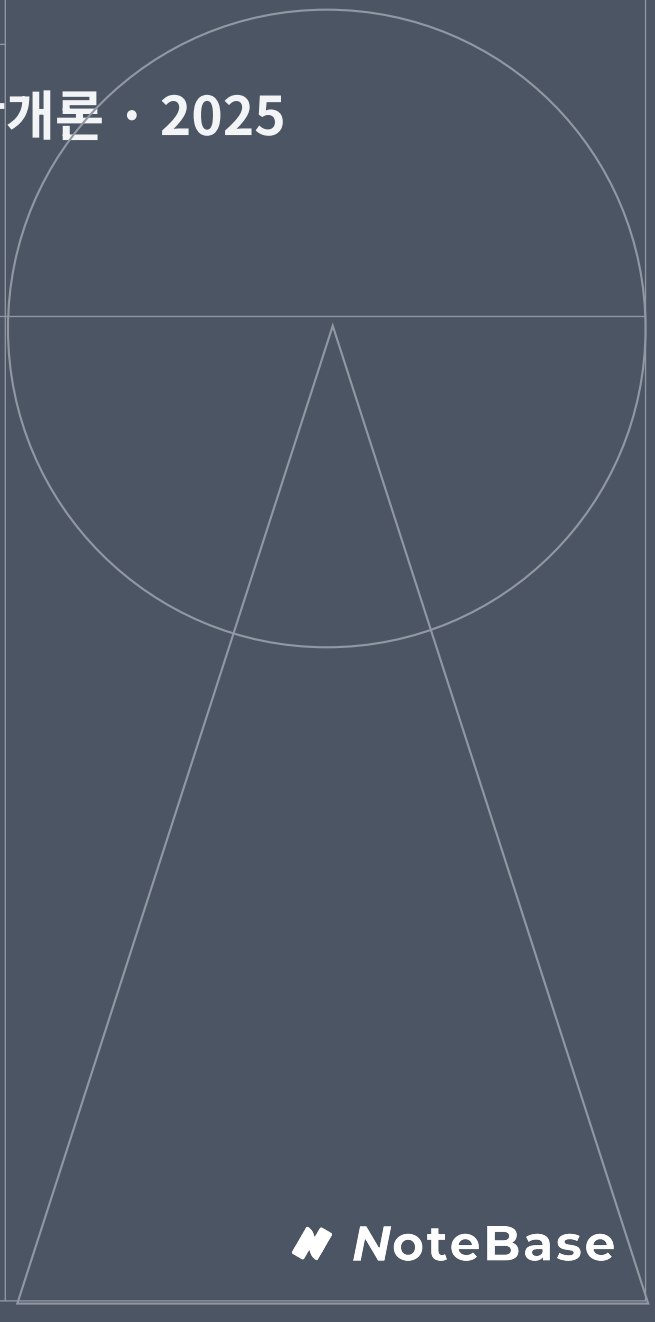
# 기출OX 해설집

공무원/국가직/9급 · 통계학개론 · 2025



NoteBase

# 기출OX 해설집

## 공무원/국가직/9급 · 통계학개론 · 2025

**공무원/국가직/9급 · 통계학개론 · 2025**

**문 1. 다음 설명 중 옳지 않은 것은?**

① 자료의 제40 백분위수는 중앙값보다 작거나 같다.

② 자료의 표준편차와 변동계수의 단위는 같다.

③ 자료 1, 2의 산술평균은 기하평균보다 크다.

④ 5% 절사평균은 전체 자료의 10%를 삭제한 자료의 평균이다.

**정답 2번**

**간단한 해설**

표준편차는 **원자료와 같은 단위**를 갖지만, 변동계수는 표준편차를 평균으로 나눈 비율이므로 **단위가 없다**. 변동계수를 백분율로 표시해도 원자료의 측정 단위를 갖는 것은 아니다.

**선지별 해설**

① O: 백분위수는 자료를 작은 값부터 정렬했을 때의 상대적 위치를 나타낸다. 중앙값은 **제50 백분위수**이며, 백분위의 순위가 커질수록 그 값은 작아지지 않는다. 따라서 제40 백분위수는 제50 백분위수인 중앙값보다 작거나 같다. 같은 값이 반복되는 자료에서는 두 값이 같을 수도 있다.

② X: 표준편차는 자료가 평균 주변에 흩어진 정도를 나타내며 원자료와 같은 단위를 갖는다. 평균이 0이 아닐 때 변동계수는 $\mathrm{CV} = \frac{s}{\bar{x}}$로 구한다. 여기서 $s$는 표준편차, $\bar{x}$는 산술평균이다. 분자와 분모의 단위가 같아 서로 약분되므로 변동계수는 **단위가 없는 상대적 산포 지표**이다. 예를 들어 원자료가 cm 단위이면 표준편차도 cm 단위이지만 변동계수에는 cm 단위가 붙지 않는다. 변동계수는 필요에 따라 100을 곱해 %로 표시한다.

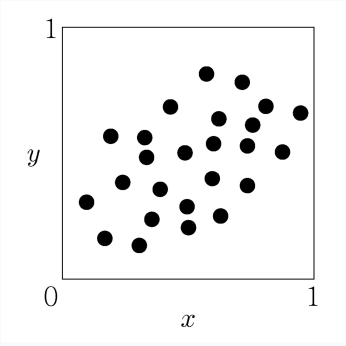
③ O: 산술평균은 자료의 합을 자료 수로 나눈 값이고, 두 양수의 기하평균은 두 값의 곱에 제곱근을 취한 값이다. 자료 1, 2의 산술평균은 $\frac{1+2}{2} = 1.5$, 기하평균은 $\sqrt{1 \cdot 2} = \sqrt{2} \approx 1.414$이므로 산술평균이 더 크다. 일반적으로 양수 자료에서는 산술평균이 기하평균보다 크거나 같으며, **모든 값이 같을 때만 두 평균이 같다**.

④ O: 5% 절사평균은 자료를 작은 값부터 정렬한 뒤 **하위 5%와 상위 5%를 각각 제외**하고 남은 자료의 산술평균을 구한 것이다. 양쪽에서 5%씩 제외하므로 삭제 비율은 전체 자료의 10%이다. 예를 들어 자료가 100개이면 가장 작은 5개와 가장 큰 5개를 빼고 남은 90개의 평균을 구한다.

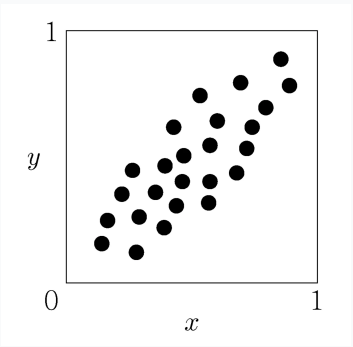
NoteBase

**문 2. 자료 $(x_i, y_i)$ $(i = 1, 2, \ldots, n)$에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$를 적합하여 분석하고자 한다. 다음 산점도 중 단순선형회귀모형의 결정계수가 가장 큰 것은?**

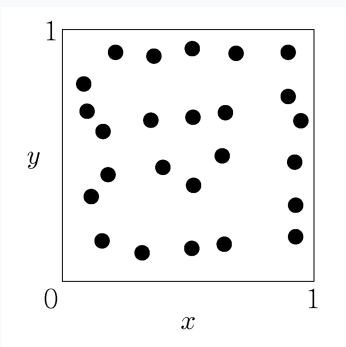
①



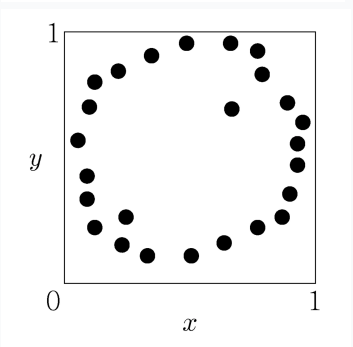

②




③




④




**정답 2번**

**간단한 해설**

②의 점들은 **우상향 직선 주위에 가장 뚜렷하게 모여** 있다. 결정계수는 직선으로 설명되는 변동의 비율이므로, 기울기의 크기보다 **직선적 관계의 강도**를 비교한다.

