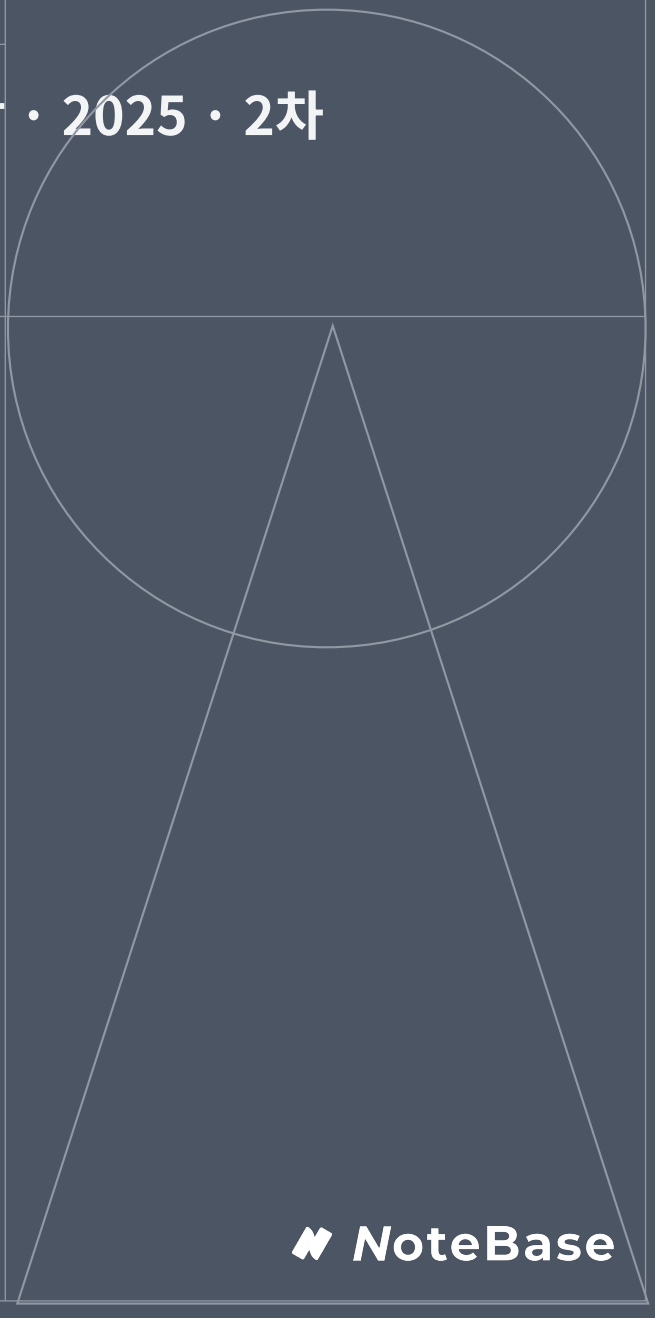
기출OX 해설집

공무원/국가직/7급 · 통계학 · 2025 · 2차



NoteBase

기출OX 해설집

공무원/국가직/7급 · 통계학 · 2025 · 2차

공무원/국가직/7급 · 통계학 · 2025 · 2차

문 1. 다음은 중학생 100명을 대상으로 하루 평균 게임 시간을 조사하여 작성한 도수분포표의 일부이다. 이에 대한 설명으로 옳은 것은?

하루 평균 게임 시간	도수	상대도수
1시간 미만	25	
1시간 이상 ~ 2시간 미만	35	
2시간 이상 ~ 3시간 미만	(가)	0.21
3시간 이상 ~ 4시간 미만	7	
4시간 이상 ~ 5시간 미만	7	
5시간 이상	(나)	
합계	100	1.00

① (가)의 값은 20이다.

② (나)의 값은 3이다.

③ 하루 평균 게임 시간이 2시간 미만인 학생들의 비율은 0.5보다 작다.

④ 하루 평균 1시간 이상 2시간 미만 게임을 하는 학생들의 비율이 가장 높다.

정답 4번

출제 포인트

이 문제는 도수와 상대도수의 관계를 이용해 빈칸 값을 구하고, 각 구간의 비율을 비교할 수 있는지를 묻는다.

NoteBase

자료 읽는 순서

전체 학생 수가 100명임을 먼저 확인한다.
상대도수 0.21은 해당 구간의 도수가 21명이라는 뜻으로 읽는다.
각 구간의 도수와 비율을 구한 뒤 선지의 수치와 크기 비교를 대조한다.

선지별 해설

1번: X. 2시간 이상 3시간 미만 구간의 상대도수가 0.21이므로 도수는 100×0.21=21이다. 선지의 '20'이 틀렸고, 올바르게는 '(가)의 값은 21이다.'이다.
2번: X. 전체 100명에서 이미 알려진 도수 25, 35, 21, 7, 7을 더하면 95이므로 남은 5시간 이상 구간의 도수는 5이다. 선지의 '3'이 틀렸고, 올바르게는 '(나)의 값은 5이다.'이다.
3번: X. 하루 평균 게임 시간이 2시간 미만인 학생은 1시간 미만 25명과 1시간 이상 2시간 미만 35명을 합한 60명이다. 비율은 0.60이므로 0.5보다 작지 않다. 선지의 '0.5보다 작다'가 틀렸고, 올바르게는 '0.5보다 크다.'이다.
4번: O. 각 구간의 비율은 차례로 0.25, 0.35, 0.21, 0.07, 0.07, 0.05이다. 이 중 1시간 이상 2시간 미만의 비율 0.35가 가장 크므로 옳다.

오답 포인트

상대도수 0.21을 그대로 21명으로 바꾸지 못하거나, 2시간 미만을 1시간 미만만으로 잘못 읽으면 ①과 ③에서 틀리기 쉽다. 도수와 비율, 구간 합산 범위를 함께 확인해야 한다.

다시 나오면 이렇게 판단

도수분포표 문제는 전체 수를 기준으로 상대도수를 도수로 바꾸고, 필요한 범위는 해당 구간들을 직접 합산한 뒤 선지의 수치와 비교하면 된다.

암기 포인트

상대도수=도수/전체도수이므로 전체가 100이면 상대도수의 소수 둘째 자리 값이 그대로 인원수로 연결된다.

NoteBase

문 2. X_1, X_2, X_3는 모평균이 μ인 모집단으로부터 추출한 확률표본일 때, μ의 불편추정량으로 옳은 것만을 모두 고르면? (단, μ는 0이 아니다)

ㄱ. $\frac{1}{3}X_1 + \frac{2}{3}X_2 - X_3$
ㄴ. $\frac{1}{3}X_1 - X_2 + \frac{2}{3}X_3$
ㄷ. $X_1 + \frac{1}{2}X_2 - \frac{1}{2}X_3$
ㄹ. $\frac{1}{2}X_1 + \frac{2}{3}X_2 - \frac{1}{6}X_3$

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄹ
④ ㄷ, ㄹ

정답 4번

출제 포인트

이 문제는 확률표본의 선형결합이 모평균 μ의 불편추정량이 되는 조건, 즉 계수합이 1인지 판단할 수 있는지를 묻는다.

조건 정리

주어진 값은 X_1, X_2, X_3가 모두 모평균 μ를 갖는 확률표본이라는 점이다.
구해야 할 것은 각 식의 기댓값이 μ와 같은지 여부이다.
주의할 조건은 $\mu \neq 0$이므로 계수합이 0이면 기댓값이 0이 되어 불편추정량이 될 수 없다는 점이다.

사용할 공식

선형성에 의해 $E(aX_1 + bX_2 + cX_3) = (a + b + c)\mu$이다.
이 공식을 쓰는 이유는 확률표본의 각 관측치가 모두 같은 평균 μ를 가지므로, 선형결합의 기댓값은 계수합에 μ를 곱한 값으로 바로 정리되기 때문이다.

계산 과정

ㄱ은 $E\left(\frac{1}{3}X_1 + \frac{2}{3}X_2 - X_3\right) = \left(\frac{1}{3} + \frac{2}{3} - 1\right)\mu = 0$ 이므로 불편추정량이 아니다.
ㄴ은 $E\left(\frac{1}{3}X_1 - X_2 + \frac{2}{3}X_3\right) = \left(\frac{1}{3} - 1 + \frac{2}{3}\right)\mu = 0$ 이므로 불편추정량이 아니다.
ㄷ은 $E\left(X_1 + \frac{1}{2}X_2 - \frac{1}{2}X_3\right) = \left(1 + \frac{1}{2} - \frac{1}{2}\right)\mu = \mu$ 이므로 불편추정량이다.
ㄹ은 $E\left(\frac{1}{2}X_1 + \frac{2}{3}X_2 - \frac{1}{6}X_3\right) = \left(\frac{1}{2} + \frac{2}{3} - \frac{1}{6}\right)\mu = \mu$ 이므로 불편추정량이다.

NoteBase

선지 판단

1번 X: ㄱ, ㄴ을 고른 조합인데 두 식의 계수합이 모두 0이어서 기댓값이 0이다. 올바르게는 불편추정량은 ㄷ, ㄹ이다.
2번 X: ㄱ, ㄷ을 고른 조합인데 ㄱ이 불편추정량이 아니다. 올바르게는 ㄱ 대신 ㄹ이 들어가야 한다.
3번 X: ㄴ, ㄹ을 고른 조합인데 ㄴ이 불편추정량이 아니다. 올바르게는 ㄴ 대신 ㄷ이 들어가야 한다.
4번 O: ㄷ, ㄹ만 고른 조합이고 두 식의 계수합이 각각 1이므로 둘 다 불편추정량이다.

오답 포인트

계수 앞 숫자만 보고 복잡한 식을 피상적으로 판단하면 틀리기 쉽다. 이 문항의 함정은 음의 계수가 있어도 계수합만 1이면 불편추정량이 될 수 있다는 점과, $\mu \neq 0$ 조건 때문에 계수합 0을 결코 μ로 착각하면 안 된다는 점이다.

다시 나오면 이렇게 판단

모평균의 불편추정량인지 묻는 선형결합 문제는 먼저 각 항의 계수를 모두 더해 1인지 확인하고, 그다음 기댓값이 정확히 μ로 남는지 검산하는 순서로 판단한다.

암기 포인트

확률표본의 선형결합에서 모평균의 불편추정량 조건은 계수합이 1이라는 점이다.

NoteBase

문 3. 다음은 n개의 자료 (x_i, y_i)에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$를 최소제곱법으로 적합하여 얻은 분산분석표의 일부이다. 옳은 것만을 모두 고르면? (단, ε_i는 $N(0, \sigma^2)$을 따르고 서로 독립이다)

요인	제곱합	자유도	평균제곱	$F-$값
회귀	SSR		MSR	F_0
잔차	SSE	10	MSE	

$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 9$, $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 12$, $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 6$일 때,

ㄱ. $SSR = 4$

ㄴ. $MSE = 0.8$

ㄷ. $F_0 = 5$

① ㄱ, ㄴ

② ㄱ, ㄷ

③ ㄴ, ㄷ

④ ㄱ, ㄴ, ㄷ

정답 4번

출제 포인트

이 문제는 단순선형회귀의 제곱합 분해와 분산분석표 공식을 이용해 SSR, MSE, F_0를 계산할 수 있는지를 묻는 문제이다.

조건 정리

주어진 값은 $S_{xx} = \sum(x_i - \bar{x})^2 = 9, S_{yy} = \sum(y_i - \bar{y})^2 = 12, S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = 6$이다.
잔차 자유도는 10이므로 단순선형회귀에서 $n - 2 = 10$이다.
구해야 할 값은 ㄱ의 SSR, ㄴ의 MSE, ㄷ의 F_0이다.

사용할 공식

단순선형회귀에서 $\hat{\beta}_1 = S_{xy}/S_{xx}, SSR = \hat{\beta}_1 S_{xy} = S_{xy}^2/S_{xx}, SSE = S_{yy} - SSR$이다.
또한 $MSR = SSR/1, MSE = SSE/(n-2), F_0 = MSR/MSE$이다.
이 공식을 쓰는 이유는 문제에서 단순선형회귀의 분산분석표 일부와 S_{xx}, S_{yy}, S_{xy}를 함께 주었기 때문이다.

NoteBase

계산 과정

1. $SSR = S_{xy}^2/S_{xx} = 6^2/9 = 36/9 = 4$이므로 ㄱ은 참이다.
2. $SSE = S_{yy} - SSR = 12 - 4 = 8$이다.
3. 잔차 자유도가 10이므로 $MSE = SSE/10 = 8/10 = 0.8$이어서 ㄴ은 참이다.
4. 회귀 자유도는 1이므로 $MSR = SSR/1 = 4$이다.
5. 따라서 $F_0 = MSR/MSE = 4/0.8 = 5$이므로 ㄷ도 참이다.

