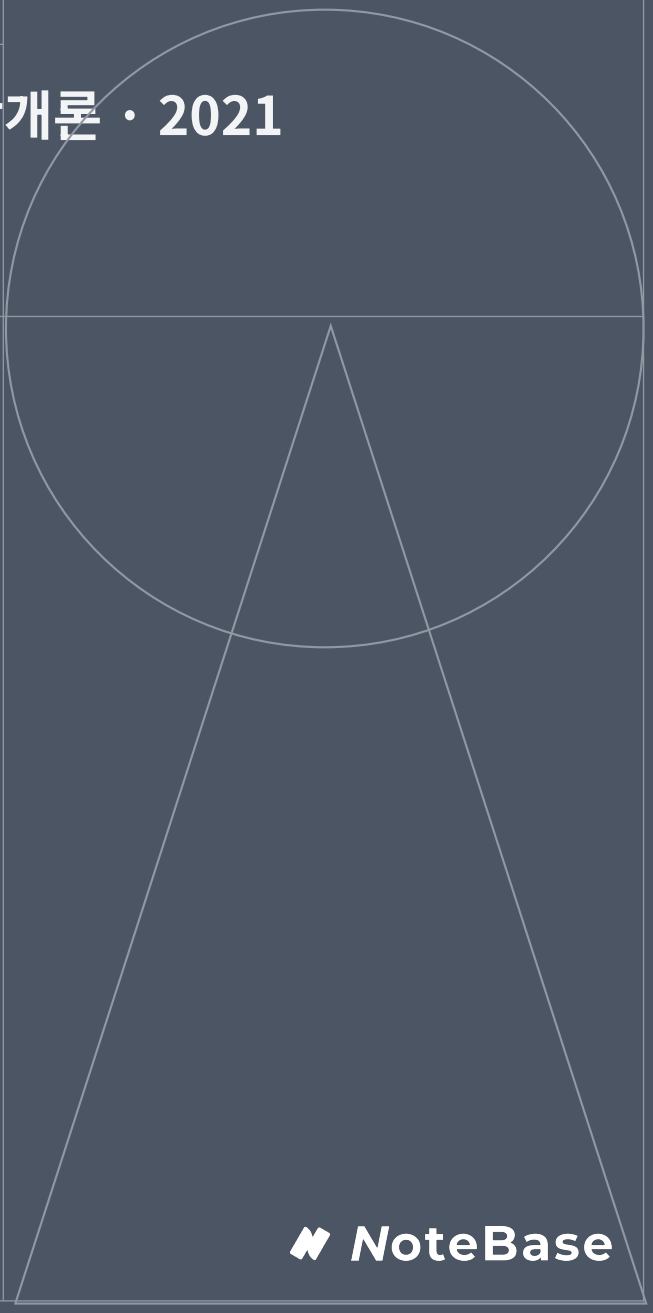
기출OX 해설집

공무원/국가직/9급 · 통계학개론 · 2021



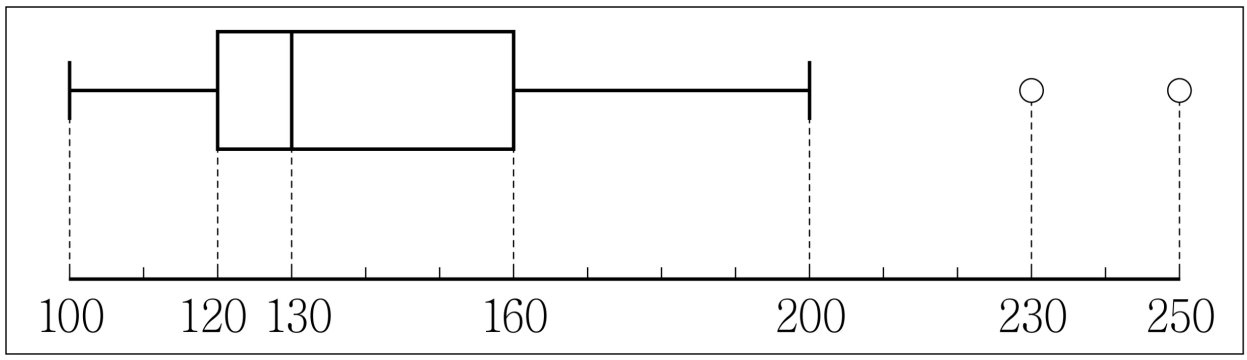
NoteBase

기출OX 해설집

공무원/국가직/9급 · 통계학개론 · 2021

공무원/국가직/9급 · 통계학개론 · 2021

문 1. 다음 상자그림에 대한 설명으로 옳은 것은?



① 자료의 표준편차는 150이다.

② 자료의 중앙값은 130이다.

③ 자료의 최빈값이 중앙값보다 크다.

④ 자료의 사분위수범위는 100이다.

정답 2번

간단한 해설

상자 안의 세로선은 자료를 크기순으로 나열했을 때 가운데에 놓이는 값인 **중앙값**을 나타낸다. 그림에서 이 선이 가리키는 값은 **130**이다.

NoteBase

선지별 해설

① X: 표준편차는 각 관측값이 평균에서 얼마나 떨어져 있는지를 나타내는 산포도이다. 상자그림에는 개별 관측값과 평균이 제시되어 있지 않으므로 **표준편차의 정확한 값은 알 수 없다**. 그림에서 최솟값은 100, 이상값까지 포함한 최댓값은 250이므로, 250 − 100 = 150은 표준편차가 아니라 범위이다.

② O: 상자 안에 그어진 세로선은 **중앙값**을 나타낸다. 이 선에서 아래로 내려온 점선이 가로축의 130에 닿으므로 자료의 중앙값은 130이다.

③ X: 최빈값은 자료에서 가장 자주 나타나는 값이다. 상자그림은 사분위수와 수염, 이상값 등으로 자료의 분포를 요약하며 **각 값의 출현 빈도는 보여 주지 않는다**. 따라서 최빈값이 얼마인지, 중앙값 130보다 큰지는 이 그림만으로 판단할 수 없다.