NoteBase

**선지별 해설**

결정계수 $R^2$는 관측된 y의 전체 변동 중 회귀직선으로 설명되는 변동의 비율이다. 절편을 포함한 단순선형회귀를 최소제곱법으로 적합하면 $R^2 = r^2$이며, r은 x와 y의 표본상관계수이다. 따라서 양의 관계인지 음의 관계인지보다 **상관계수의 절댓값이 얼마나 큰지**가 중요하다. 산점도에서는 점들이 하나의 직선 주위에 얼마나 밀집하는지를 비교한다.

①: x가 커질수록 y도 커지는 경향은 있지만 점들이 위아래로 넓게 흩어져 있다. ②보다 직선에서 벗어나는 정도가 커서, y의 전체 변동 중 직선으로 설명되는 비율이 작다.

②: 왼쪽 아래에서 오른쪽 위로 이어지는 **직선적 증가 경향이 가장 뚜렷하다**. 점들이 그 직선 주위에 상대적으로 좁게 모여 있으므로, 네 그림 중 상관계수의 절댓값과 결정계수가 가장 크다.

③: x의 여러 위치에서 y가 높은 점과 낮은 점이 함께 나타나며, 일정한 증가나 감소 경향이 뚜렷하지 않다. 직선으로 설명할 수 있는 변동이 적으므로 결정계수는 작다.

④: 점들이 원 둘레를 따라 배치되어 있지만, x가 증가할 때 y가 일관되게 증가하거나 감소하지 않는다. 원형이라는 규칙성이 있어도 **하나의 직선으로 잘 설명되는 관계는 아니므로** 단순선형회귀의 결정계수는 작다.

따라서 직선적 관계가 가장 강한 ②에서 결정계수가 가장 크다.

NoteBase

**문 3. 컴퓨터 부품의 품질은 두 인자 A, B에 따라 달라질 수 있다고 한다. 다음은 인자 A의 3가지 수준과 인자 B의 6가지 수준의 가능한 18가지 실험 조건에서 무작위 순서로 한 개씩 부품을 생산한 후, 부품의 품질을 수치화하여 얻은 분산분석표의 일부이다. 인자 A의 수준에 따라 부품의 품질에 차이가 있는지를 검정하기 위한 $F$-검정 통계량의 값은?**

| 요인 | 제곱합 | 자유도 | 평균제곱 | $F-$값 |
|---|---|---|---|---|
| 인자 A | 540 | | | |
| 인자 B | 400 | | | |
| 오차 | 100 | | | |
| 계 | 1,040 | | | |

① 4

② 8

③ 27

④ 54

**정답 3번**

**간단한 해설**

인자 A의 효과는 제곱합 자체가 아니라 **평균제곱을 오차평균제곱으로 나눈 값**으로 검정한다. 인자 A의 평균제곱은 270, 오차평균제곱은 10이므로 $F = \frac{270}{10} = 27$이다.

NoteBase

**계산 과정**

분석 조건: 인자 A는 3수준, 인자 B는 6수준이고 각 조합에서 한 번씩 관측했으므로 관측값은 총 18개이다. 주어진 표는 **두 인자의 교호작용이 없다고 가정하는** 반복 없는 이원분산분석표이다. 검정할 가설은 인자 A의 세 수준에서 부품 품질의 모평균이 모두 같다는 것이다.

인자 A의 평균제곱: 평균제곱은 제곱합을 해당 자유도로 나눈 값이다. 인자 A의 자유도는 수준 수에서 1을 뺀 3 − 1 = 2이고, 표의 인자 A 행에서 제곱합은 540이다. 따라서 인자 A의 평균제곱은 $MS_A = \frac{540}{2} = 270$이다.

오차평균제곱: 전체 자유도는 18 − 1 = 17, 인자 B의 자유도는 6 − 1 = 5이다. 전체 자유도에서 두 인자의 자유도를 빼면 오차 자유도는 17 − 2 − 5 = 10이다. 표의 오차 제곱합이 100이므로 $MS_E = \frac{100}{10} = 10$이다. 이 값은 두 인자의 효과로 설명되지 않는 **오차분산의 추정값**이다.

검정통계량: 인자 A의 효과가 없으면 인자 A의 평균제곱과 오차평균제곱은 같은 오차분산을 추정하므로, 둘의 비로 A의 효과를 검정한다. $F = \frac{MS_A}{MS_E} = \frac{270}{10} = 27$이다. 같은 단위의 평균제곱을 나누므로 F값에는 단위가 없으며, 계산값 27은 ③에 해당한다.

NoteBase

**문 4. 가설검정에 대한 설명으로 옳지 않은 것은?**

① 제2종 오류는 대립가설이 참일 때 귀무가설을 기각하지 못하는 오류이다.

② 검정력(power)은 대립가설이 참일 때 귀무가설을 기각하는 확률이다.

③ 유의확률(p-value)이 유의수준보다 작을 때 귀무가설을 기각한다.

④ 자료를 관찰한 후 귀무가설과 대립가설을 설정한다.

**정답 4번**

**간단한 해설**

④의 '자료를 관찰한 후'가 잘못되었다. 귀무가설과 대립가설은 **검정에 사용할 자료를 관찰하기 전에** 설정하여, 관찰 결과에 맞춰 가설을 정하는 일을 피해야 한다.

**선지별 해설**

① O: 제2종 오류는 **대립가설이 참인데도 귀무가설을 기각하지 못하는** 오류이다. 반대로 제1종 오류는 귀무가설이 참인데 이를 기각하는 오류이다.

② O: 검정력은 **대립가설이 참일 때 귀무가설을 기각하는 확률**이다. 대립가설이 참인 상황에서 기각하거나 기각하지 못하는 두 결과는 서로 보완적이므로, 제2종 오류 확률이 β이면 검정력은 $1-\beta$이다.

③ O: 유의확률은 귀무가설이 참이라는 전제 아래, 관측 결과 이상으로 극단적인 결과를 얻을 확률이다. 이 값이 미리 정한 유의수준보다 작으면 관측 결과를 귀무가설 아래에서 나타나기 어려운 것으로 보아 **귀무가설을 기각**한다.

④ X: 귀무가설과 대립가설은 **검정에 사용할 자료를 관찰하기 전에 설정**한다. 자료를 본 뒤 그 결과에 유리하도록 가설을 정하면 우연히 나타난 패턴까지 검정 대상으로 선택할 수 있어, 미리 정한 유의수준에 따른 오류 통제가 어려워진다.