선지 판단

① X: ㄱ과 ㄴ은 맞지만 ㄷ도 맞다. 따라서 'ㄱ, ㄴ'만 고른 것은 틀리고, 올바르게는 'ㄱ, ㄴ, ㄷ'이다.
② X: ㄱ과 ㄷ은 맞지만 ㄴ도 맞다. 따라서 'ㄱ, ㄷ'만 고른 것은 틀리고, 올바르게는 'ㄱ, ㄴ, ㄷ'이다.
③ X: ㄴ과 ㄷ은 맞지만 ㄱ도 맞다. 따라서 'ㄴ, ㄷ'만 고른 것은 틀리고, 올바르게는 'ㄱ, ㄴ, ㄷ'이다.
④ O: $SSR = 4$, $MSE = 0.8$, $F_0 = 5$이므로 ㄱ, ㄴ, ㄷ이 모두 옳다.

오답 포인트

총제곱합 S_{yy}를 바로 자유도로 나누어 MSE를 구하거나, SSR을 구하지 않고 F_0를 추정하면 틀린다. 특히 MSE의 분모는 표에 주어진 잔차 자유도 10이고, F_0는 MSR/MSE라는 점을 놓치기 쉽다.

다시 나오면 이렇게 판단

회귀 분산분석 문제는 먼저 S_{xy}, S_{xx}로 SSR을 구하고, 다음에 $SSE = S_{yy} - SSR$을 계산한 뒤, 마지막에 자유도를 반영해 MSE와 F값을 구하는 순서로 판단한다.

암기 포인트

단순회귀 ANOVA는 $SSR = S_{xy}^2/S_{xx}$, $SSE = S_{yy} - SSR$, $F = MSR/MSE$의 흐름으로 정리된다.

NoteBase

문 4. 다음은 4가지 광고유형에 따라 상품판매량의 차이가 있는지를 알아보기 위해 얻은 분산분석표의 일부이다. 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	$F-$값	$p-$값
처리	36				0.022
오차		20			
계	96				

① 처리평균제곱은 12이다.

② F-검정통계량의 값은 4이다.

③ 총자료의 개수는 23이다.

④ 유의수준 5%에서 광고유형에 따라 상품판매량은 차이가 있다.

정답 3번

출제 포인트

이 문제는 일원분산분석표에서 자유도, 평균제곱, F값, 총자료 수의 관계를 계산해 선지를 판정하는 문제이다.

조건 정리

주어진 값은 처리제곱합 36, 오차자유도 20, 총제곱합 96, 처리의 p값 0.022이다.
광고유형은 4가지이므로 집단 수 $k = 4$이다.
구해야 할 값은 처리자유도, 처리평균제곱, F값, 총자료의 개수이다.

사용할 공식

일원분산분석에서는 처리자유도는 $k - 1$, 오차제곱합은 $SS_E = SS_{Total} - SS_{Tr}$, 평균제곱은 $MS = SS/df$, 검정통계량은 $F = MS_{Tr}/MS_E$, 총자료의 개수는 $N = df_{Total} + 1$이다.
이 공식을 쓰는 이유는 분산분석표의 빈칸이 집단 수, 제곱합, 자유도 사이의 기본 관계로 연결되기 때문이다.

NoteBase

계산 과정

1. 광고유형이 4가지이므로 처리자유도는 $4-1=3$이다.
2. 오차제곱합은 $96-36=60$이다.
3. 처리평균제곱은 $36/3=12$, 오차평균제곱은 $60/20=3$이다.
4. 따라서 $F=12/3=4$이다.
5. 총자유도는 $3+20=23$이므로 총자료의 개수는 $23+1=24$이다.

선지별 해설

① O
처리자유도는 광고유형 4개이므로 $4-1=3$이다. 따라서 처리평균제곱은 $36/3=12$이므로 옳다.

② O
오차제곱합은 $96-36=60$이고 오차평균제곱은 $60/20=3$이다. 따라서 $F=12/3=4$이므로 옳다.

③ X
총자료의 개수 N은 총자유도와 같지 않고 총자유도에 1을 더한 값이다. 총자유도는 $3+20=23$이므로 총자료의 개수는 24이다. 틀린 표현은 "총자료의 개수는 23"이고, 올바르게는 "총자료의 개수는 24"이다.

④ O
p값이 0.022이므로 유의수준 5%에서 $p<0.05$이다. 따라서 귀무가설을 기각하고 광고유형에 따라 상품판매량의 평균에 차이가 있다고 판단하므로 옳다.

오답 포인트

③은 총자유도와 총자료의 개수를 같은 값으로 착각하는 오류를 노린 선지이다. 분산분석표에서는 항상 총자유도 $=N-1$이므로 자유도 23을 바로 자료 수 23으로 읽으면 틀리게 된다.

다시 나오면 이렇게 판단

분산분석표 문제는 먼저 집단 수로 처리자유도 $k-1$을 정하고, 총제곱합에서 처리제곱합을 빼서 오차제곱합을 구한 뒤, 평균제곱과 F값을 계산한다. 마지막에는 총자유도 $=N-1$ 관계로 자료 수를 한 번 더 검산한다.

암기 포인트

일원분산분석에서는 처리자유도 $=k-1$, 오차자유도 $=N-k$, 총자유도 $=N-1$이다.

NoteBase

문 5. 어느 대학교 도서관의 작년 자료에 의하면 학부생 중 도서 대출 비율은 1학년 20%, 2학년 30%, 3학년 40%, 4학년 10%라고 한다. 올해 이 도서관에서 대출한 학부생 1,000명을 임의추출하여 학년별로 대출한 학생 수를 조사한 결과가 다음과 같다. 올해 이 도서관에서 도서를 대출한 학부생들의 학년 비율이 작년과 차이가 있는지를 검정하기 위한 카이제곱 검정통계량의 값은?

학년: 1학년, 2학년, 3학년, 4학년, 계
학생 수: 210, 270, 440, 80, 1,000

① 10.1
② 11.5
③ 12.4
④ 15.0

정답 2번

출제 포인트

이 문제는 작년 비율을 기대도수로 두고 카이제곱 적합도 검정통계량 χ^2를 계산하는 문제이다.

조건 정리

주어진 값:
작년 비율은 1학년 20%, 2학년 30%, 3학년 40%, 4학년 10%이다.
올해 표본 크기는 1,000명이고 관측도수는 210명, 270명, 440명, 80명이다.
구해야 할 값:
작년 비율과 올해 학년 비율이 같은지 검정하는 카이제곱 검정통계량이다.
주의할 조건:
비율 자체를 바로 넣지 말고 각 학년의 기대도수를 먼저 200, 300, 400, 100으로 계산해야 한다.

사용할 공식

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
이 공식을 쓰는 이유:
작년의 학년 비율이 귀무가설 아래의 범주별 확률이고, 올해 표본에서 나온 관측도수와 기대도수의 차이를 범주별로 합쳐 적합도를 검정하기 때문이다.

NoteBase

계산 과정

1. 기대도수를 구하면 $1000 \times (0.2, 0.3, 0.4, 0.1) = (200, 300, 400, 100)$이다.
2. 각 항을 계산하면 $(210-200)^2/200 = 0.5$, $(270-300)^2/300 = 3$, $(440-400)^2/400 = 4$, $(80-100)^2/100 = 4$이다.
3. 따라서 $\chi^2 = 0.5 + 3 + 4 + 4 = 11.5$이다.

선지별 해설

① X: 계산 결과가 11.5이므로 10.1은 맞지 않는다. 틀린 표현은 검정통계량의 값 자체이며, 올바르게는 $\chi^2 = 11.5$이다.
② O: 기대도수를 200, 300, 400, 100으로 두고 각 범주의 $(O-E)^2/E$를 합하면 11.5가 된다.
③ X: 계산 결과가 11.5이므로 12.4는 맞지 않는다. 틀린 표현은 검정통계량의 값 자체이며, 올바르게는 $\chi^2 = 11.5$이다.
④ X: 계산 결과가 11.5이므로 15.0은 맞지 않는다. 틀린 표현은 검정통계량의 값 자체이며, 올바르게는 $\chi^2 = 11.5$이다.

오답 포인트

작년 비율 20%, 30%, 40%, 10%를 그대로 식에 넣거나 올해 표본비율 21%, 27%, 44%, 8%만 보고 판단하면 틀리기 쉽다. 적합도 검정은 먼저 기대도수를 만든 뒤 각 범주의 기여도 $(O-E)^2/E$를 더해야 한다.

다시 나오면 이렇게 판단

범주형 비율 비교에서 기준 비율이 주어지면 표본수에 그 비율을 곱해 기대도수를 먼저 만든다. 그다음 각 범주의 $(O-E)^2/E$를 계산해 모두 더한 값이 카이제곱 검정통계량이다.

암기 포인트

적합도 검정은 기준비율 → 기대도수 → $(O-E)^2/E$ 합의 순서로 푼다.

NoteBase

문 6. 기댓값(평균)이 $\frac{13}{8}$인 이산형 확률변수 X의 확률분포가 다음과 같을 때, $\frac{b}{a}$의 값은?

x	0	1	2	3	계
$P(X=x)$	$\frac{1}{4}$	a	b	$\frac{1}{4}$	1

① 1

② 2

③ 3

④ 4

정답 3번

출제 포인트

확률의 총합이 1이라는 조건과 기댓값 공식 $E(X)=\sum xP(X=x)$을 연립해 $\frac{b}{a}$를 구하는 문제이다.

조건 정리

주어진 값은 $P(X=0)=\frac{1}{4}$, $P(X=1)=a$, $P(X=2)=b$, $P(X=3)=\frac{1}{4}$, $E(X)=\frac{13}{8}$이다. 구해야 할 값은 $\frac{b}{a}$이다. 확률의 총합 1과 기댓값 식을 함께 써야 한다.

사용할 공식

확률의 총합은 1이므로 $a+b+\frac{1}{4}+\frac{1}{4}=1$이다. 기댓값은 $E(X)=\sum xP(X=x)$이므로 $0\cdot\frac{1}{4}+1\cdot a+2\cdot b+3\cdot\frac{1}{4}=\frac{13}{8}$이다. 미지수 a, b에 대해 서로 다른 식 두 개가 생기므로 이 공식을 사용한다.

계산 과정

전체 확률의 합에서 $a+b=\frac{1}{2}$를 얻는다. 기댓값 식에서 $a+2b+\frac{3}{4}=\frac{13}{8}$이므로 $a+2b=\frac{7}{8}$를 얻는다. 두 식의 차를 이용하면 $b=\frac{3}{8}$이고, 따라서 $a=\frac{1}{8}$이다. 그러므로 $\frac{b}{a}=\frac{3/8}{1/8}=3$이다.

NoteBase

선지별 해설

① X: 값 1은 틀리다. 계산하면 $b=\frac{3}{8}$, $a=\frac{1}{8}$이므로 $\frac{b}{a}=3$이다. 올바르게는 3이다.
② X: 값 2는 틀리다. $\frac{b}{a}=2$라면 $b=2a$여야 하지만, $a+b=\frac{1}{2}$와 $a+2b=\frac{7}{8}$를 함께 만족하면 $b=3a$가 된다. 올바르게는 3이다.
③ O: $a+b=\frac{1}{2}$, $a+2b=\frac{7}{8}$에서 $b=\frac{3}{8}$, $a=\frac{1}{8}$이므로 $\frac{b}{a}=3$이다.
④ X: 값 4는 틀리다. $\frac{b}{a}=4$라면 $b=4a$여야 하지만 그러면 $a+b=\frac{1}{2}$와 $a+2b=\frac{7}{8}$를 동시에 만족할 수 없다. 올바르게는 3이다.

오답 포인트

양끝 확률 $\frac{1}{4}$, $\frac{1}{4}$를 이미 사용한 뒤에 남는 확률이 $a+b=\frac{1}{2}$라는 점을 놓치기 쉽다. 또 기댓값 식에서 $0\cdot\frac{1}{4}=0$은 사라지지만 $3\cdot\frac{1}{4}=\frac{3}{4}$는 반드시 포함해야 한다.

다시 나오면 이렇게 판단

확률분포표에서 미지수가 보이면 먼저 확률의 총합 1을 식으로 세우고, 다음에 $E(X)=\sum xP(X=x)$를 적어 연립한다. 마지막에 비율을 묻더라도 각 확률값을 먼저 구한 뒤 나누면 실수가 줄어든다.

암기 포인트

확률표 미지수는 '총합 1'과 '기댓값 1개'를 먼저 연립하면 정리된다.

NoteBase

문 7. 연속형 확률변수 X의 확률밀도함수가 다음과 같을 때, $P(X > 4 \mid X > 2)$의 값은?

f_X(x)=cases(1)/(4)e^-x/4, & x>0 \ 0, & 그 외cases

① e^{-1}

② $e^{-3/4}$

③ $e^{-1/2}$

④ $e^{-1/4}$

정답 3번

출제 포인트

이 문제는 지수분포의 생존함수와 조건부확률을 이용해 $P(X > 4 \mid X > 2)$를 구하는 문제이다.