④ X: 사분위수범위는 **제3사분위수에서 제1사분위수를 뺀 값**으로, 상자의 가로 길이에 해당한다. 상자의 왼쪽 끝은 제1사분위수 120, 오른쪽 끝은 제3사분위수 160이므로 사분위수범위는 160 − 120 = 40이다. 200 − 100 = 100은 양쪽 수염 끝 사이의 차이이다.

NoteBase

문 2. n개의 자료 (x_i, y_i) $(i = 1, 2, \ldots, n)$의 표본상관계수는 0.2이다. $v_i = x_i + 1,\ w_i = 2y_i - 1$이라고 할 때, 변환된 n개의 자료 (v_i, w_i) $(i = 1, 2, \ldots, n)$의 표본상관계수는?

① 0.2

② 0.3

③ 0.4

④ 0.5

정답 1번

간단한 해설

각 변수에 상수를 더하거나 빼고 **양의 상수**를 곱해도 표본상관계수는 변하지 않는다. 따라서 변환된 자료의 표본상관계수도 0.2이다.

핵심 풀이

상관계수는 두 변수의 공분산을 각 변수의 표준편차로 나누어 표준화한 값이다. 각 자료에 같은 상수를 더하거나 빼면 평균도 그만큼 이동하므로 **평균으로부터의 편차는 변하지 않는다**. 따라서 x_i에 1을 더하고 $2y_i$에서 1을 빼는 것은 상관계수에 영향을 주지 않는다.

y_i에 2를 곱하면 두 변수의 공분산은 2배가 되지만, y_i의 표본표준편차도 2배가 된다. 원래 표본공분산을 s_{xy}, 각 변수의 표본표준편차를 s_x, s_y라고 하면, 변환 후에는 $s_{vw} = 2s_{xy}$, $s_v = s_x$, $s_w = 2s_y$이다. 따라서 상관계수에서는 **분자와 분모의 배율이 상쇄된다**.