NoteBase

**문 5. 어느 회사에서 남성 150명과 여성 150명을 임의추출한 후 세 종류의 신상품 A, B, C에 대한 선호도를 조사하여 다음과 같은 분할표를 얻었다. "성별에 따른 신상품 선호도에 차이가 없다."라는 귀무가설을 검정하기 위한 카이제곱검정통계량의 값과 유의수준 5%에서 검정한 결과를 바르게 연결한 것은? (단, $\chi^2_\alpha(k)$는 자유도가 $k$인 카이제곱분포의 제$100 \times (1-\alpha)$ 백분위수를 나타내고, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(6) = 12.59$이다)**

| 성별 \ 신상품 | A | B | C | 계 |
|---|---|---|---|---|
| 남성 | 60 | 40 | 50 | 150 |
| 여성 | 40 | 60 | 50 | 150 |
| 계 | 100 | 100 | 100 | 300 |

① 카이제곱검정통계량의 값: 8, 검정 결과: 귀무가설을 기각함

② 카이제곱검정통계량의 값: 8, 검정 결과: 귀무가설을 기각하지 못함

③ 카이제곱검정통계량의 값: 15, 검정 결과: 귀무가설을 기각함

④ 카이제곱검정통계량의 값: 15, 검정 결과: 귀무가설을 기각하지 못함

**정답 1번**

**간단한 해설**

귀무가설 아래에서는 여섯 칸의 **기대도수가 모두 50명**이므로 카이제곱검정통계량은 8이다. 자유도는 2이고, 8이 5% 임계값 5.99보다 크므로 **귀무가설을 기각**한다.

NoteBase

**계산 과정**

1. 검정할 가설과 자료: 귀무가설은 남성과 여성의 A·B·C 선호 비율이 같다는 것이고, 대립가설은 적어도 한 상품의 선호 비율이 다르다는 것이다. 표의 관측도수는 A·B·C 순서로 남성이 60명·40명·50명, 여성이 40명·60명·50명이다. 두 성별의 표본 크기는 각각 150명이고, 각 상품의 합계는 모두 100명으로 전체는 300명이다.

2. 기대도수: 귀무가설 아래의 공통 선호 비율은 전체 표본에서 각 상품이 차지하는 비율로 추정한다. 각 칸의 기대도수는 행 합계에 열 합계를 곱한 뒤 전체 합계로 나눈 값이므로 $E_{ij} = \frac{150 \cdot 100}{300} = 50$명이다. 여기서 i는 성별, j는 상품을 뜻한다. 즉, 선호 비율이 같다면 각 성별에서 상품별로 50명씩 관측될 것으로 기대한다. **모든 기대도수가 5 이상**이므로 카이제곱 근사를 적용하기에 충분하다.

3. 검정통계량: 관측도수를 $O_{ij}$, 기대도수를 $E_{ij}$라 하면 $\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$이다. 각 칸에서 관측도수와 기대도수의 차이를 제곱하고, 기대도수로 나누어 상대적인 차이를 합한다. A·B의 네 칸은 차이가 각각 10명 또는 −10명이므로 각 기여분은 $\frac{(\pm 10)^2}{50} = 2$이다. C의 두 칸은 관측도수와 기대도수가 같아 기여분이 0이다. 따라서 검정통계량은 $2 \times 4 + 0 \times 2 = 8$이며, 단위는 없다.

4. 자유도와 기각 여부: 성별이 2개, 상품이 3개이므로 자유도는 $(2-1) \times (3-1) = 2$이다. 행·열 합계가 주어지면 독립적으로 정할 수 있는 도수는 두 개이므로, 칸 수 6을 자유도로 쓰지 않는다. 이 검정은 관측도수와 기대도수의 차이가 클수록 귀무가설에 불리한 **오른쪽 꼬리 검정**이다. 자유도 2의 5% 임계값은 5.99이고 $8 > 5.99$이므로 귀무가설을 기각한다. 즉, 성별에 따른 신상품 선호도에 차이가 있다는 통계적 근거가 있다.

검정통계량 8과 귀무가설 기각을 연결한 ①에 해당한다.

NoteBase

**문 6. 세 자료 A, B, C의 변동계수를 각각 $a$, $b$, $c$라 할 때, 옳은 것은?**

◯ 자료 A: 10, 20, 30, 40, 50
◯ 자료 B: 5, 15, 25, 35, 45
◯ 자료 C: 5, 10, 15, 20, 25

① a=c이고 a<b

② a=c이고 b<a

③ a<b<c

④ c<b<a

**정답 1번**

**간단한 해설**

변동계수는 **평균에 대한 표준편차의 비율**이다. C는 A의 각 값을 절반으로 줄인 자료이므로 평균과 표준편차가 함께 절반이 되어 변동계수가 같다. B는 A의 각 값에서 5를 뺀 자료이므로 표준편차는 같지만 평균이 작아져 변동계수가 더 크다.

**공통 해설**

변동계수는 CV=표준편차평균로, 평균의 크기를 기준으로 자료가 얼마나 흩어져 있는지 나타낸다. 백분율로 표시할 때는 이 비율에 100을 곱하며, 대소 관계는 달라지지 않는다. 세 자료는 모두 평균이 양수이고 표준편차도 양수이므로, **분자인 표준편차와 분모인 평균이 어떻게 달라지는지**로 변동계수를 판단한다.

C의 각 값은 A의 각 값의 절반이다. 평균도 절반이 되고, 각 값과 평균의 차이도 절반이 되므로 표준편차 역시 절반이 된다. 따라서 표준편차를 평균으로 나눈 비율은 그대로여서 $a = c$이다.

B의 각 값은 A의 각 값에서 5를 뺀 것이다. 평균도 5만큼 작아지므로 각 값과 평균의 차이는 A와 같고, 표준편차는 변하지 않는다. 변동계수의 **분자는 같고 양수인 분모만 작아지므로** B의 변동계수가 더 크다. 따라서 $a < b$이며, 세 자료의 관계는 $a = c < b$이다.

NoteBase

**문 7. 두 확률변수 $X$와 $Y$의 결합확률분포가 다음 표와 같다. $X$의 기댓값이 0일 때, 옳은 것은?**

| $Y$ \ $X$ | $-1$ | $1$ |
|---|---|---|
| $-1$ | $\frac{3}{8}$ | ( ) |
| $0$ | $\frac{1}{8}$ | $\frac{1}{4}$ |
| $1$ | ( ) | $\frac{1}{8}$ |

① $Var(X) = \frac{1}{4}$

② $E(Y) = \frac{3}{8}$

③ $E(XY) = \frac{1}{4}$

④ $P(X + Y = 0) = \frac{1}{8}$

**정답 4번**

**간단한 해설**

$X + Y = 0$이 되는 경우는 **두 값이 −1과 1로 서로 반대인 경우**이다. 조건 $E(X) = 0$으로 표의 빈칸을 구하면 두 경우의 확률이 각각 0과 $\frac{1}{8}$이므로, 합이 0일 확률은 $\frac{1}{8}$이다.

NoteBase

**선지별 해설**

표에서 열은 $X$, 행은 $Y$의 값이다. $X$는 −1과 1만 취하므로 $E(X)=0$이면 **두 열의 확률 합이 각각 1/2**이다. 따라서 왼쪽 아래 빈칸은 $P(X=-1,Y=1)=\frac{1}{2}-\frac{3}{8}-\frac{1}{8}=0$, 오른쪽 위 빈칸은 $P(X=1,Y=-1)=\frac{1}{2}-\frac{1}{4}-\frac{1}{8}=\frac{1}{8}$이다.

① X: $X$가 −1이든 1이든 **제곱한 값은 항상 1**이므로 $E(X^2)=1$이다. 분산은 $Var(X)=E(X^2)-\{E(X)\}^2$ 이므로, 주어진 기댓값 0을 대입하면 분산은 1이다. 따라서 $\frac{1}{4}$이 아니다.