조건 정리

주어진 확률밀도함수는 $f_X(x) = \frac{1}{4}e^{-x/4}$ $(x > 0)$이다.
구해야 할 값은 $P(X > 4 \mid X > 2)$이다.
조건 $X > 2$가 이미 주어졌으므로 무조건확률 $P(X > 4)$가 아니라 2를 초과한 뒤의 추가 초과확률을 계산해야 한다.

사용할 공식

조건부확률 공식은 $P(X > 4 \mid X > 2) = \frac{P(X > 4 \cap X > 2)}{P(X > 2)} = \frac{P(X > 4)}{P(X > 2)}$ 이다.
또한 이 밀도함수는 지수분포이므로 생존함수는 $P(X > t) = e^{-t/4}$이다.
이 공식을 쓰는 이유는 $X > 4$이면 자동으로 $X > 2$도 만족하므로 분자가 $P(X > 4)$로 단순화되기 때문이다.

계산 과정

1. 생존함수를 쓰면 $P(X > 4) = e^{-4/4} = e^{-1}$이고, $P(X > 2) = e^{-2/4} = e^{-1/2}$이다.
2. 따라서 $P(X > 4 \mid X > 2) = \frac{e^{-1}}{e^{-1/2}} = e^{-1 + 1/2} = e^{-1/2}$ 이다.
3. 지수분포의 무기억성으로 보아도 $P(X > 4 \mid X > 2) = P(X > 2) = e^{-1/2}$가 된다.

NoteBase

선지별 해설

① X

e^{-1}은 $P(X>4)$라는 무조건확률이다. 조건 $X>2$를 반영하지 않은 값이라는 점이 틀렸다. 올바르게는 $P(X>4 \mid X>2)=e^{-1/2}$이다.

② X

지수 계산은 $-1-(-1/2)=-1/2$가 되어야 하므로 $e^{-3/4}$는 나오지 않는다. 지수의 차를 잘못 계산한 점이 틀렸고, 올바르게는 $e^{-1/2}$이다.

③ O

$P(X>4 \mid X>2)=e^{-(4-2)/4}=e^{-1/2}$이므로 계산 결과와 정확히 일치한다.

④ X

$e^{-1/4}$는 초과 구간의 길이를 $4-2=2$가 아니라 1로 잘못 본 값이다. 조건부확률에서 남은 구간은 2만큼이므로 올바르게는 $e^{-2/4}=e^{-1/2}$이다.

오답 포인트

조건부확률을 보자마자 $P(X>4)$를 그대로 쓰거나, 지수의 차 $(4-2)/4$를 잘못 계산하면 오답이 된다. 이 문항의 함정은 조건 $X>2$를 빼먹는 실수와 지수분포의 무기억성을 떠올리지 못하는 데 있다.

다시 나오면 이렇게 판단

지수분포에서 $P(X>a \mid X>b)$ 꼴이 나오면 먼저 $a>b$인지 확인한 뒤 $\frac{P(X>a)}{P(X>b)}$ 또는 $P(X>a-b)$ 형태로 바꾼다. 이후 생존함수 $P(X>t)=e^{-t/\theta}$에서 남은 길이만 대입해 판단한다.

암기 포인트

지수분포는 무기억성을 가져 $P(X>s+t \mid X>s)=P(X>t)$가 성립한다.

문 8. 두 확률변수 X와 Y는 서로 독립이며 각각 성공의 확률이 $\frac{1}{2}$인 베르누이(Bernoulli) 분포를 따른다고 할 때, 확률변수 $W = X + Y$의 확률질량함수는?

① $P(W=k) = \frac{2!}{k!}\left(\frac{1}{2}\right)^k,\ k=0,1,2$

② $P(W=k) = \frac{2!}{k!}\left(\frac{1}{2}\right)^2,\ k=0,1,2$

③ $P(W=k) = \frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^k,\ k=0,1,2$

④ $P(W=k) = \frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^2,\ k=0,1,2$

정답 4번

출제 포인트

독립인 두 베르누이 확률변수의 합이 이항분포를 따른다는 점과 이항분포의 확률질량함수를 정확히 쓰는지를 묻는 문제이다.

조건 정리

주어진 값: X, Y는 서로 독립이고 각각 성공확률이 1/2인 베르누이분포를 따른다.
구해야 할 값: $W = X + Y$의 확률질량함수 $P(W=k)$이다.
주의할 조건: 독립인 두 번의 성공 횟수의 합이므로 $W \sim Bin(2, 1/2)$이고 $k = 0, 1, 2$이다.

사용할 공식

이항분포의 확률질량함수는 $P(W=k) = \binom{2}{k} p^k (1-p)^{2-k}$이다.
이 공식을 쓰는 이유: 성공확률이 1/2인 독립 베르누이 시행 2회의 성공 횟수 합이므로 $n=2$인 이항분포 상황이기 때문이다.

계산 과정

1. X와 Y가 각각 베르누이(1/2)이고 서로 독립이므로 $W = X + Y \sim Bin(2, 1/2)$이다.
2. 따라서 $P(W=k) = \binom{2}{k}(1/2)^k(1-1/2)^{2-k}$이다.
3. $1 - 1/2 = 1/2$이므로 $P(W=k) = \binom{2}{k}(1/2)^k(1/2)^{2-k} = \binom{2}{k}(1/2)^2 = \frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^2$이다.

선지별 해설

① X. $\frac{2!}{k!}\left(\frac{1}{2}\right)^k$는 조합항에서 $(2-k)!$가 빠졌고, 실패확률 항 $\left(\frac{1}{2}\right)^{2-k}$도 빠졌다. 올바르게는 $\frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^2$이다.
② X. $\frac{2!}{k!}\left(\frac{1}{2}\right)^2$는 지수 부분은 단순화 결과와 같지만, 여전히 조합항의 분모 $(2-k)!$가 빠져 있다. 올바르게는 $\frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^2$이다.
③ X. $\frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^k$는 조합항은 맞지만 실패확률 항 $\left(\frac{1}{2}\right)^{2-k}$를 누락했다. 올바르게는 $\frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{2-k}$이고, 이를 정리하면 $\frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^2$이다.
④ O. $\frac{2!}{k!\,(2-k)!}\left(\frac{1}{2}\right)^2$는 $\binom{2}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{2-k}$를 정리한 식과 같다. 따라서 $k=0, 1, 2$에서의 확률질량함수로 맞다.

오답 포인트

$p = 1/2$라는 이유만 보고 p^k만 남긴 채 $(1-p)^{n-k}$를 지우기 쉽다. 이 문항의 함정은 $p = 1 - p$라는 사실과 $(1-p)^{n-k}$가 없어지는 것은 다르다는 점이다. 또한 조합항 $\binom{2}{k}$를 $2!/k!$로 잘못 쓰면 경우의 수를 빠뜨리게 된다.

다시 나오면 이렇게 판단

베르누이 확률변수 여러 개의 합이 나오면 먼저 시행 횟수 n과 성공확률 p를 확인해 이항분포인지 판단한다. 그다음 $\binom{n}{k}p^k(1-p)^{n-k}$를 그대로 쓴 뒤, 마지막 단계에서만 $p = 1/2$ 같은 값을 대입해 정리하면 항 누락을 막을 수 있다.

암기 포인트

독립인 베르누이의 합은 이항분포이고, 확률질량함수는 항상 $\binom{n}{k}p^k(1-p)^{n-k}$ 꼴이다.

NoteBase

문 9. 10개의 자료 (x_i, y_i)에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$를 최소제곱법으로 적합할 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ε_i는 $N(0, \sigma^2)$을 따르고 서로 독립이다)

ㄱ. 설명변수와 반응변수의 표본평균을 각각 $\bar{x}$와 $\bar{y}$라 할 때, 회귀직선은 항상 $(\bar{x}, \bar{y})$를 지난다.
ㄴ. 0이 아닌 총제곱합(SST)이 일정할 때, 잔차제곱합(SSE)이 작아지면 결정계수도 작아진다.
ㄷ. 가설 $H_0 : \beta_1 = 0$ 대 $H_1 : \beta_1 \neq 0$을 검정하기 위한 F분포의 분자와 분모의 자유도는 각각 1과 8이다.

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

정답 2번

출제 포인트

이 문제는 단순선형회귀에서 최소제곱추정량의 기본 성질, 결정계수와 잔차제곱합의 관계, 기울기 검정의 F분포 자유도를 함께 묻는 문제이다.

핵심 개념

- 최소제곱 회귀직선은 항상 표본평균점 $(\bar{x}, \bar{y})$을 지난다.
- 결정계수는 전체변동 중 회귀모형이 설명하는 비율이므로 $R^2 = SSR/SST = 1 - SSE/SST$이다.
- 단순선형회귀에서 $H_0 : \beta_1 = 0$ 검정의 F통계량 자유도는 분자 1, 분모 $n - 2$이다. 여기서는 $n = 10$이므로 분모 자유도는 8이다.

NoteBase

선지별 해설

① X
- ①은 ㄱ, ㄴ을 옳다고 본 조합인데, ㄱ은 맞고 ㄴ은 틀리므로 오답이다.
- 틀린 표현은 ㄴ의 'SSE이 작아지면 결정계수도 작아진다'이다.
- 실제로는 SST가 일정할 때 $R^2 = 1 - SSE/SST$이므로 SSE가 작아질수록 결정계수는 커진다.

② O
- ㄱ은 최소제곱법으로 적합한 단순선형회귀직선의 기본 성질이므로 옳다. 회귀직선은 항상 $(\bar{x}, \bar{y})$를 지난다.
- ㄷ도 옳다. 단순선형회귀에서 기울기 하나를 검정하므로 분자 자유도는 1이고, 오차자유도는 $n - 2 = 8$이다.
- 따라서 옳은 것은 ㄱ, ㄷ뿐이므로 ②가 정답이다.

③ X
- ③은 ㄴ, ㄷ을 옳다고 본 조합인데, ㄷ은 맞지만 ㄴ이 틀리므로 오답이다.
- 틀린 표현은 ㄴ의 '결정계수도 작아진다'이다.
- 올바르게는 SST가 고정되면 SSE가 감소할수록 설명된 변동 비율인 결정계수는 증가한다.

④ X
- ④는 ㄱ, ㄴ, ㄷ이 모두 옳다고 본 조합인데, ㄴ이 틀리므로 오답이다.
- ㄱ과 ㄷ은 각각 회귀직선의 평균점 통과 성질과 F검정 자유도에 관한 올바른 설명이다.
- 올바른 조합은 ㄱ, ㄷ이다.

오답 포인트

ㄴ에서 SSE와 결정계수의 방향을 거꾸로 연결하면 틀리기 쉽다. 결정계수는 설명되지 않은 변동이 아니라 설명된 변동의 비율이므로, SSE가 줄어들면 모형 적합이 좋아져 R^2는 커진다.

다시 나오면 이렇게 판단

회귀문제에서는 먼저 $(\bar{x}, \bar{y})$ 통과 여부, $R^2 = 1 - SSE/SST$ 관계, 기울기 검정 자유도 $(1, n-2)$를 바로 대입해 본다. 이 세 가지 기본 성질만 확인해도 조합형 선지는 빠르게 걸러진다.

암기 포인트

회귀직선은 평균점을 지나고, SSE ↓이면 R^2 ↑이며, 단순회귀 기울기 F검정 자유도는 $(1, n-2)$이다.

NoteBase

문 10. 모평균과 모분산이 알려지지 않은 정규모집단으로부터 임의추출한 21개의 표본에 대한 표본분산이 16이다. 모분산이 40보다 작은지를 유의수준 5%에서 검정하고자 할 때, 카이제곱 검정통계량의 값과 검정 결과를 바르게 연결한 것은? (단, $\chi^2_\alpha(k)$는 자유도가 k인 카이제곱 분포의 제$100\times(1-\alpha)$ 백분위수를 나타내고, $\chi^2_{0.05}(20)=31.41$, $\chi^2_{0.95}(20)=10.85$, $\chi^2_{0.05}(21)=32.67$, $\chi^2_{0.95}(21)=11.59$이다)

① 카이제곱 검정통계량의 값: 8, 검정 결과: 모분산이 40보다 작다고 할 수 없다

② 카이제곱 검정통계량의 값: 8, 검정 결과: 모분산이 40보다 작다고 할 수 있다

③ 카이제곱 검정통계량의 값: 11, 검정 결과: 모분산이 40보다 작다고 할 수 없다

④ 카이제곱 검정통계량의 값: 11, 검정 결과: 모분산이 40보다 작다고 할 수 있다

정답 2번

출제 포인트

이 문제는 정규모집단의 모분산에 대한 단측 카이제곱검정에서 검정통계량과 기각역을 정확히 정하는지를 묻는다.