$r_{vw} = \frac{2s_{xy}}{s_x \cdot 2s_y} = r_{xy} = 0.2$이므로 변환 후 표본상관계수는 ①의 0.2이다. 공분산이 2배가 된다고 해서 표준화한 상관계수까지 2배인 0.4가 되는 것은 아니다.

NoteBase

문 3. 어느 회사에서 차량 내비게이션을 연령대가 30대부터 50대까지인 고객에게 한 달 동안 사용하게 한 후 구매 의사를 알아보기 위해 240명을 임의추출하여 다음과 같은 분할표를 작성하였다. 연령대와 구매 의사는 서로 독립이라는 귀무가설을 검정하기 위한 카이제곱통계량의 p-값(유의확률)이 0.128일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면?

구매 의사 \ 연령대	30대	40대	50대	계
있음	40	60	40	140
없음	20	40	40	100
계	60	100	80	240

ㄱ. 귀무가설이 참일 때 30대이며 구매 의사가 있는 고객의 수의 기대도수는 35이다.

ㄴ. 귀무가설이 참일 때 카이제곱통계량의 자유도는 2이다.

ㄷ. 유의수준 5%에서 귀무가설을 기각한다.

① ㄴ

② ㄱ, ㄴ

③ ㄱ, ㄷ

④ ㄱ, ㄴ, ㄷ

정답 2번

간단한 해설

유의확률 0.128이 유의수준 0.05보다 크므로, 연령대와 구매 의사가 독립이라는 귀무가설을 **기각하지 못한다**.

NoteBase

보기별 해설

ㄱ. O: 두 변수가 독립이면 각 칸에 속할 확률은 **해당 행과 열의 주변비율을 곱한 값**이다. 따라서 기대도수는 행 합계×열 합계전체 합계로 구한다. 표에서 구매 의사가 있는 고객은 140명, 30대 고객은 60명이므로 해당 칸의 기대도수는 $\frac{140 \times 60}{240} = 35$이다. 이는 독립을 가정할 때 예상되는 도수이며, 실제 관측도수 40명과 구별한다.

ㄴ. O: 독립성 검정의 카이제곱통계량은 귀무가설 아래에서 자유도가 (행 수 − 1) × (열 수 − 1)인 카이제곱분포를 근사적으로 따른다. **합계 행과 열을 제외**하면 구매 의사는 2개 범주, 연령대는 3개 범주이므로 자유도는 (2 − 1) × (3 − 1) = 2이다.

ㄷ. X: 유의확률이 유의수준 이하이면 귀무가설을 기각한다. 이 문제에서는 0.128 > 0.05이므로 유의수준 5%에서 귀무가설을 **기각하지 못한다**. 즉, 두 변수가 독립이 아니라고 판단할 통계적 근거가 부족하다. 이것이 독립임을 증명했다는 뜻은 아니다.

옳은 보기는 ㄱ, ㄴ이며, 이를 모두 포함한 조합은 ②이다.

NoteBase

문 4. 단순선형회귀모형 $y_i = \alpha + \beta x_i + \varepsilon_i$ $(i = 1, 2, \ldots, n)$에서 최소제곱법에 의하여 추정된 회귀계수를 $\hat{\alpha}$와 $\hat{\beta}$, y_i의 예측값을 $\hat{y}_i$, 잔차를 $e_i = y_i - \hat{y}_i$, 총제곱합(SST)을 $\sum_{i=1}^{n}(y_i - \bar{y})^2$, 회귀제곱합(SSR)을 $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, 잔차제곱합(SSE)을 $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$이라고 할 때, 이에 대한 설명으로 옳지 않은 것은? (단, ε_i는 $N(0, \sigma^2)$을 따르고 서로 독립이다)

① SSE는 σ^2의 불편추정량(unbiased estimator)이다.