② X: 각 행의 확률을 더하면 $P(Y=-1)=\frac{1}{2}$, $P(Y=0)=\frac{3}{8}$, $P(Y=1)=\frac{1}{8}$이다. 기댓값은 각 값에 그 확률을 곱해 더하므로 $E(Y)=(-1)\cdot\frac{1}{2}+0\cdot\frac{3}{8}+1\cdot\frac{1}{8}=-\frac{3}{8}$이다. 선지는 **기댓값의 부호**가 잘못되었다.

③ X: $E(XY)$는 각 칸에서 **두 값의 곱에 결합확률을 곱한 뒤** 모두 더한다. $Y=0$인 행은 곱이 0이고, $(X,Y)=(-1,1)$인 칸은 확률이 0이다. 나머지 세 칸의 기여를 합하면 $E(XY)=\frac{3}{8}-\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$이다. 따라서 $\frac{1}{4}$이 아니다.

④ O: 합이 0인 경우는 $(X,Y)=(-1,1)$과 $(1,-1)$이다. 두 경우는 **동시에 일어날 수 없으므로 확률을 더한다**. 따라서 $P(X+Y=0)=0+\frac{1}{8}=\frac{1}{8}$이다.

NoteBase

**문 8. 확률변수 $X$에 대한 누적 확률 $P(X \le x)$가 다음과 같을 때, 옳지 않은 것은?**

P(X≤ x)= cases 0, & x<0\ (1)/(6)x, & 0≤ x<2\ (1)/(2), & 2≤ x<4\ (1)/(6)x, & 4≤ x<6\ 1, & x≥ 6 cases

① $P(X \le a) = \frac{1}{4}$을 만족시키는 $a$의 값은 $\frac{3}{2}$이다.

② $P(X = 3) = 0$

③ $P(X \ge 5) = \frac{2}{3}$

④ $P(1 < X \le 4) = \frac{1}{2}$

**정답 3번**

**간단한 해설**

③의 확률은 **5 이상인 오른쪽 꼬리확률**이다. 주어진 누적확률은 5에서 연속이므로 $P(X \ge 5) = 1 - P(X \le 5)$이며, 그 값은 $1 - \frac{5}{6} = \frac{1}{6}$이다.

**선지별 해설**

누적분포함수를 $F(x) = P(X \le x)$라 하자. $F(x)$는 $x$ 이하의 확률을 모두 합한 값이므로, 구간확률은 누적확률의 차로 구한다. 한 점의 확률은 그 점에서 누적분포함수가 뛰어오르는 크기이다.

① O: $\frac{1}{4}$은 $0 \le x < 2$ 구간의 누적확률 범위에 있으므로 $F(a) = \frac{a}{6}$을 이용한다. $\frac{a}{6} = \frac{1}{4}$을 풀면 $a = \frac{3}{2}$이고, 이 값은 해당 구간에 속한다.

② O: 3을 포함하는 $2 \le x < 4$ 구간에서 누적분포함수는 $\frac{1}{2}$로 일정하다. 따라서 **3에서 누적확률이 뛰어오르지 않으므로** $P(X = 3) = 0$이다.

③ X: $4 \le x < 6$에서 $F(x) = \frac{x}{6}$이므로 $F(5) = \frac{5}{6}$이다. 5에서는 누적분포함수가 연속이어서 $P(X = 5) = 0$이므로, **5 이상일 확률과 5를 초과할 확률이 같다**. 따라서 $P(X \ge 5) = 1 - F(5)$이고, $1 - \frac{5}{6} = \frac{1}{6}$이다. 선지의 $\frac{2}{3}$은 $\frac{1}{6}$로 고쳐야 한다.

④ O: $1 < X \le 4$는 **1을 제외하고 4를 포함하는 구간**이므로 $P(1 < X \le 4) = F(4) - F(1)$이다. 4에는 $4 \le x < 6$의 식을 적용하여 $F(4) = \frac{4}{6} = \frac{2}{3}$이고, $F(1) = \frac{1}{6}$이다. 따라서 구간확률은 $\frac{2}{3} - \frac{1}{6} = \frac{1}{2}$이다.

NoteBase

**문 9. 확률변수의 성질에 대한 설명으로 옳은 것만을 모두 고르면?**

ㄱ. 확률변수 $X$가 정규분포를 따를 때 $5X + 100$은 정규분포를 따른다.
ㄴ. 확률변수 $X$가 평균이 1이고 분산이 1인 정규분포를 따를 때 $(X-1)^2$은 자유도가 1인 카이제곱분포를 따른다.
ㄷ. 확률변수 $X$와 $Y$가 서로 독립이고 표준정규분포를 따를 때 $\frac{X}{Y}$는 분자와 분모의 자유도가 각각 1, 1인 $F$ 분포를 따른다.

① ㄱ
② ㄱ, ㄴ
③ ㄱ, ㄷ
④ ㄴ, ㄷ

**정답 2번**

**간단한 해설**

서로 독립인 두 표준정규변수의 비는 **자유도 1인 t분포**를 따른다. 분자와 분모의 자유도가 각각 1인 F분포는 그 비를 **제곱했을 때** 얻어진다.

**보기별 해설**

ㄱ. O: 정규분포를 따르는 확률변수에 **0이 아닌 상수를 곱하고 상수를 더해도** 정규분포를 따른다. $X \sim N(\mu, \sigma^2)$이면 $5X + 100 \sim N(5\mu + 100, 25\sigma^2)$이다. 100을 더하면 평균만 이동하고, 5를 곱하면 분산은 25배가 된다.

ㄴ. O: 평균이 1이고 분산이 1인 정규변수에서 평균 1을 빼면 $X - 1 \sim N(0, 1)$이다. **표준정규변수 하나의 제곱**은 자유도가 1인 카이제곱분포를 따르므로 $(X-1)^2 \sim \chi^2(1)$이다.

ㄷ. X: 서로 독립인 표준정규변수 $X$와 $Y$의 비 $\frac{X}{Y}$는 자유도가 1인 t분포, 즉 표준 코시분포를 따른다. F분포는 독립인 두 카이제곱변수를 각각의 자유도로 나눈 값들의 비로 정의된다. 여기서는 $X^2$과 $Y^2$이 서로 독립인 자유도 1의 카이제곱변수이므로 $\frac{X^2}{Y^2} \sim F(1, 1)$이다. 따라서 원문의 비를 **제곱한 값**이 F분포를 따른다. $\frac{X}{Y}$는 음수가 될 수 있지만 F분포를 따르는 변수는 음수가 될 수 없다는 점에서도 구별된다.

따라서 옳은 보기는 ㄱ, ㄴ이며, ②의 조합에 해당한다.

NoteBase

**문 10. 어떤 방송사의 저녁 뉴스 시청률을 알아보기 위해 400가구를 조사했다. 이 방송사 저녁 뉴스를 시청하는 비율 $p$에 대한 가설 $H_0 : p = 0.1$ 대 $H_1 : p > 0.1$에 대한 검정에서 $Z$-검정통계량의 값이 2일 때, 조사한 400가구 중 이 방송사 저녁 뉴스를 시청한 가구 수와 유의확률을 바르게 연결한 것은? (단, 확률변수 $Z$가 표준정규분포를 따를 때 $P(Z \le 2) = 0.9772$이다)**

① 시청한 가구 수: 50, 유의확률: 0.0228

② 시청한 가구 수: 50, 유의확률: 0.0456

③ 시청한 가구 수: 52, 유의확률: 0.0228

④ 시청한 가구 수: 52, 유의확률: 0.0456

**정답 3번**

### 간단한 해설

검정통계량 2는 표본비율이 가정한 비율 0.1보다 표준오차 0.015의 두 배만큼 크다는 뜻이므로, 시청 가구 수는 **52가구**이다. 대립가설이 $p > 0.1$인 **우측 단측검정**이므로 유의확률은 1 − 0.9772 = 0.0228이다.