조건 정리

주어진 값: 표본크기 21, 표본분산 16, 검정할 기준 모분산 40이다.
구해야 할 값: 카이제곱 검정통계량과 유의수준 5%에서의 검정 결론이다.
주의할 조건: 대립가설이 모분산이 40보다 작다는 내용이므로 좌측단측검정이며 자유도는 20이다.

사용할 공식

검정통계량은 $\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$ 이다.
이 공식을 쓰는 이유: 모평균과 모분산이 알려지지 않은 정규모집단에서 모분산 σ^2를 검정할 때 $\frac{(n-1)S^2}{\sigma_0^2}$가 자유도 $n-1$인 카이제곱분포를 따르기 때문이다.

계산 과정

1. 가설을 $H_0:\sigma^2=40, H_1:\sigma^2<40$로 둔다.
2. 검정통계량은 $\chi^2=\frac{(21-1)\cdot 16}{40}=\frac{320}{40}=8$ 이다.
3. 자유도 20의 좌측단측검정이므로 기각역은 $\chi^2\le\chi^2_{0.95}(20)=10.85$ 이다.
4. 계산된 값 8은 10.85보다 작으므로 H_0를 기각한다. 따라서 모분산이 40보다 작다고 할 수 있다.

NoteBase

선지 판단

① X: 검정통계량 8은 맞지만 결론이 틀리다. 8은 기각값 10.85보다 작으므로 귀무가설을 기각한다. 틀린 표현은 '모분산이 40보다 작다고 할 수 없다'이고, 올바르게는 '모분산이 40보다 작다고 할 수 있다'이다.
② O: 검정통계량이 8이고, 8 \le 10.85이므로 유의수준 5%에서 모분산이 40보다 작다고 할 수 있다.
③ X: 결론은 '모분산이 40보다 작다고 할 수 없다'로 잘못 연결되었고 검정통계량도 틀리다. 올바른 검정통계량은 11이 아니라 8이며, 올바르게는 '8, 모분산이 40보다 작다고 할 수 있다'이다.
④ X: 결론은 맞지만 검정통계량이 틀리다. 올바른 검정통계량은 $\frac{20 \cdot 16}{40} = 8$이며, 11은 이 문제의 공식 대입으로 나오지 않는다. 올바르게는 '8, 모분산이 40보다 작다고 할 수 있다'이다.

오답 포인트

좌측단측검정인데도 우측 임계값 $\chi^2_{0.05}(20) = 31.41$를 쓰거나, 카이제곱분포 표의 아래꼬리 임계값이 $\chi^2_{0.95}(20) = 10.85$라는 점을 놓치면 결론을 반대로 낼 수 있다.

다시 나오면 이렇게 판단

모분산 검정은 먼저 대립가설 방향을 보고 좌측인지 우측인지 정한 다음, $\frac{(n-1)S^2}{\sigma_0^2}$를 계산하고 그 방향에 맞는 임계값과 비교하면 된다.

암기 포인트

모분산 좌측검정은 '작다'이면 작은 카이제곱값에서 기각하며, 검정통계량은 항상 $(n-1)S^2/\sigma_0^2$이다.

NoteBase

문 11. 자료의 개수가 100인 어느 자료에서 최댓값 a가 $a+10$으로 되었을 때, 값이 변하지 않는 측도만을 모두 고르면? (단, 자료의 값은 모두 다르다)

ㄱ. 평균
ㄴ. 중앙값
ㄷ. 분산
ㄹ. 사분위수범위

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄹ
④ ㄷ, ㄹ

정답 3번

출제 포인트

이 문제는 한 개의 극단값이 변할 때 위치 기반 측도와 값의 크기를 직접 반영하는 측도를 구분하는 문제이다.

핵심 개념

평균은 모든 자료값의 합을 자료 수로 나눈 값이므로 한 값이 바뀌면 함께 변한다.
중앙값은 순서대로 세웠을 때 가운데 위치의 값으로 정해지므로 최댓값만 바뀌고 순위가 그대로이면 변하지 않는다.
분산은 각 값이 평균에서 얼마나 떨어져 있는지를 제곱해 반영한 산포도이므로 한 값과 평균이 바뀌면 달라진다.
사분위수범위는 $Q_3 - Q_1$이며 가운데 50% 구간의 폭이므로 최댓값 하나의 변화만으로 Q_1과 Q_3가 바뀌지 않으면 그대로이다.

NoteBase

선지별 해설

① X
'ㄱ, ㄴ'에서 틀린 부분은 'ㄱ. 평균'이다. 최대값 1개가 10 커지면 전체 합이 10 커져 평균은 10/100 = 0.1만큼 증가한다. 중앙값은 100개 자료의 50번째와 51번째 값으로 정해지므로 그대로이다. 올바르게는 'ㄴ, ㄹ'이다.

② X
'ㄱ, ㄷ'는 둘 다 변하는 측도이므로 발문의 '값이 변하지 않는 측도'와 맞지 않는다. 평균은 0.1 증가하고, 분산도 최대값 자체와 평균이 함께 바뀌어 달라진다. 올바르게는 'ㄴ, ㄹ'이다.

③ O
'ㄴ, ㄹ'이 변하지 않는다. 최대값만 100번째 위치에서 커지고 나머지 1번째부터 99번째까지의 순서는 그대로이므로 중앙값은 변하지 않는다. 사분위수범위도 Q_1과 Q_3가 가운데 순위에서 정해져 최댓값 변화의 영향을 받지 않는다.

④ X
'ㄷ, ㄹ'에서 틀린 부분은 'ㄷ. 분산'이다. 분산은 극단값 변화에 민감한 측도이므로 최대값이 10 커지면 반드시 달라진다. 변하지 않는 것은 'ㄹ. 사분위수범위'와 'ㄴ. 중앙값'이다. 올바르게는 'ㄴ, ㄹ'이다.

오답 포인트

중앙값과 사분위수범위는 순위로 정해지고, 평균과 분산은 실제 값의 크기를 반영한다는 점을 섞어 놓은 함정이다. 특히 분산은 나머지 99개가 그대로여도 정의상 극단값 1개의 변화만으로 달라진다.

다시 나오면 이렇게 판단

한 값만 바뀌는 문제에서는 먼저 그 변화가 자료의 순위를 바꾸는지 본다. 순위가 안 바뀌면 중앙값과 사분위수 같은 위치 통계량은 그대로이고, 합이나 편차를 쓰는 평균과 분산은 변한다고 판단한다.

암기 포인트

극단값 변화에서는 평균과 분산은 민감하고, 중앙값과 사분위수범위는 순위가 그대로이면 유지된다.

NoteBase

문 12. 두 확률변수 X와 Y의 결합확률분포가 다음과 같다. 이에 대한 설명으로 옳은 것만을 모두 고르면?

X \ Y	-1	0	1
-1	0	0.4	0
1	0.3	0	0.3

ㄱ. $E(XY)$는 $E(X) \times E(Y)$와 같다.

ㄴ. X와 Y는 독립이다.

ㄷ. X와 Y의 상관계수는 0이다.

① ㄱ, ㄴ

② ㄱ, ㄷ

③ ㄴ, ㄷ

④ ㄱ, ㄴ, ㄷ

정답 2번

출제 포인트

이 문제는 결합확률분포표에서 주변분포를 구한 뒤 $E(X)$, $E(Y)$, $E(XY)$, 공분산과 상관계수를 판단하는 문제이다.

조건 정리

주어진 값:

X는 -1, 1을 갖고 Y는 -1, 0, 1을 갖는다.

$P(X=-1, Y=0)=0.4$, $P(X=1, Y=-1)=0.3$, $P(X=1, Y=1)=0.3$이고 나머지는 0이다.

구해야 할 값:

ㄱ의 $E(XY)=E(X)E(Y)$ 성립 여부, ㄴ의 독립 여부, ㄷ의 상관계수 0 여부를 판단한다.

주의할 조건:

독립은 $E(XY)=E(X)E(Y)$ 하나만으로 판단할 수 없고, 상관계수는 공분산이 0인지까지 계산해야 한다.

NoteBase

사용할 공식

주변기댓값은 $E(X)=\sum_x xP(X=x)$, $E(Y)=\sum_y yP(Y=y)$이다.
곱의 기댓값은 $E(XY)=\sum_x\sum_y xyP(X=x,y)$이다.
상관계수는 $\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$, 공분산은 $\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$이다.
이 공식을 쓰는 이유:
문제에서 결합확률분포표를 직접 주었으므로 주변분포와 곱의 기댓값을 계산하면 세 문장을 바로 검증할 수 있다.

계산 과정

1. 주변분포를 구하면 $P(X=-1)=0.4$, $P(X=1)=0.6$, $P(Y=-1)=0.3$, $P(Y=0)=0.4$, $P(Y=1)=0.3$이다.
2. 따라서 $E(X)=(-1)\cdot 0.4+1\cdot 0.6=0.2$, $E(Y)=(-1)\cdot 0.3+0\cdot 0.4+1\cdot 0.3=0$이다.
3. $E(XY)=(-1)(0)\cdot 0.4+1(-1)\cdot 0.3+1(1)\cdot 0.3=0$이다. 따라서 $E(X)E(Y)=0.2\cdot 0=0$이므로 ㄱ은 참이다.
4. 독립이면 모든 (x,y)에서 $P(X=x,Y=y)=P(X=x)P(Y=y)$가 성립해야 한다. 그런데 $P(X=-1,Y=0)=0.4$인데 $P(X=-1)P(Y=0)=0.4\cdot 0.4=0.16$이므로 ㄴ은 거짓이다.
5. 공분산은 $\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=0-0=0$이다. $\sigma_X>0$, $\sigma_Y>0$이므로 상관계수는 0이고 ㄷ은 참이다.

선지 판단

① X: ㄱ은 참이지만 ㄴ이 거짓이므로 틀리다. 틀린 표현은 'X와 Y는 독립이다'이고, 올바르게는 'X와 Y는 독립이 아니다'이다.
② O: ㄱ과 ㄷ이 모두 참이므로 맞다. $E(XY)=E(X)E(Y)=0$이고 공분산도 0이어서 상관계수는 0이다.
③ X: ㄴ이 거짓이므로 틀리다. 틀린 표현은 'X와 Y는 독립이다'이고, 올바르게는 'X와 Y는 독립이 아니다'이다.
④ X: ㄴ이 거짓이므로 틀리다. 틀린 표현은 'ㄱ, ㄴ, ㄷ이 모두 옳다'는 판단이고, 올바르게는 'ㄱ과 ㄷ만 옳다'이다.

오답 포인트

이 문항의 함정은 $E(XY)=E(X)E(Y)$ 또는 상관계수 0을 곧바로 독립으로 오해하는 데 있다. 독립이면 무상관이지만, 무상관이라고 해서 반드시 독립은 아니다.

다시 나오면 이렇게 판단

결합분포 문제는 주변분포 계산 -> $E(X)$, $E(Y)$, $E(XY)$ 계산 -> 공분산 확인 -> 독립성은 $P(X,Y)=P(X)P(Y)$ 직접 대조의 순서로 판단한다. $E(XY)=E(X)E(Y)$나 상관계수 0만 보고 독립이라고 결론내리면 안 된다.

NoteBase

암기 포인트

독립이면 무상관이지만, 무상관이라고 해서 독립은 아니다.

NoteBase

문 13. 자유도가 2보다 큰 t분포에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. 표준정규분포보다 t분포의 분산이 더 크다.
ㄴ. t분포는 자유도가 커질수록 표준정규분포에 수렴한다.
ㄷ. t분포의 확률밀도함수는 0을 중심으로 좌우 대칭이고 종 모양이다.

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

정답 4번

출제 포인트

자유도가 2보다 큰 t분포의 분산, 자유도 증가에 따른 극한분포, 0 대칭성을 함께 묻는 문제이다.

핵심 개념

자유도 ν인 t분포는 0을 중심으로 좌우 대칭인 연속확률분포이다.
자유도 $\nu > 2$이면 분산은 $\frac{\nu}{\nu-2}$이고, 이는 항상 1보다 크므로 표준정규분포보다 분산이 크다.
또한 자유도가 커질수록 꼬리가 얇아지면서 t분포는 표준정규분포에 수렴한다.

선지별 해설

① X: ㄱ과 ㄴ은 옳지만 ㄷ도 옳다. t분포의 확률밀도함수는 0을 중심으로 한 좌우 대칭의 종 모양이므로, 올바르게는 ㄱ·ㄴ·ㄷ을 모두 골라야 한다.
② X: ㄱ과 ㄷ은 옳지만 ㄴ도 옳다. 자유도가 커질수록 t분포는 표준정규분포에 수렴하므로, 올바르게는 ㄱ·ㄴ·ㄷ을 모두 골라야 한다.
③ X: ㄴ과 ㄷ은 옳지만 ㄱ도 옳다. 자유도 $\nu > 2$이면 분산이 $\frac{\nu}{\nu-2}$로 1보다 크므로, 올바르게는 ㄱ·ㄴ·ㄷ을 모두 골라야 한다.
④ O: ㄱ은 표준정규분포의 분산 1보다 t분포의 분산 $\frac{\nu}{\nu-2}$가 더 크다는 뜻이므로 옳다. ㄴ은 자유도가 증가할수록 t분포가 표준정규분포에 가까워진다는 뜻이므로 옳다. ㄷ은 t분포가 0 대칭의 종 모양 밀도함수를 가진다는 뜻이므로 옳다.