② 결정계수(R^2)는 $\frac{SSR}{SST}$이다.

③ $\sum_{i=1}^{n} e_i = 0$이다.

④ $\sum_{i=1}^{n} x_i e_i = 0$이다.

정답 1번

간단한 해설

SSE는 잔차를 제곱하여 더한 값이므로, 오차분산을 추정하려면 **잔차 자유도**로 나누어야 한다. 절편과 기울기를 추정한 단순선형회귀에서 σ^2의 불편추정량은 $\frac{SSE}{n-2}$이다.

선지별 해설

① X: 불편추정량은 기댓값이 추정하려는 모수와 같은 통계량이다. 절편과 기울기 두 개를 추정하므로 **잔차 자유도는** $n-2$이며, 주어진 모형에서 $E(SSE) = (n-2)\sigma^2$이다. 따라서 $n > 2$일 때 σ^2의 불편추정량은 SSE 자체가 아니라 잔차평균제곱 $MSE = \frac{SSE}{n-2}$이다.

② O: **절편을 포함한 최소제곱 회귀**에서는 전체 변동이 회귀식으로 설명되는 변동과 설명되지 않는 잔차 변동으로 나뉘어 $SST = SSR + SSE$가 성립한다. 결정계수는 전체 변동 중 회귀식이 설명하는 비율이므로, $SST > 0$일 때 $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$이다.

③ O: 최소제곱법은 잔차제곱합을 가장 작게 하는 절편과 기울기를 구한다. 잔차제곱합을 **절편에 대해 편미분**하여 0으로 놓으면 $-2\sum_{i=1}^{n} e_i = 0$이라는 정규방정식을 얻는다. 따라서 $\sum_{i=1}^{n} e_i = 0$이며, 이는 오차의 기댓값에 관한 진술이 아니라 적합 후 잔차들이 실제로 만족하는 관계이다.

④ O: 잔차제곱합을 **기울기에 대해 편미분**하여 0으로 놓으면 $-2\sum_{i=1}^{n} x_i e_i = 0$이라는 정규방정식을 얻는다. 따라서 $\sum_{i=1}^{n} x_i e_i = 0$이다. 이는 각 잔차가 0이라는 뜻이 아니라, 설명변수와 잔차의 곱을 모두 더하면 0이 된다는 뜻이다.

NoteBase

문 5. 어느 대학교의 4개 단과대학인 인문대, 사회대, 자연대, 공대에서 각각 50명씩 임의추출하여 남학생과 여학생의 수를 조사한 분할표가 다음과 같다. 4개 단과대학의 남학생 비율이 모두 같다는 귀무가설을 검정하기 위한 카이제곱통계량의 값과 유의수준 5%에서 검정 결과를 옳게 짝 지은 것은? (단, $\chi^2_\alpha(k)$는 자유도가 k인 카이제곱분포의 제$100 \times (1-\alpha)$ 백분위수를 나타낼 때 $\chi^2_{0.05}(3) = 7.81$, $\chi^2_{0.05}(4) = 9.49$이다)

성별 \ 단과대학	인문대	사회대	자연대	공대	계
남학생	20	25	25	30	100
여학생	30	25	25	20	100
계	50	50	50	50	200

① 카이제곱통계량의 값: 4, 검정 결과: 귀무가설을 기각한다.

② 카이제곱통계량의 값: 4, 검정 결과: 귀무가설을 기각하지 못한다.

③ 카이제곱통계량의 값: 8, 검정 결과: 귀무가설을 기각한다.

④ 카이제곱통계량의 값: 8, 검정 결과: 귀무가설을 기각하지 못한다.

정답 2번

간단한 해설

귀무가설 아래에서 각 단과대학의 남학생·여학생 기대도수는 모두 25명이며, 카이제곱통계량은 4이다. **자유도 3의 임계값 7.81보다 작으므로** 유의수준 5%에서 귀무가설을 기각하지 못한다.