### 계산 과정

표준오차: 조사 가구 수는 n = 400이고, 귀무가설에서 가정한 시청 비율은 $p_0 = 0.1$이다. 모비율의 Z-검정에서는 **귀무가설의 비율로 표준오차를 계산**한다. 따라서 표본비율의 표준오차는 $\sqrt{\frac{0.1 \cdot 0.9}{400}} = 0.015$이다. 귀무가설 아래 시청 가구 수의 기댓값은 40, 미시청 가구 수의 기댓값은 360으로 모두 충분히 커 정규근사를 적용할 수 있다.

시청 가구 수: 표본비율을 $\hat{p}$라 하면, 검정통계량은 $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$이다. 주어진 값 Z = 2를 대입하면 $2 = \frac{\hat{p} - 0.1}{0.015}$이므로 $\hat{p} = 0.1 + 2 \cdot 0.015 = 0.13$이다. 이는 조사 가구의 13%가 시청했다는 뜻이므로, 시청 가구 수는 400 × 0.13 = 52가구이다.

유의확률: 대립가설이 $H_1 : p > 0.1$이므로 표본비율이 클수록 대립가설을 지지한다. 따라서 귀무가설 아래에서 관측값 2 이상으로 큰 검정통계량이 나올 **오른쪽 꼬리확률**을 구한다. 주어진 표준정규분포 확률을 이용하면 P(Z≥2)=1-P(Z≤2)이므로 유의확률은 1 − 0.9772 = 0.0228이다. 양쪽 꼬리확률을 더하지 않는다.

시청 가구 수 52가구와 유의확률 0.0228을 연결한 것은 ③이다.

NoteBase

**문 11. 다음은 자료 $(x_i, y_i)$ $(i = 1, 2, \ldots, n)$에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$를 최소제곱법으로 적합하여 얻은 결과이다. 이 단순선형회귀모형의 잔차제곱합(SSE)은? (단, $\hat{\beta}_1$은 $\beta_1$의 최소제곱추정량이다)**

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 30, \quad \sum_{i=1}^{n}(y_i - \bar{y})^2 = 360, \quad \hat{\beta}_1 = 2$$

① 90

② 120

③ 240

④ 270

**정답 3번**

**간단한 해설**

잔차제곱합은 **전체제곱합에서 회귀제곱합을 뺀 값**이다. 회귀제곱합이 $2^2 \cdot 30 = 120$이므로 잔차제곱합은 360 − 120 = 240이다.

**계산 과정**

전체제곱합 SST는 관측값이 평균에서 벗어난 정도를 제곱하여 더한 값으로, $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 360$이다. **절편을 포함한 최소제곱 회귀**에서는 $SST = SSR + SSE$가 성립한다. 여기서 SSR은 회귀식이 설명한 변동인 회귀제곱합이고, SSE는 관측값과 적합값의 차이인 잔차를 제곱하여 더한 값이다.

적합값을 $\hat{y}_i$라 하면 회귀직선이 $(\bar{x}, \bar{y})$를 지나므로 $\hat{y}_i - \bar{y} = \hat{\beta}_1(x_i - \bar{x})$이다. 이를 제곱하여 더하면 $SSR = \hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$가 된다. 따라서 주어진 기울기 추정값 2와 편차제곱합 30을 대입하면 $SSR = 2^2 \cdot 30 = 120$이다. 이 120은 전체 변동 중 **회귀식으로 설명되는 부분**이다.

잔차제곱합은 설명되지 않고 남은 변동이므로 SSE = SST − SSR = 360 − 120 = 240이다. 이는 ③의 240에 해당한다.

NoteBase

**문 12. 자료 $(x_i, y_i)$ $(i = 1, 2, \ldots, 20)$에 절편이 없는 단순선형회귀모형 $y_i = \beta x_i + \varepsilon_i$를 적합하려고 한다. $\beta$에 대한 최소제곱추정량은? (단, $x_i$는 서로 다르다)**

① (Σ _i=1^20x_i y_i)/(Σ _i=1^20x_i^2)

② Σ _i=1^20x_i y_i Σ _i=1^20(x_i-x)^2

③ Σ _i=1^20(x_i-x)(y_i-y) Σ _i=1^20x_i^2

④ Σ _i=1^20(x_i-x)(y_i-y) Σ _i=1^20(x_i-x)^2

**정답 1번**

**간단한 해설**

절편이 없는 회귀에서는 직선이 **원점을 지나도록** 고정되므로 자료에서 표본평균을 빼지 않는다. 잔차제곱합을 최소화하면 $\hat{\beta} = \frac{\sum_{i=1}^{20} x_i y_i}{\sum_{i=1}^{20} x_i^2}$를 얻는다.

**핵심 풀이**

절편이 없는 모형의 적합값은 $\hat{y}_i = b x_i$이다. 최소제곱법은 관측값과 적합값의 차이인 잔차의 제곱합 $S(b) = \sum_{i=1}^{20} (y_i - b x_i)^2$을 가장 작게 하는 기울기를 찾는다. 여기서는 **절편을 0으로 고정**하므로 기울기만 추정한다.

잔차제곱합을 기울기 $b$로 미분하여 0으로 놓으면 $S'(b) = -2 \sum_{i=1}^{20} x_i (y_i - b x_i) = 0$이다. 이를 정리하면 $b \sum_{i=1}^{20} x_i^2 = \sum_{i=1}^{20} x_i y_i$가 된다. 주어진 $x_i$가 서로 다르므로 모두 0일 수 없어 $\sum_{i=1}^{20} x_i^2 > 0$이다. 따라서 $S(b)$는 이차항의 계수가 양수인 이차함수이며, 이 식을 만족하는 기울기에서 유일한 최솟값을 갖는다.

따라서 최소제곱추정량은 $\hat{\beta} = \frac{\sum_{i=1}^{20} x_i y_i}{\sum_{i=1}^{20} x_i^2}$이며, ①의 식에 해당한다.

④처럼 설명변수와 반응변수에서 각각 표본평균을 뺀 식은 **절편도 함께 추정하는 모형**의 기울기 공식이다. 이 문제는 원점을 지나는 직선을 적합하므로 그 공식을 그대로 적용하지 않는다. ②와 ③처럼 분자와 분모 중 한쪽에만 평균을 뺀 항을 사용하는 식도 위 최소제곱 조건에서 도출되는 일반식이 아니다.

NoteBase

**문 13. 다음은 자료 $(x_i, y_i)$ $(i = 1, 2, \ldots, 12)$에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$를 최소제곱법으로 적합하여 얻은 분산분석표의 일부이다. 이 단순선형회귀모형에서 잔차제곱합의 자유도와 결정계수의 합은? (단, $\varepsilon_i$는 정규분포 $N(0, \sigma^2)$을 따르고 서로 독립이다)**

| 요인 | 제곱합 | 자유도 | 평균제곱 | $F-$값 |
|---|---|---|---|---|
| 회귀 | | | | 10 |
| 잔차 | | | 13 | |
| 계 | | | | |

① $\frac{17}{2}$
② $\frac{19}{2}$
③ $\frac{21}{2}$
④ $\frac{23}{2}$

**정답 3번**

**간단한 해설**

절편과 기울기를 추정하므로 **잔차 자유도는 12 − 2 = 10**이다. 표의 F값과 잔차 평균제곱으로 구한 회귀제곱합과 잔차제곱합이 같아 **결정계수는 0.5**이며, 두 값의 합은 $\frac{21}{2}$이다.