오답 포인트

자유도가 커지면 표준정규분포에 수렴한다는 사실을 보고 분산도 곧바로 1이라고 착각하기 쉽다. 그러나 유한한 자유도 $\nu > 2$에서는 분산이 여전히 $\frac{\nu}{\nu-2} > 1$이다. 또한 대칭성, 종 모양, 분산 크기는 서로 다른 성질이므로 따로 확인해야 한다.

NoteBase

다시 나오면 이렇게 판단

t분포 문제는 세 축으로 판단한다. 첫째, 자유도 조건에 따라 평균과 분산 존재 여부를 본다. 둘째, 자유도가 커질수록 표준정규분포에 수렴하는지를 본다. 셋째, 밀도함수가 0 중심의 좌우 대칭 종 모양인지를 확인한다. 이 세 가지가 모두 맞으면 ㄱ·ㄴ·ㄷ을 모두 고르는 선택지가 정답이다.

암기 포인트

t_ν는 0 대칭, $\nu > 2$이면 분산은 $\frac{\nu}{\nu-2} > 1$, $\nu \to \infty$이면 $N(0, 1)$로 간다.

문 14. 어느 회사의 연마제 품질이 인자 A의 3가지 수준과 인자 B의 6가지 수준에 따라 차이가 있는지를 알아보고자 한다. 다음은 인자 A와 B에 대하여 반복이 없는 이원배치법으로부터 얻은 분산분석표의 일부이다. 이에 대한 설명으로 옳은 것은? (단, $F_\alpha(k_1, k_2)$는 분자와 분모의 자유도가 각각 k_1, k_2인 F분포의 제 $100 \times (1-\alpha)$ 백분위수를 나타낼 때, $F_{0.05}(2, 10) = 4.10$, $F_{0.05}(5, 10) = 3.33$, $F_{0.05}(3, 18) = 3.16$, $F_{0.05}(6, 18) = 2.66$이다)

요인	제곱합	자유도	평균제곱	$F-$값
인자 A	14			
인자 B	28			
오차	20			

① 평균제곱오차(MSE)는 2.0이다.

② 오차제곱합(SSE)의 자유도는 총실험횟수에서 1을 뺀 값이다.

③ 유의수준 5%에서 인자 A의 수준에 따라 연마제 품질에 차이가 있다.

④ 유의수준 5%에서 인자 B의 수준에 따라 연마제 품질에 차이가 있다.

정답 1번

출제 포인트

반복이 없는 이원배치 분산분석에서 자유도, 평균제곱, F통계량을 계산해 각 선지의 옳고 그름을 판단하는 문제이다.

조건 정리

- 인자 A의 수준 수는 3이다.
- 인자 B의 수준 수는 6이다.
- 제곱합은 $SSA = 14$, $SSB = 28$, $SSE = 20$이다.
- 반복이 없는 이원배치법이므로 오차 자유도는 $(a-1)(b-1)$이다.
- 유의수준 5%의 임계값은 $F_{0.05}(2, 10) = 4.10$, $F_{0.05}(5, 10) = 3.33$이다.

NoteBase

사용할 공식

자유도는 $df_A = a - 1$, $df_B = b - 1$, $df_E = (a-1)(b-1)$이다.
평균제곱은 $MSA = SSA/df_A$, $MSB = SSB/df_B$, $MSE = SSE/df_E$이다.
검정통계량은 $F_A = MSA/MSE$, $F_B = MSB/MSE$이다.
이 공식을 쓰는 이유는 반복이 없는 이원배치 분산분석에서는 인자 A, 인자 B의 효과를 오차평균제곱과 비교해 유의성을 판단하기 때문이다.

계산 과정

1. 자유도를 구하면 $df_A = 3 - 1 = 2$, $df_B = 6 - 1 = 5$, $df_E = (3-1)(6-1) = 10$이다.
2. 평균제곱을 구하면 $MSE = 20/10 = 2.0$, $MSA = 14/2 = 7$, $MSB = 28/5 = 5.6$이다.
3. 검정통계량을 구하면 $F_A = 7/2.0 = 3.5$, $F_B = 5.6/2.0 = 2.8$이다.
4. $3.5 < 4.10$, $2.8 < 3.33$이므로 유의수준 5%에서 인자 A, 인자 B 모두 유의하지 않다.

선지별 해설

① O
- $MSE = SSE/df_E = 20/10 = 2.0$이므로 옳다.

② X
- 오차제곱합의 자유도는 총실험횟수에서 1을 뺀 값이 아니라 반복이 없는 이원배치법에서 $(a-1)(b-1)$이다.
- 이 문항에서는 $(3-1)(6-1) = 10$이다.
- 틀린 표현은 "총실험횟수에서 1을 뺀 값"이고, 올바르게는 "총실험횟수에서 1을 뺀 값은 총자유도이며, 오차자유도는 $(a-1)(b-1)$이다"이다.

③ X
- 인자 A의 검정통계량은 $F_A = 7/2.0 = 3.5$이고 임계값은 $F_{0.05}(2, 10) = 4.10$이다.
- $3.5 < 4.10$이므로 귀무가설을 기각할 수 없어 인자 A 수준에 따른 차이가 있다고 볼 수 없다.
- 틀린 표현은 "차이가 있다"이고, 올바르게는 "유의수준 5%에서 차이가 있다고 할 수 없다"이다.

④ X
- 인자 B의 검정통계량은 $F_B = 5.6/2.0 = 2.8$이고 임계값은 $F_{0.05}(5, 10) = 3.33$이다.
- $2.8 < 3.33$이므로 귀무가설을 기각할 수 없어 인자 B 수준에 따른 차이가 있다고 볼 수 없다.
- 틀린 표현은 "차이가 있다"이고, 올바르게는 "유의수준 5%에서 차이가 있다고 할 수 없다"이다.

오답 포인트

오차 자유도를 총자유도와 혼동하거나, 평균제곱을 구하지 않고 제곱합만 비교하면 오답이 된다. 또 F값이 임계값보다 작은데도 곧바로 차이가 있다고 판단하기 쉽다.

NoteBase

다시 나오면 이렇게 판단

반복이 없는 이원배치 분산분석 문항은 먼저 자유도 $a-1$, $b-1$, $(a-1)(b-1)$를 정한 뒤 MS와 F를 계산하고, 마지막에 F값과 임계값의 대소를 비교해 유의 여부를 판단한다.

암기 포인트

반복 없음 이원배치법은 오차 자유도 $(a-1)(b-1)$, 검정통계량은 $F = MS/MSE$, 유의 판정은 F가 임계값보다 커야 성립한다.

NoteBase

문 15. 연속형 확률변수 X의 누적분포함수가 다음과 같을 때, X의 분산은?

F_X(x)=cases0, & x ≤ 0 \ x^2, & 0 < x ≤ 1 \ 1, & x > 1cases

① $\frac{1}{18}$
② $\frac{4}{9}$
③ $\frac{1}{2}$
④ $\frac{2}{3}$

정답 1번

출제 포인트

이 문제는 누적분포함수를 미분해 밀도함수를 구한 뒤, $E[X]$, $E[X^2]$, $Var(X)$를 계산할 수 있는지를 묻는다.

조건 정리

주어진 누적분포함수는 $F_X(x) = 0\ (x \le 0)$, $F_X(x) = x^2\ (0 < x \le 1)$, $F_X(x) = 1\ (x > 1)$이다.
구해야 할 값은 X의 분산이다.
주의할 점은 누적분포함수 자체가 아니라 이를 미분한 밀도함수로 기댓값을 계산해야 한다는 점이다.

사용할 공식

밀도함수는 $f_X(x) = F'_X(x)$이므로 $0 < x < 1$에서 $f_X(x) = 2x$이다.
분산 공식은 $Var(X) = E[X^2] - (E[X])^2$이다.
이 공식을 쓰는 이유는 연속형 확률변수의 분산이 2차 적률과 1차 적률로 바로 계산되기 때문이다.

계산 과정

1. 누적분포함수를 미분하면 $f_X(x) = 2x\ (0 < x < 1)$이다.
2. 기댓값은 $E[X] = \int_0^1 x \cdot 2x dx = 2\int_0^1 x^2 dx = \frac{2}{3}$이다.
3. 제곱의 기댓값은 $E[X^2] = \int_0^1 x^2 \cdot 2x dx = 2\int_0^1 x^3 dx = \frac{1}{2}$이다.
4. 따라서 $Var(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$이다.

선지 판단

① O: 계산 결과 분산은 $\frac{1}{18}$이므로 옳다.
② X: $\frac{4}{9}$는 $(E[X])^2 = \left(\frac{2}{3}\right)^2$에 해당한다. 틀린 표현은 '분산이 $\frac{4}{9}$이다'이고, 올바르게는 '분산 계산 중 빼 주는 항인 $(E[X])^2$가 $\frac{4}{9}$이다'이다.
③ X: $\frac{1}{2}$는 $E[X^2]$ 값이다. 틀린 표현은 '분산이 $\frac{1}{2}$이다'이고, 올바르게는 '제곱의 기댓값 $E[X^2]$가 $\frac{1}{2}$이다'이다.
④ X: $\frac{2}{3}$는 기댓값 $E[X]$이다. 틀린 표현은 '분산이 $\frac{2}{3}$이다'이고, 올바르게는 '평균 $E[X]$가 $\frac{2}{3}$이다'이다.

NoteBase

오답 포인트

②, ③, ④는 각각 $(E[X])^2$, $E[X^2]$, $E[X]$를 분산과 혼동하도록 만든 선택지이다. 누적분포함수를 미분하지 않고 바로 값을 대입하거나, 분산 공식을 끝까지 정리하지 않으면 이 함정에 걸리기 쉽다.

다시 나오면 이렇게 판단

연속형 분산 문제는 누적분포함수이면 먼저 밀도함수로 바꾸고, 그다음 $E[X]$와 $E[X^2]$를 각각 구한 뒤 마지막에 $E[X^2] - (E[X])^2$를 적용하면 된다.

암기 포인트

누적분포함수 $\rightarrow$ 밀도함수 미분, 분산 $= E[X^2] - (E[X])^2$이다.

NoteBase

문 16. 다음은 반복이 3회 있는 이원배치법으로부터 얻은 분산분석표의 일부이다. 인자 A의 수준은 4이고 인자 B의 수준은 5일 때, 옳은 것은?

요인	제곱합	자유도	평균제곱	F-값
인자 A	18	(가)		(라)
인자 B	48			(마)
교호작용 $A \times B$	36	(나)		
오차	80		(다)	

① (가)의 값과 (라)의 값은 같다.

② (나)의 값과 (마)의 값은 같다.

③ (다)의 값과 (라)의 값은 같다.

④ (라)의 값과 (마)의 값은 같다.

정답 1번

출제 포인트

이 문제는 반복이 있는 이원배치 분산분석표에서 자유도, 평균제곱, F-값의 관계를 계산하는 문제이다.

조건 정리

주어진 값은 인자 A의 수준 수 4, 인자 B의 수준 수 5, 반복 수 3, 제곱합 $SSA = 18$, $SSB = 48$, $SS_{A \times B} = 36$, $SSE = 80$이다.
구해야 할 값은 (가), (나), (다), (라), (마)의 수치와 각 보기의 참거짓이다.
주의할 조건은 반복이 있으므로 오차 자유도가 $ab(r-1)$라는 점과 F-값은 평균제곱을 오차평균제곱으로 나눈 값이라는 점이다.

사용할 공식

자유도는 $df_A = a-1$, $df_B = b-1$, $df_{A \times B} = (a-1)(b-1)$, $df_E = ab(r-1)$이다.
평균제곱은 $MS = SS/df$이다.
검정통계량은 $F_A = MS_A/MS_E$, $F_B = MS_B/MS_E$이다.
이 공식을 쓰는 이유는 표의 빈칸이 모두 이원배치 분산분석의 기본 관계식으로 연결되기 때문이다.

NoteBase

계산 과정

1. 인자 A의 자유도는 $4-1=3$이므로 (가)=3이다.
2. 교호작용의 자유도는 $(4-1)(5-1)=3\times4=12$이므로 (나)=12이다.
3. 오차 자유도는 $4\times5\times(3-1)=40$이므로 오차평균제곱은 (다)=80/40=2이다.
4. 인자 A의 평균제곱은 $18/3=6$이므로 (라)=F_A=6/2=3이다.
5. 인자 B의 자유도는 $5-1=4$, 평균제곱은 $48/4=12$이므로 (마)=F_B=12/2=6이다.