NoteBase

계산 과정

1. 기대도수: 네 단과대학의 남학생 비율이 같다는 귀무가설을 카이제곱 동질성 검정으로 검정한다. 표의 남학생 합계는 100명, 여학생 합계도 100명이므로, 공통 남학생 비율의 추정값은 p=(100)/(200)=12이다. 각 대학의 표본이 50명이므로 남학생 기대도수는 50×12=25명이고 여학생도 25명이다. 따라서 **8개 칸의 기대도수가 모두 25명**으로, 카이제곱 근사를 적용하기에 충분히 크다.

2. 통계량: 관측도수를 O_{ij}, 귀무가설 아래의 기대도수를 E_{ij}라 하면 $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$이다. 각 칸에서 관측도수와 기대도수의 차이를 제곱한 뒤 기대도수로 나누어 합하므로, 귀무가설에서 벗어난 정도를 측정한다. 인문대의 남녀 관측도수는 20명·30명, 공대는 30명·20명으로 네 칸 모두 기대도수와 5명씩 차이 난다. 이 네 칸은 각각 $\frac{5^2}{25} = 1$만큼 기여한다. 사회대와 자연대는 남녀 모두 25명으로 기여도가 0이므로, 통계량은 1 + 1 + 1 + 1 = 4이다.

3. 자유도와 기각 여부: 성별이 2개 범주, 단과대학이 4개 범주이므로 자유도는 $(2-1)\times(4-1) = 3$이다. 관측도수가 기대도수에서 크게 벗어날수록 통계량이 커지므로 **오른쪽 꼬리를 기각역**으로 삼는다. 유의수준 5%의 임계값은 문제에 주어진 $\chi^2_{0.05}(3) = 7.81$이다. $4 < 7.81$이므로 귀무가설을 기각하지 못한다. 이는 네 대학의 남학생 비율이 다르다고 판단할 증거가 충분하지 않다는 뜻이며, 비율이 모두 같다고 입증한 것은 아니다.

통계량 4와 귀무가설을 기각하지 못한다는 검정 결과를 함께 제시한 선택지는 ②이다.

NoteBase

문 6. 확률변수 X에 대하여 $E[X(X-1)]=3$, $E[X(X+1)]=5$일 때, X의 분산은?

① 3

② 4

③ 5

④ 6

정답 1번

간단한 해설

분산은 **제곱의 기댓값에서 평균의 제곱을 뺀 값**이다. $E[X^2]=4$, $E[X]=1$이므로 분산은 3이다.

계산 과정

분산은 평균에서 벗어난 정도를 제곱하여 평균한 값으로, $\mathrm{Var}(X)=E[X^2]-(E[X])^2$로 계산한다. 주어진 곱을 전개하고 기댓값의 선형성을 적용하면 $E[X^2]-E[X]=3$, $E[X^2]+E[X]=5$를 얻는다.

두 식을 더하면 **평균에 해당하는 항이 소거**되어 $2E[X^2]=3+5=8$이다. 따라서 $E[X^2]=4$이며, 이는 확률변수의 값을 제곱한 뒤 기댓값을 구한 것이다.

두 번째 식에서 첫 번째 식을 빼면 $2E[X]=5-3=2$이다. 따라서 $E[X]=1$이며, 이는 확률변수의 평균이다.

분산 공식에 대입하면 $\mathrm{Var}(X)=4-1^2=3$이다. 분산을 구할 때는 **평균 자체가 아니라 평균의 제곱**을 뺀다. 계산한 분산 3은 ①의 값과 일치한다.

NoteBase

문 7. 갑, 을 두 사람이 가위바위보 게임을 10회 시행하여 갑이 이기는 횟수를 확률변수 X라 할 때, 이에 대한 설명으로 옳은 것은? (단, 갑, 을은 각자 임의로 가위, 바위, 보 중에서 하나를 내고, 비기는 경우도 1회 시행으로 간주한다)

① X의 확률분포는 대칭이다.

② 확률변수 $Y = 10 - X$는 10회의 시행에서 갑이 지는 횟수이다.