NoteBase

**계산 과정**

잔차 자유도: 관측값은 12개이고, 단순선형회귀모형에서 절편 $\beta_0$와 기울기 $\beta_1$ 두 모수를 추정한다. 따라서 잔차제곱합의 자유도는 $n-2$, 즉 $12-2=10$이다. 설명변수가 하나이므로 회귀제곱합의 자유도는 1이다.

회귀제곱합과 잔차제곱합: 평균제곱은 해당 제곱합을 자유도로 나눈 값이다. 표에서 잔차 평균제곱 MSE는 13이므로 잔차제곱합 SSE는 $13 \times 10 = 130$이다. F값은 **회귀 평균제곱을 잔차 평균제곱으로 나눈 값**이므로 $F = \frac{\text{MSR}}{\text{MSE}}$이다. F = 10에서 회귀 평균제곱 MSR는 $10 \times 13 = 130$이고, 회귀 자유도가 1이므로 회귀제곱합 SSR는 $130 \times 1 = 130$이다.

결정계수: 절편을 포함한 최소제곱 회귀에서는 총제곱합 SST가 회귀제곱합과 잔차제곱합의 합이므로 $130 + 130 = 260$이다. 결정계수는 **전체 변동 중 회귀모형이 설명하는 비율**이므로 $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{130}{260} = \frac{1}{2}$이다. 즉, 반응변수의 전체 변동 중 50%를 모형이 설명한다.

요구한 합: 잔차제곱합의 자유도와 결정계수를 더하면 $10 + \frac{1}{2} = \frac{21}{2}$이며, ③의 값과 일치한다.

NoteBase

**문 14. 주사위 한 개를 던져서 나온 눈의 수를 $N$이라고 하자. 1부터 $N$까지의 자연수가 하나씩 적힌 공 $N$개를 주머니에 넣은 후 임의로 한 개를 뽑을 때, 뽑힌 공에 적힌 숫자가 4일 확률은? (단, 각 공에는 숫자가 하나만 적혀 있다)**

① $\frac{31}{360}$
② $\frac{37}{360}$
③ $\frac{43}{360}$
④ $\frac{49}{360}$

**정답 2번**

### 간단한 해설

숫자 4가 적힌 공을 뽑으려면 주사위 눈이 **4·5·6 중 하나**여야 한다. 각 경우에 4를 뽑을 확률에 그 눈이 나올 확률을 곱해 더하면 $\frac{37}{360}$이다.

### 계산 과정

가능한 경우: 뽑힌 공에 적힌 숫자가 4인 사건을 $A$라 하자. 주사위 눈 $N$이 $1\cdot2\cdot3$이면 주머니에 4가 적힌 공이 없으므로 사건 $A$는 일어날 수 없다. 따라서 **주사위 눈이 $4\cdot5\cdot6$인 경우만** 계산한다.

각 경우의 확률: $N=n$으로 정해지면 주머니에는 1부터 $n$까지 적힌 공이 하나씩 있고, 각 공을 뽑을 가능성은 같다. $n$이 $4\cdot5\cdot6$이면 4가 적힌 공은 각각 한 개이므로 $P(A\mid N=n)=\frac{1}{n}$이다. 즉, 4를 뽑을 확률은 주사위 눈이 4일 때 $\frac{1}{4}$, 5일 때 $\frac{1}{5}$, 6일 때 $\frac{1}{6}$이다.

전체 확률: 주사위의 각 눈이 나올 확률은 $\frac{1}{6}$이다. 각 경우가 일어나고 이어서 4가 적힌 공을 뽑을 확률은 **그 눈이 나올 확률과 그 조건에서 4를 뽑을 확률의 곱**이다. 주사위 눈이 $4\cdot5\cdot6$인 경우는 서로 겹치지 않으므로 이 확률들을 더한다. 따라서 $P(A)=\frac{1}{6}\left(\frac{1}{4}+\frac{1}{5}+\frac{1}{6}\right)$이다. 괄호 안을 통분하면 $\frac{15+12+10}{60}=\frac{37}{60}$이므로, $P(A)=\frac{1}{6}\cdot\frac{37}{60}=\frac{37}{360}$이다.

구하는 확률은 $\frac{37}{360}$이며, ②의 값과 일치한다.

NoteBase

**문 15. 어느 도시 직장인 1,500명을 임의추출하여 성별과 출근 교통수단은 해당 항목을 선택하게 하고, 출근 거리와 소요 시간은 직접 작성하도록 하는 다음과 같은 설문조사를 실시했다.**
**이에 대한 설명으로 옳지 않은 것은?**

성별: (가) 남성 (나) 여성
교통수단: (가) 대중교통 (나) 자가용 (다) 기타
출근 거리: ( ) km
소요 시간: ( ) 분

① 교통수단에 따라 소요 시간에 차이가 있는지를 일표본(one sample) $t$-검정으로 확인할 수 있다.

② 성별과 교통수단이 관련이 있는지를 독립성검정으로 확인할 수 있다.

③ 출근 거리와 소요 시간 사이에 선형관계가 있는지를 상관분석으로 확인할 수 있다.

④ 교통수단이 대중교통, 자가용, 기타인 비율이 6:3:1인지를 적합도검정으로 확인할 수 있다.

**정답 1번**

**간단한 해설**

일표본 t-검정은 **한 모집단의 평균을 특정 기준값과 비교**하는 방법이다. 대중교통·자가용·기타의 세 집단 간 평균 소요 시간을 비교하려면 **일원분산분석**을 사용한다.

NoteBase

**선지별 해설**

① X: 일표본 t-검정은 한 모집단의 평균이 미리 정한 기준값과 같은지를 검정하므로, 교통수단별 집단의 평균을 비교하는 데 적합하지 않다. 이 설문은 교통수단이 대중교통·자가용·기타의 세 범주이므로, 집단별 관측값의 독립성·정규성·등분산성 등 기본 가정 아래 **일원분산분석으로 세 집단의 평균 소요 시간**을 비교한다. 이때 귀무가설은 세 집단의 모평균이 모두 같다는 것이다.

② O: 성별과 교통수단은 모두 범주형 변수이다. 성별 2개 범주와 교통수단 3개 범주를 교차한 분할표를 만들고, 카이제곱 독립성검정으로 **두 범주형 변수의 관련성**을 확인할 수 있다. 성별과 교통수단이 독립이라는 귀무가설 아래의 기대도수와 실제 관측도수를 비교한다.

③ O: 출근 거리와 소요 시간은 모두 수치로 측정하는 양적 변수이다. 피어슨 상관분석은 두 변수 사이의 **선형관계의 방향과 강도**를 나타내며, 모집단 상관계수가 0인지 검정하여 선형관계가 있는지를 판단할 수 있다. 여기서는 거리와 시간이 한 사람에게서 함께 측정된 값의 쌍으로 분석된다.

④ O: 교통수단이라는 하나의 범주형 변수에서 대중교통·자가용·기타의 모집단 비율이 각각 0.6, 0.3, 0.1인지를 확인하는 문제이다. 카이제곱 적합도검정은 **관측된 범주별 도수가 지정한 비율에 부합하는지**를 검정한다. 표본 1,500명에 그 비율을 적용한 기대도수는 각각 900명, 450명, 150명이며, 이를 실제 관측도수와 비교한다.