선지별 해설

① O: (가)=3, (라)=3이므로 두 값은 같다. 보기의 서술이 계산 결과와 일치하므로 옳다.
② X: (나)=12, (마)=6이므로 두 값은 같지 않다. 틀린 표현은 "같다"이고, 올바르게는 "(나)의 값은 12이고 (마)의 값은 6이므로 서로 다르다"이다.
③ X: (다)=2, (라)=3이므로 두 값은 같지 않다. 틀린 표현은 "같다"이고, 올바르게는 "(다)의 값은 2이고 (라)의 값은 3이므로 서로 다르다"이다.
④ X: (라)=3, (마)=6이므로 두 값은 같지 않다. 틀린 표현은 "같다"이고, 올바르게는 "(라)의 값은 3이고 (마)의 값은 6이므로 서로 다르다"이다.

오답 포인트

오차 자유도를 $ab(r-1)$가 아니라 전체 자유도나 단순 곱셈 자유도로 잘못 잡으면 (다)부터 모두 틀리게 된다. 또 SS, MS, F를 섞어 보면 ②, ③, ④처럼 서로 다른 종류의 값을 단순 비교하는 함정에 걸리기 쉽다.

다시 나오면 이렇게 판단

반복이 있는 이원배치 분산분석표는 먼저 수준 수로 자유도를 채우고, 다음으로 오차평균제곱을 구한 뒤,
마지막에 각 요인의 평균제곱을 오차평균제곱으로 나누어 F-값을 계산하는 순서로 판단한다.

암기 포인트

반복이 있는 이원배치법에서는 $df_E=ab(r-1)$, $MS=SS/df$, F=MS_요인/MS_E의 세 줄이 핵심이다.

NoteBase

문 17. 다음은 모분산이 알려지지 않고 서로 독립인 두 정규모집단 A와 B로부터 임의추출한 표본을 조사한 결과이다. 두 모집단 A와 B의 평균 차이인 $\mu_A - \mu_B$에 대한 95% 신뢰구간은? (단, $t_\alpha(k)$는 자유도가 k인 t 분포의 제$100 \times (1-\alpha)$ 백분위수를 나타내고, $t_{0.025}(8) = 2.306$, $t_{0.025}(9) = 2.262$이다)

모집단 (모평균, 모분산)	표본의 크기	표본평균	표본표준편차
$A\ (\mu_A,\ \sigma^2)$	4	22	3
$B\ (\mu_B,\ \sigma^2)$	6	18	1

① $4 \pm 2.262 \times 4\sqrt{\frac{5}{12}}$
② $4 \pm 2.262 \times 2\sqrt{\frac{5}{12}}$
③ $4 \pm 2.306 \times 4\sqrt{\frac{5}{12}}$
④ $4 \pm 2.306 \times 2\sqrt{\frac{5}{12}}$

정답 4번

출제 포인트

이 문제는 모분산이 같고 알려지지 않은 두 정규모집단의 평균 차이 $\mu_A - \mu_B$에 대한 95% 신뢰구간을 공통분산 추정 t구간으로 구하는 문제이다.

조건 정리

주어진 값: $n_A = 4, n_B = 6, \bar{x}_A = 22, \bar{x}_B = 18, s_A = 3, s_B = 1$이다.
구해야 할 값: $\mu_A - \mu_B$의 95% 신뢰구간이다.
주의할 조건: 두 모집단은 서로 독립이고 모분산이 같으며 알려지지 않았으므로 합동표준편차와 자유도 $n_A + n_B - 2$를 써야 한다.

사용할 공식

사용 공식은
$(\bar{x}_A - \bar{x}_B) \pm t_{0.025}(n_A + n_B - 2)s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$
이고,
$s_p^2 = \frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A + n_B - 2}$
이다.
이 공식을 쓰는 이유는 두 정규모집단이 서로 독립이고 공통 모분산 σ^2를 가진다고 주어졌기 때문이다.

NoteBase

계산 과정

1. 표본평균 차이는 $\bar{x}_A - \bar{x}_B = 22 - 18 = 4$이다.

2. 합동분산은
$s_p^2 = \frac{(4-1)\cdot 3^2 + (6-1)\cdot 1^2}{4+6-2} = \frac{27+5}{8} = 4$
이므로 합동표준편차는 $s_p = 2$이다.

3. 자유도는 $4 + 6 - 2 = 8$이므로 $t_{0.025}(8) = 2.306$을 쓴다.

4. 표준오차는
$s_p\sqrt{\frac{1}{4} + \frac{1}{6}} = 2\sqrt{\frac{5}{12}}$
이다.

5. 따라서 95% 신뢰구간은
$4 \pm 2.306 \times 2\sqrt{\frac{5}{12}}$
이다.

선지별 해설

① X: 중심값 4는 맞지만 $t_{0.025}(9) = 2.262$를 써서 자유도를 잘못 잡았고, 합동분산 $s_p^2 = 4$를 그대로 넣어 합동표준편차 $s_p = 2$로 바꾸지 못했다. 틀린 표현은 $2.262 \times 4\sqrt{\frac{5}{12}}$이고, 올바르게는 $2.306 \times 2\sqrt{\frac{5}{12}}$이다.
② X: 합동표준편차 $s_p = 2$는 맞지만 자유도를 9로 잘못 보아 $t_{0.025}(9) = 2.262$를 사용했다. 틀린 표현은 2.262이고, 올바르게는 자유도 8에 대한 2.306이다.
③ X: 자유도 8을 써서 2.306을 선택한 것은 맞지만, 합동분산 4를 표준편차처럼 사용해 계수를 4로 두었다. 틀린 표현은 $4\sqrt{\frac{5}{12}}$이고, 올바르게는 $s_p = \sqrt{4} = 2$이므로 $2\sqrt{\frac{5}{12}}$이다.
④ O: 자유도 8이므로 $t_{0.025}(8) = 2.306$을 쓰고, 합동표준편차는 $s_p = 2$이므로 신뢰구간은 $4 \pm 2.306 \times 2\sqrt{\frac{5}{12}}$가 된다.

오답 포인트

이 문제의 함정은 합동분산 s_p^2와 합동표준편차 s_p를 구분하지 못하게 만드는 점과, 자유도를 $n_A + n_B - 2$가 아니라 $n_A + n_B - 1$처럼 잘못 세게 만드는 점이다. ①과 ③은 제곱근을 빠뜨린 오류이고, ①과 ②는 자유도 선택 오류이다.

다시 나오면 이렇게 판단

두 집단 독립, 정규, 모분산 동일, 모분산 미지라는 조건이 보이면 먼저 공통분산 추정 t구간인지 확인하고, 다음으로 s_p^2를 구한 뒤 반드시 $s_p = \sqrt{s_p^2}$로 바꾸며, 마지막에 자유도 $n_1 + n_2 - 2$를 대입해 임계값을 고르면 된다.

암기 포인트

등분산 미지 두 표본 평균차 신뢰구간은 합동표준편차 s_p와 자유도 $n_1 + n_2 - 2$를 쓴다.

NoteBase

문 18. n개의 자료 (x_i, y_i)에 단순선형회귀모형을 적합하여 얻은 추정식이 $\hat{y}_i = 4 + 6x_i$일 때, 관측값 $(x_1, y_1) = (3, 24)$에 대한 잔차의 제곱값은?

① 1

② 2

③ 4

④ 9

정답 3번

출제 포인트

이 문제는 단순선형회귀식에서 예측값을 구한 뒤 잔차를 제곱해 잔차제곱값을 구하는 문제이다.

조건 정리

주어진 값: 추정식 $\hat{y}_i = 4 + 6x_i$, 관측값 $(x_1, y_1) = (3, 24)$이다.
구해야 할 값: 관측값 $(3, 24)$에서의 잔차의 제곱값이다.
주의할 조건: 잔차는 실제값에서 예측값을 뺀 값이고, 문제는 잔차가 아니라 잔차의 제곱값을 묻는다.

사용할 공식

사용할 공식은 잔차 $e_i = y_i - \hat{y}_i$와 잔차제곱 e_i^2이다.

이 공식을 쓰는 이유: 회귀식이 주어졌으므로 먼저 해당 x에서의 예측값을 구하고, 실제값과의 차이인 잔차를 계산한 뒤 이를 제곱하면 문제의 요구값이 된다.

계산 과정

1. $x_1 = 3$을 추정식에 대입하면 $\hat{y}_1 = 4 + 6 \cdot 3 = 22$이다.
2. 잔차는 $e_1 = y_1 - \hat{y}_1 = 24 - 22 = 2$이다.
3. 잔차의 제곱값은 $e_1^2 = 2^2 = 4$이다.

NoteBase

선지별 해설

① X: 1은 2^2를 잘못 계산한 값이라서 잔차제곱값이 아니다. 틀린 표현은 값이 1이라는 점이며, 올바르게는 잔차제곱값이 4이다.
② X: 2는 잔차 $24 - 22$ 자체의 값이다. 틀린 표현은 잔차와 잔차제곱을 같은 값으로 본 점이며, 올바르게는 잔차는 2이고 잔차제곱값은 4이다.
③ O: 예측값이 22, 잔차가 2이므로 잔차의 제곱값은 $2^2 = 4$가 된다. 따라서 ③이 맞다.
④ X: 9는 $(24 - 21)^2$처럼 예측값을 잘못 계산했을 때 나올 수 있는 값이다. 틀린 표현은 회귀식 대입 결과를 잘못 잡은 점이며, 올바르게는 $x = 3$에서 예측값은 22이고 잔차제곱값은 4이다.

오답 포인트

회귀식 문제에서는 예측값, 잔차, 잔차제곱을 서로 다른 값으로 구분해야 한다. 이 문항은 ②처럼 잔차 2를 그대로 고르거나, 회귀식 대입을 잘못해 ④ 같은 값을 만드는 실수를 노린 문제이다.

다시 나오면 이렇게 판단

회귀 계산형 문제는 주어진 x를 회귀식에 먼저 대입해 예측값을 구하고, 실제값과의 차이로 잔차를 계산한 뒤, 문제에서 잔차인지 잔차제곱인지 마지막에 한 번 더 확인하면 된다.

암기 포인트

잔차는 실제값 - 예측값, 잔차제곱은 그 차이를 한 번 더 제곱한 값이다.

NoteBase

문 19. 절편이 있는 단순선형회귀모형을 최소제곱법으로 적합할 때, 결정계수에 대한 설명으로 옳은 것만을 모두 고르면? (단, 총제곱합(SST)은 0이 아니다)

ㄱ. 결정계수의 값은 0보다 크거나 같고 1보다 작거나 같다.
ㄴ. 결정계수의 값이 1이면 회귀제곱합(SSR)이 0이다.
ㄷ. 결정계수는 설명변수와 반응변수의 표본상관계수와 동일하다.

① ㄱ
② ㄱ, ㄴ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

정답 1번

출제 포인트

절편이 있는 단순선형회귀에서 결정계수 R^2의 범위, $R^2 = 1$일 때의 의미, 그리고 표본상관계수와의 관계를 묻는 문제이다.

핵심 개념

결정계수 R^2: 반응변수 총변동 중 회귀식이 설명한 비율이다. $R^2 = SSR/SST = 1 - SSE/SST$이므로 절편이 있고 $SSTeq0$이면 $0 \le R^2 \le 1$이다.
회귀제곱합 SSR: 회귀식이 설명한 변동이다. $R^2 = 1$이면 설명하지 못한 변동 SSE가 0이라는 뜻이지 SSR이 0이라는 뜻이 아니다.
표본상관계수 r: 설명변수와 반응변수의 선형관계 방향과 강도를 나타낸다. 단순선형회귀에서는 $R^2 = r^2$가 성립한다.

NoteBase

선지별 해설

① O: ㄱ만 옳게 고른 선택지이다. ㄱ은 결정계수의 기본 성질을 묻는 진술로, 절편이 있는 단순선형회귀에서는 $R^2 = 1 - SSE/SST$이므로 항상 0 이상 1 이하이다.
② X: 이 선택지는 ㄴ을 함께 옳다고 본 점이 틀리다. ㄴ의 틀린 표현은 '결정계수의 값이 1이면 회귀제곱합(SSR)이 0이다'이다. 실제로는 $R^2 = 1$이면 $SSE = 0$이고, $SSR = SST$가 된다. 올바르게는 ㄱ만 옳으므로 ①이 맞다.
③ X: 이 선택지는 ㄴ과 ㄷ을 옳다고 본 점이 틀리다. ㄴ은 위와 같이 $SSR = 0$이 아니라 $SSE = 0$이어야 하고, ㄷ의 틀린 표현은 '결정계수는 … 표본상관계수와 동일하다'이다. 실제로 결정계수는 표본상관계수 자체가 아니라 그 제곱인 r^2이다. 올바르게는 ㄱ만 옳다.
④ X: 이 선택지는 ㄴ과 ㄷ을 포함해서 틀리다. ㄴ은 $R^2 = 1 \Rightarrow SSE = 0$으로 고쳐야 하고, ㄷ은 '결정계수는 표본상관계수와 동일하다'가 아니라 '결정계수는 표본상관계수의 제곱과 같다'로 고쳐야 한다. 올바르게는 ㄱ만 고른 ①이다.