③ $Var(X) = \frac{5}{2}$이다.

④ $E(3X + 5) = 15$이다.

정답 4번

간단한 해설

한 번의 시행에서 갑이 이길 확률은 $\frac{1}{3}$이므로 $E(X) = \frac{10}{3}$이다. **기댓값의 선형성**에 따라
$E(3X + 5) = 3E(X) + 5 = 15$가 된다.

선지별 해설

두 사람이 각각 임의로 손 모양을 내면 가능한 조합은 9가지이고, 갑의 승리·패배·무승부에 해당하는 조합은 각각 3가지이다. 따라서 한 번의 시행에서 갑이 이길 확률은 $p = \frac{1}{3}$이다. 매번 독립적으로 같은 방식의 시행을 반복하므로, 승리 횟수 X는 이항분포 $B(10, \frac{1}{3})$을 따른다.

① X: 이항분포가 대칭이 되려면 **성공확률이 1/2**이어야 한다. 여기서는 $p = \frac{1}{3}$이므로 대칭분포가 아니다. 실제로 가능한 양 끝값인 0회와 10회의 확률은 각각 $P(X = 0) = (\frac{2}{3})^{10}$, $P(X = 10) = (\frac{1}{3})^{10}$으로 서로 다르다.

② X: 전체 10회에서 이긴 횟수를 빼면 **진 횟수와 비긴 횟수의 합**이 남는다. 따라서 $Y = 10 - X$는 갑이 지는 횟수가 아니라 이기지 못한 횟수이다. 무승부도 1회 시행으로 세므로 이를 패배와 구분해야 한다.

③ X: 이항분포의 분산은 **시행 횟수 × 성공확률 × 실패확률**이다. 여기서 실패는 승리하지 못하는 경우이므로 그 확률은 $1 - p = \frac{2}{3}$이다. 따라서 $\mathrm{Var}(X) = np(1 - p)$에 대입하면 $\mathrm{Var}(X) = 10 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{20}{9}$이며, $\frac{5}{2}$가 아니다.

④ O: 이항분포의 기댓값은 $E(X) = np = 10 \cdot \frac{1}{3} = \frac{10}{3}$이다. 이는 같은 조건의 10회 시행을 반복할 때 갑의 평균 승리 횟수를 뜻한다. **기댓값의 선형성** $E(aX + b) = aE(X) + b$를 적용하면 $E(3X + 5) = 3 \cdot \frac{10}{3} + 5 = 15$이다.

NoteBase

문 8. 단순선형회귀모형 $y_i = \alpha + \beta x_i + \varepsilon_i$ $(i = 1, 2, \ldots, n)$에 최소제곱법을 적용하여 얻은 분산분석표의 일부가 다음과 같을 때, 이에 대한 설명으로 옳지 않은 것은? (단, ε_i는 $N(0, \sigma^2)$을 따르고 서로 독립이다)

요인	제곱합	자유도	평균제곱	$F-$값	$p-$값
회귀	10	(㉠)	10	5	0.017
잔차	36	18	(㉡)		
계	46				

① ㉠과 ㉡의 값은 각각 1과 2이다.

② F-값은 회귀평균제곱(MSR)을 잔차평균제곱(MSE)으로 나눈 값이다.

③ 전체 자료의 수 n은 19이다.

④ 유의수준 5%에서 회귀모형이 유의하다.

정답 3번

간단한 해설

절편과 기울기를 추정하는 단순선형회귀에서 **잔차 자유도는 자료 수보다 2 작다**. 잔차 자유도가 18이므로 자료 수는 $n = 18 + 2 = 20$이며, 19는 자료 수가 아니라 전체 자유도이다.