NoteBase

**문 16. 다음은 반복 수가 같은 일원배치법으로부터 얻은 분산분석표의 일부이다. 처리 효과가 없다는 귀무가설을 유의수준 5%에서 검정하는 방법으로 옳은 것은? (단, $F_\alpha(k_1, k_2)$는 분자와 분모의 자유도가 각각 $k_1, k_2$인 $F$분포의 제$100 \times (1-\alpha)$ 백분위수를 나타내고, $F_{0.95}(3, 20) = 0.12$, $F_{0.95}(20, 3) = 0.32$, $F_{0.05}(20, 3) = 8.66$이다)**

| 요인 | 제곱합 | 자유도 | 평균제곱 | $F-$값 |
|---|---|---|---|---|
| 처리 | | 3 | 40 | $F_0$ |
| 오차 | 1,000 | 20 | | |
| 계 | | | | |

① $F_0 = \frac{5}{4} > 0.12$이므로 귀무가설을 기각함

② $F_0 = \frac{5}{4} < 8.66$이므로 귀무가설을 기각하지 못함

③ $F_0 = \frac{4}{5} > \frac{1}{8.66}$이므로 귀무가설을 기각함

④ $F_0 = \frac{4}{5} < \frac{1}{0.32}$이므로 귀무가설을 기각하지 못함

**정답 4번**

**간단한 해설**

분산분석의 검정통계량은 **처리 평균제곱을 오차 평균제곱으로 나눈 값**이므로 $F_0 = \frac{4}{5}$이다. 이 값은 상위 5% 임계값 $\frac{1}{0.32} = 3.125$보다 작으므로 귀무가설을 기각하지 못한다.

NoteBase

**계산 과정**

검정 원리: 귀무가설은 '처리 효과가 없다', 즉 모든 처리의 모평균이 같다는 것이다. 오차의 독립성·정규성·등분산성을 전제로 하는 일원배치 분산분석에서는 처리 평균제곱을 오차 평균제곱과 비교한다. 처리 효과가 있으면 이 비가 커지는 경향이 있으므로 **오른쪽 꼬리에서 기각**한다.

오차 평균제곱: 표에서 오차 제곱합은 1,000, 오차 자유도는 20이다. 평균제곱은 제곱합을 해당 자유도로 나눈 값이므로 $MS_E = \frac{1000}{20} = 50$이다. 이는 처리 내의 오차 변동을 나타내며, 검정통계량의 분모가 된다.

검정통계량: 표의 처리 평균제곱은 이미 40으로 주어져 있다. 따라서 $F_0 = \frac{MS_T}{MS_E} = \frac{40}{50} = \frac{4}{5}$이다. 두 평균제곱의 단위가 같으므로 이 비에는 단위가 없다. 귀무가설 아래에서 분자의 처리 자유도 3과 분모의 오차 자유도 20을 사용하므로, 기준 분포는 $F(3, 20)$이다.

임계값: 문제의 $F_\alpha$는 오른쪽 꼬리 확률이 $\alpha$인 값이다. 유의수준 5%에서는 $F_{0.05}(3, 20)$을 써야 한다. F분포의 역수는 **분자와 분모의 자유도가 바뀐 F분포**를 따르므로 $F_{0.05}(3, 20) = \frac{1}{F_{0.95}(20, 3)}$이다. 주어진 값 0.32를 대입하면 임계값은 $\frac{1}{0.32} = 3.125$이다.

검정 결론: $F_0 = \frac{4}{5} < \frac{1}{0.32}$이므로 기각역에 들어가지 않아 귀무가설을 기각하지 못한다. 이 계산과 판단이 ④에 해당한다. 이는 처리 효과가 없음을 확정한 것이 아니라, 처리 효과가 있다고 판단할 근거가 충분하지 않다는 뜻이다.

NoteBase

**문 17. 모평균과 모분산 $\sigma^2$이 알려지지 않은 정규모집단에서 추출한 13개의 확률표본 $X_1, X_2, \ldots, X_{13}$의 표본분산을 $S^2$이라 하자. 가설 $H_0 : \sigma^2 = 4$ 대 $H_1 : \sigma^2 > 4$에 대한 검정에서 기각역으로 $S^2 > 6$을 사용하고자 한다. 확률변수 $V$는 자유도가 12인 카이제곱분포를 따를 때, 이 검정에서 제1종 오류의 확률과 같은 것은?**

① $P(V > 6)$

② $P(V > 12)$

③ $P(V > 18)$

④ $P(V > 24)$

**정답 3번**

### 간단한 해설

제1종 오류는 **귀무가설이 참인데도 기각하는 것**이다. 모분산이 4일 때 기각조건 $S^2 > 6$을 자유도 12인 카이제곱변수로 바꾸면 $V > 18$이므로, 그 확률은 $P(V > 18)$이다.

### 계산 과정

제1종 오류는 귀무가설이 참일 때 이를 기각하는 오류이다. 따라서 구할 확률은 **모분산이 4인 조건에서 기각역에 들어갈 확률**, 즉 $\alpha = P(S^2 > 6 \mid \sigma^2 = 4)$이다.

정규모집단의 크기 $n = 13$인 확률표본에서 표본분산은 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$이다. 이때 $\frac{(n-1)S^2}{\sigma^2}$은 자유도 $n - 1$인 카이제곱분포를 따른다. 모평균을 모르므로 표본평균을 이용하며, 표본평균으로부터의 편차 합이 0이라는 제약 때문에 **자유도는 13이 아니라 12**이다.

귀무가설 아래에서 $V = \frac{12S^2}{4} = 3S^2$로 둘 수 있다. 따라서 기각조건 $S^2 > 6$의 양변에 3을 곱하면 $V > 18$이 된다. 18은 표본분산의 경계값 6을 카이제곱변수의 척도로 바꾼 값이다.

그러므로 제1종 오류의 확률은 $\alpha = P(V > 18)$이며, 이는 ③의 표현과 일치한다.

NoteBase

**문 18. 어떤 회사에서 생산하는 건전지의 평균 수명은 300일이라고 알려져 있다. 이를 확인하기 위하여 건전지 15개를 임의추출하여 구한 평균 수명(단위: 일)에 대한 95% 신뢰구간은 (296, 308)이었다. 귀무가설 "건전지의 평균 수명은 300일이다."와 대립가설 "건전지의 평균 수명은 300일이 아니다."에 대한 검정에서 $t$-검정통계량의 값과 같은 것은? (단, 건전지의 수명은 정규분포를 따른다고 하며, $t_\alpha(k)$는 자유도가 $k$인 $t$ 분포의 제$100 \times (1-\alpha)$ 백분위수를 나타낸다.)**

① $\frac{1}{3} \times t_{0.025}(14)$

② $\frac{2}{3} \times t_{0.025}(14)$

③ $t_{0.025}(14)$

④ $\frac{4}{3} \times t_{0.025}(14)$

**정답 1번**

**간단한 해설**

신뢰구간의 **중심은 302일, 오차한계는 6일**이다. 표본평균과 귀무가설의 평균 300일의 차이가 오차한계의 1/3이므로, 검정통계량은 $\frac{1}{3}t_{0.025}(14)$이다.

**계산 과정**

1. 적용할 분포와 공식: 정규모집단에서 15개를 임의추출하고 모분산을 모르는 경우, 표본표준편차를 이용하는 평균의 신뢰구간과 검정에는 **자유도 15 − 1 = 14인 t분포**를 사용한다. 표본평균을 $\bar{x}$, 표본표준편차를 $s$, 추정 표준오차를 $e = \frac{s}{\sqrt{15}}$라 하자. 95% 양측 신뢰구간은 양쪽 꼬리에 각각 0.025를 남기므로, 임계값을 $c = t_{0.025}(14)$라 두면 신뢰구간은 $\bar{x} \pm ce$이다.