오답 포인트

ㄴ에서 $R^2 = 1$을 '회귀가 아무것도 설명하지 못한다'로 거꾸로 읽으면 $SSR = 0$이라고 착각하기 쉽다. 또한 ㄷ에서 단순선형회귀의 관계식 $R^2 = r^2$를 $R^2 = r$로 잘못 외우면 상관계수의 부호까지 그대로 가져오게 된다.

다시 나오면 이렇게 판단

결정계수 문제가 나오면 먼저 $R^2 = SSR/SST = 1 - SSE/SST$를 떠올린다. 그다음 $R^2 = 1$은 오차가 0이라는 뜻인지, 단순선형회귀에서는 상관계수 자체인지 제곱인지까지 순서대로 확인하면 된다.

암기 포인트

결정계수는 설명비율이므로 $0 \le R^2 \le 1$, 완전적합이면 $SSE = 0$, 단순회귀에서는 $R^2 = r^2$이다.

NoteBase

문 20. 어느 우체국에 한 시간 동안 들어가는 사람의 수 Y는 평균이 λ인 포아송 분포를 따르는 확률변수이다. 가설 $H_0 : \lambda = 3$ 대 $H_1 : \lambda < 3$에 대한 검정에서 기각역으로 '$Y \leq 1$'을 사용할 때, 제1종 오류의 확률은?

① $2e^{-3}$

② $3e^{-3}$

③ $4e^{-3}$

④ $5e^{-3}$

정답 3번

출제 포인트

이 문제는 포아송분포에서 제1종 오류를 H_0가 참일 때의 기각확률로 계산하는 문제이다.

조건 정리

주어진 값: Y는 평균이 λ인 포아송분포를 따른다.
귀무가설: $H_0 : \lambda = 3$
대립가설: $H_1 : \lambda < 3$
기각역: $Y \leq 1$
구해야 할 값: 제1종 오류의 확률

사용할 공식

포아송분포 확률질량함수는 $P(Y = y) = e^{-\lambda}\lambda^{y}/y!$이다.
제1종 오류 확률은 α=P(기각 H_0 H_0가 참)이다.
이 문제에서는 $H_0 : \lambda = 3$이므로 $\lambda = 3$을 대입한 뒤 기각역 $Y \leq 1$의 확률을 합하면 된다.

계산 과정

1. 제1종 오류는 H_0가 참인데도 H_0를 기각하는 확률이므로 $P(Y \leq 1 \mid \lambda = 3)$을 구한다.
2. $P(Y = 0) = e^{-3}3^{0}/0! = e^{-3}$이다.
3. $P(Y = 1) = e^{-3}3^{1}/1! = 3e^{-3}$이다.
4. 따라서

$P(Y \leq 1 \mid \lambda = 3) = P(Y = 0) + P(Y = 1) = e^{-3} + 3e^{-3} = 4e^{-3}$

이다.

NoteBase

선지 판단

① X
- $2e^{-3}$는 $P(Y=0)+e^{-3}$처럼 한 항을 잘못 더한 값이다. 기각역 $Y \le 1$에는 $Y=0, 1$이 모두 들어가므로 부족하다.
- 틀린 표현 → 올바르게는 $P(Y \le 1 \mid \lambda = 3) = 4e^{-3}$이다.

② X
- $3e^{-3}$는 $P(Y=1)$만 계산한 값이다. 제1종 오류는 $Y=1$만이 아니라 $Y=0$도 함께 포함한다.
- 틀린 표현 → 올바르게는 $P(Y=0)+P(Y=1) = 4e^{-3}$이다.

③ O
- H_0 아래에서 기각역 $Y \le 1$의 확률은 $e^{-3}+3e^{-3} = 4e^{-3}$이므로 맞다.

④ X
- $5e^{-3}$는 $P(Y=2) = \frac{9}{2}e^{-3}$를 잘못 처리하거나 불필요한 항을 더한 값이다. 기각역은 $Y \le 1$이므로 $Y=2$는 포함되지 않는다.
- 틀린 표현 → 올바르게는 $Y=2$를 제외하고 $Y=0, 1$만 합한 $4e^{-3}$이다.

계산 결과 $4e^{-3}$이므로 정답은 3번이다.

오답 포인트

제1종 오류를 H_0 아래에서 계산해야 한다는 점과 기각역 $Y \le 1$에 포함되는 값을 끝까지 세어야 한다는 점이 함정이다. 특히 ②처럼 $P(Y=1)$만 구하거나 ④처럼 기각역 밖의 값을 섞으면 오답이 된다.

다시 나오면 이렇게 판단

가설검정의 확률 문제는 먼저 제1종 오류인지 제2종 오류인지 확인하고, 그다음 해당 확률을 어느 가설 아래에서 계산하는지 정한다. 이후 기각역에 포함되는 값들을 하나씩 나열해 정확히 합하면 된다.

암기 포인트

제1종 오류는 H_0가 참일 때의 기각확률이고, 포아송분포에서 $Y \le k$는 0부터 k까지의 확률을 합한 값이다.

NoteBase

문 21. 사교육이 A 대학 합격 여부와 관계가 있는지를 알아보기 위해 A 대학 지원자 300명을 대상으로 설문조사를 실시하여 다음과 같은 분할표를 얻었다. 사교육을 받은 지원자 중에서 A 대학에 합격할 오즈(odds) 값은?

A 대학 합격 여부: 합격, 불합격
사교육 받음: 합격 83, 불합격 131
사교육 받지 않음: 합격 7, 불합격 79

① $\frac{83}{7}$
② $\frac{83}{90}$
③ $\frac{83}{131}$
④ $\frac{83}{214}$

정답 3번

출제 포인트

이 문제는 2 × 2 분할표에서 사교육을 받은 집단의 합격 오즈를 확률과 구분하여 구하는 문제이다.

조건 정리

주어진 값:
사교육을 받은 지원자 중 합격은 83명, 불합격은 131명이다.
사교육을 받은 지원자 수는 모두 214명이다.
구해야 할 값:
사교육을 받은 지원자 중 A 대학에 합격할 오즈이다.
주의할 조건:
오즈는 확률이 아니라 같은 집단 안에서 성공 수와 실패 수를 비교한 값이다.

사용할 공식

오즈는 사건이 일어날 확률을 p라고 할 때 $\frac{p}{1-p}$이다.
분할표에서는 같은 집단 안의 성공 수와 실패 수를 바로 써서 성공 수실패 수로 계산할 수 있다.
이 공식을 쓰는 이유:
문제는 사교육을 받은 지원자라는 한 집단을 기준으로 합격과 불합격을 비교한 오즈를 묻고 있으므로, 받은 집단 안의 합격 수 83과 불합격 수 131을 비교해야 한다.

NoteBase

계산 과정

1. 사교육을 받은 지원자의 전체 수는 $83 + 131 = 214$이다.
2. 사교육을 받은 지원자의 합격 확률은 $\frac{83}{214}$, 불합격 확률은 $\frac{131}{214}$이다.
3. 따라서 합격 오즈는
$\frac{83/214}{131/214} = \frac{83}{131}$
이다.

선지별 해설

① X: $\frac{83}{7}$은 사교육을 받은 집단의 오즈가 아니라 사교육을 받은 합격자 수와 받지 않은 합격자 수를 비교한 값이다. 틀린 표현은 '받은 집단 내 합격 오즈'를 집단 간 합격자 비교로 바꾼 점이다. 올바르게는 같은 집단 안에서 합격 83과 불합격 131을 비교해야 한다.
② X: $\frac{83}{90}$은 전체 합격자 90명 중 사교육을 받은 사람이 83명인 비율이다. 틀린 표현은 오즈를 '합격자 중 사교육 받은 비율'로 바꾼 점이다. 올바르게는 받은 집단 안의 합격 대 불합격 비인 $\frac{83}{131}$을 써야 한다.
③ O: 사교육을 받은 집단에서 합격이 83명, 불합격이 131명이므로 합격 오즈는 $\frac{83}{131}$이다.
④ X: $\frac{83}{214}$는 사교육을 받은 지원자 중 합격 확률이다. 틀린 표현은 오즈를 확률로 바꾼 점이다. 올바르게는 확률 $\frac{83}{214}$를 오즈로 바꾸어 $\frac{83}{131}$로 계산해야 한다.

오답 포인트

이 문항의 함정은 오즈와 확률을 혼동하거나, 기준 집단을 사교육을 받은 집단으로 고정하지 않고 전체 합격자 수 90 또는 받은 집단 총수 214를 분모에 넣는 데 있다. 특히 ④는 확률값이라 가장 흔한 오답이다.

다시 나오면 이렇게 판단

오즈를 묻는 문항은 먼저 기준 집단을 한 줄로 고정한 뒤, 그 집단 안에서 사건 발생 수와 사건 미발생 수를 나눈다. 확률이면 성공 수를 전체 수로 나누고, 오즈이면 성공 수를 실패 수로 나눈다.

암기 포인트

오즈는 성공/실패, 확률은 성공/전체이다.

NoteBase

문 22. 연속형 확률변수 X의 확률밀도함수가 다음과 같을 때, X의 중앙값은? (단, log는 자연로그함수이다)

f_X(x)=cases e^-x, & x>0 \ 0, & 그 외 cases

① $\log\left(\frac{1}{2}\right)$
② $\log(2)$
③ $e^{\frac{1}{2}}$
④ e^2

정답 2번

출제 포인트

이 문제는 지수분포의 누적분포함수를 이용해 중앙값을 구하는 문제이다.

조건 정리

주어진 값: 확률밀도함수는 $f_X(x) = e^{-x}$ for $x > 0$, 그 외에는 0이다.
구해야 할 값: 중앙값 m이다.
주의할 조건: 중앙값은 평균이 아니라 $P(X \le m) = 1/2$를 만족하는 값이다.

사용할 공식

$F(x) = P(X \le x) = \int_0^x e^{-t}dt = 1 - e^{-x} \quad (x > 0)$
$F(m) = \frac{1}{2}$
이 공식을 쓰는 이유는 중앙값이 누적확률을 정확히 1/2로 만드는 지점이기 때문이다.

계산 과정

1. 누적분포함수는 $F(x) = 1 - e^{-x}$이다.
2. 중앙값 m은 $F(m) = 1/2$를 만족하므로 $1 - e^{-m} = 1/2$이다.
3. 따라서 $e^{-m} = 1/2$, $-m = \log(1/2)$이므로 $m = -\log(1/2) = \log 2$이다.

NoteBase

선지별 해설

① X: $\log(1/2) = -\log 2 < 0$이므로 중앙값이 될 수 없다. 이 분포는 $x > 0$에서만 확률밀도가 있으므로 중앙값도 양수여야 한다. 틀린 표현은 $m = \log(1/2)$이고, 올바르게는 $m = \log 2$이다.
② O: 계산 결과 중앙값은 $m = \log 2$이다. 누적확률이 1/2가 되는 지점을 정확히 나타내므로 옳다.
③ X: $e^{1/2}$는 중앙값 조건 $1 - e^{-m} = 1/2$를 풀어 얻는 값이 아니다. 중앙값은 지수함수값이 아니라 x의 값이므로, 틀린 표현은 $m = e^{1/2}$이고 올바르게는 $m = \log 2$이다.
④ X: e^{2}는 $\log 2$를 다시 지수형으로 잘못 바꾼 값이다. 중앙값은 $e^{-m} = 1/2$를 만족하는 m 자체이므로, 틀린 표현은 $m = e^{2}$이고 올바르게는 $m = \log 2$이다.

오답 포인트

중앙값 조건 $F(m) = 1/2$를 세운 뒤 마지막에 음수 부호를 놓치면 ①로 잘못 간다. 또 $e^{-m} = 1/2$에서 미지수 m를 구해야 하는데 함수값을 다시 지수형으로 바꾸면 ③이나 ④ 같은 오답이 나온다.

다시 나오면 이렇게 판단

연속형 분포의 중앙값 문제는 먼저 누적분포함수 $F(x)$를 구한 뒤 $F(m) = 1/2$를 세워 푼다고 판단하면 된다. 계산 끝에서는 정의역과 부호가 맞는지도 함께 확인하게 된다.

암기 포인트

중앙값은 밀도함수값이 아니라 누적확률이 1/2가 되는 x값이며, Exp(1)의 중앙값은 $\log 2$이다.

NoteBase

문 23. 두 사람 A와 B가 번갈아 가면서 공정한 주사위를 던져 처음으로 6의 눈이 나온 사람이 이기는 게임을 한다. 첫 번째로 A가 던질 때, B가 이길 확률은?

① $\frac{4}{9}$
② $\frac{5}{11}$
③ $\frac{1}{2}$
④ $\frac{5}{9}$

정답 2번

출제 포인트

이 문제는 번갈아 시행하는 확률 게임을 점화식으로 세워 후공인 B의 승률을 구하는 문제이다.

