased) 추정량이다.
- ④ $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$

40. 평균이 μ , 분산이 σ^2 인 정규모집단으로부터 크기 n 인 확률표본 $X_1, \dots, X_n$ 을 추출하였을 때, 표본평균 $\bar{X}_n$, 표본분산 S_n^2 , 표본표준편차 S_n 이 다음과 같다.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \\ S_n = \sqrt{S_n^2}, S_n > 0.$$

다음 설명 중 옳지 않은 것은?

- ① S_n^2 은 σ^2 의 불편(비편향)(unbiased) 추정량이다.
- ② $\frac{(n-1)S_n^2}{\sigma^2}$ 은 자유도가 $n-1$ 인 카이제곱 분포를 따른다.
- ③ $\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$ 은 표준정규분포를 따른다.
- ④ $\bar{X}_n$ 은 μ 의 불편 추정량이다.