2. 표본평균과 오차한계: 신뢰구간이 표본평균을 중심으로 대칭이므로 $\bar{x} = \frac{296+308}{2} = 302$일이다. 구간의 반폭인 오차한계는 $\frac{308-296}{2} = 6$일이다.

3. 표준오차 구하기: 오차한계 6일은 **임계값과 표준오차를 곱한 값**이다. 따라서 $ce = 6$이고, 추정 표준오차는 $e = \frac{6}{c}$일이다. 6일 자체를 표준오차로 사용해서는 안 된다.

4. 검정통계량 계산: 귀무가설의 모평균은 $\mu_0 = 300$일이다. t검정통계량은 표본평균과 가설값의 차이를 추정 표준오차로 나눈 $t = \frac{\bar{x}-\mu_0}{e}$이다. 평균의 차이는 302 − 300 = 2일이므로 $t = \frac{2}{6/c} = \frac{c}{3}$이다. 분자와 분모의 단위가 모두 일이므로 검정통계량에는 단위가 없으며, 표본평균이 300일보다 커서 부호는 양수이다.

따라서 검정통계량은 $\frac{1}{3}t_{0.025}(14)$이며, ①의 식과 일치한다.

NoteBase

**문 19. 정규분포 $N(\mu, \sigma^2)$를 따르는 모집단에서 추출한 25개의 확률표본 $X_1, \ldots, X_{25}$의 표본평균을 $\overline{X}$라고 하자. $X_1$이 다음 조건을 만족한다. 다음 중 옳은 것만을 모두 고르면?**

$P(X_1 \le 4) + P(X_1 \le 6) = 1, \quad E(X_1^2) = 34$

ㄱ. $P(\overline{X} \le 4) = P(\overline{X} \ge 6)$

ㄴ. $P(\overline{X} \ge 6) = P(X_1 \ge 10)$

ㄷ. $P(X_1 \ge 8) + P(\overline{X} \ge 4.4) = 1$

① ㄱ, ㄴ

② ㄱ, ㄷ

③ ㄴ, ㄷ

④ ㄱ, ㄴ, ㄷ

**정답 4번**

**간단한 해설**

조건에서 모평균은 5이고, 원래 변수의 표준편차는 3, 표본평균의 표준편차는 $\frac{3}{5}$이다. ㄷ의 8과 4.4를 각 분포에서 표준화하면 1과 −1이므로, **대칭성에 의해 두 오른쪽 꼬리확률의 합은 1**이다.

**보기별 해설**

정규분포의 대칭성에 의해 $P(X_1 \le 4) + P(X_1 \le 6) = 1$이면 4와 6은 평균을 중심으로 대칭인 지점이다. 따라서 $\mu = \frac{4+6}{2} = 5$이다. 분산은 $\sigma^2 = E(X_1^2) - \{E(X_1)\}^2$이므로 $\sigma^2 = 34 - 25 = 9$이다. 25개의 확률표본은 서로 독립이고 같은 정규분포를 따르므로, 표본평균도 정규분포를 따르며 평균은 5, 분산은 $\frac{9}{25}$이다. 즉, $X_1$의 표준편차는 3이고 $\overline{X}$의 표준편차는 $\frac{3}{5}$이다.

ㄱ. O: $\overline{X}$는 평균 5를 중심으로 대칭이다. 4와 6은 평균에서 각각 1만큼 떨어져 있으므로 **왼쪽 꼬리와 오른쪽 꼬리의 면적이 같다**. 따라서 $P(\overline{X} \le 4) = P(\overline{X} \ge 6)$이다.

ㄴ. O: $Z \sim N(0, 1)$이라 하자. $\overline{X} = 6$의 표준화 값은 $\frac{6-5}{3/5} = \frac{5}{3}$이고, $X_1 = 10$의 표준화 값도 $\frac{10-5}{3} = \frac{5}{3}$이다. 두 확률 모두 **같은 표준화 값 이상일 확률**이므로 $P(\overline{X} \ge 6) = P(Z \ge \frac{5}{3})$이고, 이는 $P(X_1 \ge 10)$과 같다.

ㄷ. O: $X_1 = 8$의 표준화 값은 $\frac{8-5}{3} = 1$이고, $\overline{X} = 4.4$의 표준화 값은 $\frac{4.4-5}{3/5} = -1$이다. 따라서 두 확률은 각각 $P(Z \ge 1)$과 $P(Z \ge -1)$이다. 표준정규분포의 대칭성과 연속성에 의해 $P(Z \ge -1) = P(Z \le 1)$이며, **한 점의 확률은 0**이므로 $P(Z \ge 1) + P(Z \le 1) = 1$이다. 따라서 주어진 두 확률의 합은 1이다.

ㄱ, ㄴ, ㄷ이 모두 성립하므로 세 보기를 모두 포함한 ④와 일치한다.

NoteBase

**문 20. 어느 극장에서 월요일부터 금요일까지 관람객 수의 비율이 1 : 1 : 1 : 1 : 2인지를 검정하고자 평일 관람객 1,200명을 조사하여 다음 표와 같이 정리했다. 카이제곱검정에서 $a$의 값이 0부터 600까지 증가할 때, 유의확률에 대한 설명으로 옳은 것은?**

| 요일 | 월 | 화 | 수 | 목 | 금 | 계 |
|---|---|---|---|---|---|---|
| 관람객 수 | 170 | 210 | 220 | $a$ | $600-a$ | 1,200 |

① 유의확률은 증가하다가 감소한다.

② 유의확률은 감소하다가 증가한다.

③ 유의확률은 지속적으로 증가한다.

④ 유의확률은 지속적으로 감소한다.

**정답 1번**

**간단한 해설**

$a = 200$일 때 목요일과 금요일의 관측도수가 기대도수와 일치하여 **카이제곱 검정통계량이 최소**가 된다.
유의확률은 검정통계량의 오른쪽 꼬리확률이므로, a가 0에서 600까지 커지면 **증가하다가 감소한다**.

**공통 해설**

카이제곱 적합도검정은 관측도수가 가정한 비율에 따른 기대도수에서 얼마나 벗어나는지를 판단한다.
관측도수를 $O_i$, 기대도수를 $E_i$라 하면 검정통계량은 $\chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i}$이다. **기대도수와의 차이를 제곱**하므로, 어느 방향으로 벗어나든 차이가 커지면 검정통계량이 커진다.

관람객 1,200명을 비율 1 : 1 : 1 : 1 : 2에 따라 나누면 월요일부터 목요일까지의 기대도수는 각각 200명이고, 금요일은 400명이다. 표에서 월·화·수의 관측도수 170명, 210명, 220명은 고정되어 있다. 목요일과 금요일의 관측도수는 각각 a명, $600 - a$명이므로, 기대도수와의 차이는 각각 $a - 200$, $200 - a$이다. 따라서 a가 0에서 200까지 증가하면 두 차이의 절댓값이 함께 줄어 검정통계량이 감소하고, 200에서 600까지 증가하면 두 절댓값이 함께 커져 검정통계량이 증가한다.

가정한 비율을 자료에서 추정하지 않았으므로 자유도는 $5 - 1 = 4$이다. 유의확률은 귀무가설 아래 자유도 4인 카이제곱분포에서 **관측된 검정통계량 이상인 영역의 확률**이다. 검정통계량이 작아지면 이 오른쪽 꼬리확률은 커지고, 검정통계량이 커지면 작아진다. 그러므로 유의확률은 a가 0에서 200까지 증가할 때 커지고, 200에서 600까지 증가할 때 작아진다.

NoteBase