조건 정리

주어진 값:
- A와 B는 번갈아 공정한 주사위를 던진다.
- 각 차례에 6이 나올 확률은 $\frac{1}{6}$, 6이 아닐 확률은 $\frac{5}{6}$이다.
- A가 먼저 던진다.

구해야 할 값:
- B가 먼저 6을 만들어 이길 확률

주의할 조건:
- A와 B가 한 번씩 던졌는데 둘 다 실패하면 처음과 같은 상태로 게임이 다시 시작된다.

사용할 공식

$p = \frac{5}{6} \cdot \frac{1}{6} + \left(\frac{5}{6} \cdot \frac{5}{6}\right)p$

이 공식을 쓰는 이유:

B가 이기려면 먼저 A가 실패하고 B가 성공해야 한다. 또는 A와 B가 모두 실패한 뒤 처음과 같은 상태에서 다시 B가 이겨야 하므로 같은 확률 p가 반복된다.

NoteBase

계산 과정

1. p를 B가 이길 확률이라고 둔다.
2. 첫 순환에서 바로 B가 이길 확률은 $\frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$이다.
3. A와 B가 모두 실패할 확률은 $\frac{5}{6} \cdot \frac{5}{6} = \frac{25}{36}$이고, 그 뒤에는 다시 같은 게임이 반복된다.
4. 따라서

$p = \frac{5}{36} + \frac{25}{36}p$

이고, 이를 정리하면

$\frac{11}{36}p = \frac{5}{36}$

이므로

$p = \frac{5}{11}$

이다.

선지 판단

① X

- 틀린 표현은 B의 승률이 $\frac{4}{9}$라는 값이다. 계산 결과는 $\frac{5}{11}$이므로 올바르게는 B가 이길 확률이 $\frac{5}{11}$이다.

② O

- 점화식 $p = \frac{5}{36} + \frac{25}{36}p$를 풀면 $p = \frac{5}{11}$이므로 맞는 선지이다.

③ X

- 틀린 표현은 두 사람의 승률이 같아 $\frac{1}{2}$라는 점이다. A가 먼저 던지므로 선공 이점이 있고, 올바르게는 B의 승률이 $\frac{5}{11}$, A의 승률이 $\frac{6}{11}$이다.

④ X

- 틀린 표현은 B의 승률이 $\frac{5}{9}$라는 값이다. 첫 순환만 보거나 반복 구조를 잘못 합치면 이런 오답이 나오지만, 올바르게는 반복 실패 상황까지 반영한 $\frac{5}{11}$이다.

계산 결과 $\frac{5}{11}$이므로 정답은 2번이다.

오답 포인트

③처럼 공정한 주사위라는 말만 보고 승률이 $\frac{1}{2}$라고 생각하기 쉽다. 그러나 A가 먼저 던지는 순간 두 사람의 기회는 대칭이 아니므로 선공 이점을 반영해야 한다. 또 ①, ④처럼 한 번의 왕복만 보고 끝내면 둘 다 실패한 뒤 게임이 다시 시작되는 구조를 놓치게 된다.

다시 나오면 이렇게 판단

번갈아 시도하는 확률 게임은 한 순환 뒤 처음 상태로 돌아오는지를 먼저 본다. 처음 상태가 반복되면 '이번 순환에서 바로 성공하는 확률 + 모두 실패한 뒤 원래 확률이 다시 붙는 항'으로 점화식을 세워 판단한다.

암기 포인트

선공이 있는 반복 확률게임은 '즉시 성공 확률 + 재시작 확률 × 원래 확률'로 점화식을 세운다.

NoteBase

문 24. 서로 독립인 추정량 $\hat{\theta}_1$과 $\hat{\theta}_2$은 각각 모수 θ의 불편추정량이고 $\mathrm{Var}(\hat{\theta}_1) = 3\mathrm{Var}(\hat{\theta}_2)$을 만족한다. 추정량 $\hat{\theta}_3 = a\hat{\theta}_1 + b\hat{\theta}_2$이 θ의 불편추정량일 때, $\hat{\theta}_3$의 분산을 최소화하기 위한 상수 a와 b의 값을 바르게 연결한 것은? (단, θ는 0이 아니고 $\mathrm{Var}(\hat{\theta}_1)$과 $\mathrm{Var}(\hat{\theta}_2)$은 각각 0보다 크다)

① a: $\frac{1}{4}$, b: $\frac{3}{4}$
② a: $\frac{1}{3}$, b: $\frac{2}{3}$
③ a: $\frac{2}{3}$, b: $\frac{1}{3}$
④ a: $\frac{3}{4}$, b: $\frac{1}{4}$

정답 1번

출제 포인트

이 문제는 서로 독립인 두 불편추정량의 선형결합에서 불편성 조건을 유지한 채 분산을 최소화하는 가중치를 구하는 문제이다.

조건 정리

주어진 값: $Var(\hat{\theta}_1) = 3Var(\hat{\theta}_2)$, $\hat{\theta}_1$과 $\hat{\theta}_2$는 서로 독립이다.
구해야 할 값: $\hat{\theta}_3 = a\hat{\theta}_1 + b\hat{\theta}_2$가 불편추정량이면서 분산이 최소가 되게 하는 a, b이다.
주의할 조건: 독립이므로 공분산항이 0이고, 불편성 때문에 가중치 합 $a + b = 1$이 성립한다.

사용할 공식

$E(\hat{\theta}_3) = aE(\hat{\theta}_1) + bE(\hat{\theta}_2) = (a + b)\theta$
$Var(\hat{\theta}_3) = a^2Var(\hat{\theta}_1) + b^2Var(\hat{\theta}_2)$
이 공식을 쓰는 이유: 두 추정량이 서로 독립인 불편추정량이므로 기대값으로 불편성 조건을 만들고, 분산식에서는 공분산항 없이 가중치 최적화를 할 수 있다.

계산 과정

1. $\hat{\theta}_3$가 불편추정량이려면 $E(\hat{\theta}_3) = (a + b)\theta = \theta$이므로 $a + b = 1$이다.
2. $Var(\hat{\theta}_2) = v$라고 두면 $Var(\hat{\theta}_1) = 3v$이어서 $Var(\hat{\theta}_3) = 3a^2v + b^2v$이다.
3. $b = 1 - a$를 대입하면 $Var(\hat{\theta}_3) = v\{3a^2 + (1 - a)^2\} = v(4a^2 - 2a + 1)$이다.
4. $v > 0$이므로 $4a^2 - 2a + 1$을 최소화하면 되고, 미분하면 $8a - 2 = 0$이어서 a=14이다.
5. 따라서 b=1-a=34이다.

NoteBase

선지 판단

① O: a=14, b=34이면 $a+b=1$을 만족해 불편성이 유지되고, 분산이 더 작은 $\hat{\theta}_2$에 더 큰 가중치를 두어 최소분산 조건에도 맞는다.
② X: 틀린 표현은 a=13, b=23의 조합이다. 올바르게는 a=14, b=34이다. ②는 가중치 합은 1이지만 큰 분산을 가진 $\hat{\theta}_1$에 필요한 것보다 큰 가중치를 준다.
③ X: 틀린 표현은 a=23, b=13의 조합이다. 올바르게는 a=14, b=34이다. ③은 분산이 더 큰 $\hat{\theta}_1$에 더 큰 가중치를 두므로 최소분산 추정량이 될 수 없다.
④ X: 틀린 표현은 a=34, b=14의 조합이다. 올바르게는 a=14, b=34이다. ④는 큰 분산 추정량 쪽으로 가장 많이 치우친 선택지이다.

오답 포인트

불편성 조건 $a+b=1$만 확인하고 끝내면 틀린다. 이 문제의 함정은 합이 1이라는 조건 뒤에 역분산 가중치를 한 번 더 적용해야 한다는 점이다. 분산이 큰 추정량일수록 가중치는 작아져야 하는데 이를 반대로 이해하면 ③, ④ 같은 오답으로 간다.

다시 나오면 이렇게 판단

최소분산 선형결합 문제는 먼저 불편성으로 $a+b=1$을 세우고, 다음에 독립성을 이용해 분산식을 만든 뒤, 마지막으로 큰 분산 추정량의 가중치가 더 작아지는지 확인하면 된다.

암기 포인트

서로 독립인 불편추정량의 최소분산 가중치는 역분산 비례이다. 분산 비가 3 : 1이면 가중치 비는 1 : 3이 된다.

NoteBase

문 25. 체비셰프 부등식(Chebyshev's inequality) 정리는 '확률변수 X의 평균이 μ이고 분산이 $\sigma^2 (< \infty)$일 때, 모든 $k \geq 1$에 대하여 $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$이 성립한다'이다. 이산형 확률변수 X의 확률분포가 다음과 같을 때, 체비셰프 부등식 정리에 의한 $P(|X| \geq 1)$의 상한값은?

x: -1, 0, 1, 계
$P(X=x)$: $\frac{1}{8}$, $\frac{6}{8}$, $\frac{1}{8}$, 1

① $\frac{1}{4}$
② $\frac{1}{2}$
③ $\frac{3}{4}$
④ 1

정답 1번

출제 포인트

이 문제는 체비셰프 부등식에서 평균과 표준편차를 먼저 구한 뒤 $P(|X| \geq 1)$를 $P(|X-\mu| \geq k\sigma)$ 꼴로 바꾸어 상한을 구하는 문제이다.

조건 정리

주어진 값: $P(X=-1)=\frac{1}{8}$, $P(X=0)=\frac{6}{8}$, $P(X=1)=\frac{1}{8}$이다.
구해야 할 값: 체비셰프 부등식에 의한 $P(|X| \geq 1)$의 상한값이다.
주의할 조건: 먼저 μ와 σ를 구한 뒤 $|X| \geq 1$을 $|X-\mu| \geq k\sigma$ 형태로 맞춰야 한다.

사용할 공식

$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$
이 공식을 쓰는 이유: 문제에서 직접 체비셰프 부등식에 의한 상한값을 묻고 있으므로, 사건 $|X| \geq 1$을 평균과 표준편차 기준 사건으로 바꾸어 k를 찾으면 된다.

계산 과정

1. 평균은 $\mu = E(X) = (-1)\cdot\frac{1}{8} + 0\cdot\frac{6}{8} + 1\cdot\frac{1}{8} = 0$이다.
2. 분산은 $\sigma^2 = E(X^2) - \mu^2 = 1\cdot\frac{1}{8} + 0 + 1\cdot\frac{1}{8} = \frac{1}{4}$이므로 $\sigma = \frac{1}{2}$이다.
3. 따라서 $|X| \geq 1$은 $|X-0| \geq 1 = 2\cdot\frac{1}{2} = 2\sigma$와 같으므로 $k=2$이다.
4. 체비셰프 부등식을 적용하면 $P(|X| \geq 1) = P(|X-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = \frac{1}{4}$이다.

NoteBase

선지별 해설

① O: $k=2$이므로 상한은 $\frac{1}{k^2}=\frac{1}{4}$이다. 계산 결과와 정확히 일치한다.
② X: $\frac{1}{2}$는 $k=2$에서 체비세프 부등식을 $\frac{1}{k}$로 잘못 적용할 때 나오는 값이다. 상한값은 $\frac{1}{2}$가 아니라 $\frac{1}{4}$이다.
③ X: $\frac{3}{4}$는 $1-\frac{1}{4}$처럼 여사건과 혼동한 값이다. 체비세프 부등식은 $P(|X|\ge 1)$ 자체의 상한을 주므로 올바른 값은 $\frac{3}{4}$가 아니라 $\frac{1}{4}$이다.
④ X: 1은 확률의 자명한 상한일 뿐 체비세프 부등식을 적용한 결과가 아니다. 이 문항에서 체비세프 부등식으로 얻는 구체적 상한은 1이 아니라 $\frac{1}{4}$이다.

오답 포인트

분산 $\sigma^2=\frac{1}{4}$와 표준편차 $\sigma=\frac{1}{2}$를 혼동하면 k를 잘못 잡기 쉽다. 또한 체비세프 부등식의 우변이 $\frac{1}{k}$가 아니라 $\frac{1}{k^2}$라는 점, 그리고 여사건 확률 $1-\frac{1}{4}$와 뒤섞지 않는 점이 핵심 함정이다.

다시 나오면 이렇게 판단

체비세프 부등식 문제는 평균 μ와 표준편차 σ를 먼저 구하고, 묻는 사건을 $|X-\mu|\ge k\sigma$ 꼴로 바꾼 뒤 마지막에 $\frac{1}{k^2}$를 대입하는 순서로 판단한다.

암기 포인트

체비세프 상한은 사건을 $|X-\mu|\ge k\sigma$로 맞춘 뒤 $\frac{1}{k^2}$를 쓰며, 분산이 아니라 표준편차로 k를 잡는다.

NoteBase