NoteBase

선지별 해설

① O: 단순선형회귀의 회귀 자유도는 설명변수가 하나이므로 1이다. 표에서 회귀제곱합과 회귀평균제곱이 모두 10인 점으로도 $\frac{10}{10} = 1$임을 확인할 수 있다. 잔차평균제곱은 **잔차제곱합을 잔차 자유도로 나눈 값**이므로 $\text{MSE} = \frac{36}{18} = 2$이다. 따라서 ㉠은 1, ㉡은 2이다.

② O: 회귀모형의 유의성을 검정하는 F-값은 **회귀평균제곱을 잔차평균제곱으로 나눈 값**이다. 회귀가 설명하는 변동을 오차 변동과 비교하는 것으로, 이 표에서는 $F = \frac{\text{MSR}}{\text{MSE}} = \frac{10}{2} = 5$가 된다.

③ X: 절편이 있는 단순선형회귀에서는 절편과 기울기라는 두 모수를 추정하므로 **잔차 자유도는 n − 2**이다. 표의 잔차 자유도가 18이므로 $n - 2 = 18$에서 $n = 20$이다. 전체 자유도는 $n - 1 = 19$이므로, 선지는 전체 자유도를 자료 수로 잘못 해석했다. 자료 수는 19가 아니라 20이다.

④ O: 회귀모형의 유의성 검정은 기울기가 0이라는 귀무가설 $H_0 : \beta = 0$을 검정한다. 표에 제시된 p-값 0.017은 **유의수준 0.05보다 작으므로 귀무가설을 기각**한다. 따라서 유의수준 5%에서 설명변수와 종속변수 사이의 선형관계가 통계적으로 유의하다고 판단한다.

NoteBase

문 9. 정규분포 $N(\mu, \sigma^2)$을 따르는 모집단에서 n개의 표본을 임의 추출하여 구한 μ의 95% 신뢰구간이 $(-0.1, 0.7)$일 때, 이에 대한 설명으로 옳지 않은 것은?

① 가설 $H_0 : \mu = 0$ 대 $H_1 : \mu \neq 0$을 검정할 때 유의수준 5%에서 귀무가설을 기각하지 못한다.

② 표본평균은 0.3이다.

③ μ가 구간 $(-0.1, 0.7)$에 속하지 않을 확률은 0.05이다.

④ μ의 90% 신뢰구간의 길이는 0.8보다 작다.

정답 3번

간단한 해설

95% 신뢰수준은 같은 방법으로 표본 추출과 구간 추정을 반복할 때, 만들어진 구간이 모평균을 포함하는 **장기적인 비율**을 뜻한다. 모평균은 고정된 값이므로, 이미 구한 구간에 속하지 않을 확률이 0.05라고 해석할 수 없다.

선지별 해설

① O: 모평균의 양측 95% 신뢰구간과 대응하는 양측검정의 유의수준은 5%이다. 이 검정에서는 **귀무가설의 모평균 값이 신뢰구간 안에 있으면 기각하지 못한다**. 주어진 구간 $(-0.1, 0.7)$에 0이 포함되므로, $H_0 : \mu = 0$ 대 $H_1 : \mu \neq 0$의 검정에서 귀무가설을 기각하지 못한다.

② O: 정규모집단 모평균의 통상적인 양측 신뢰구간은 **표본평균을 중심으로 좌우 대칭**이다. 따라서 표본평균은 두 끝점의 중점인 $\bar{x} = \frac{-0.1 + 0.7}{2} = 0.3$이다.

③ X: 빈도주의 추론에서 모평균 μ는 고정된 모수이고, 표본에 따라 달라지는 것은 신뢰구간의 양 끝점이다. 표본을 관측하여 구간 $(-0.1, 0.7)$을 얻은 뒤에는 그 구간도 고정되므로, 모평균이 그 안에 속하거나 속하지 않는다는 사실에 각각 0.95와 0.05의 확률을 부여하지 않는다. 95% 신뢰구간의 올바른 해석은 같은 방법으로 표본 추출과 구간 추정을 반복할 때 **만들어진 구간의 약 95%가 참 모평균을 포함한다**는 것이다. 약 5%는 모평균을 포함하지 못하는 구간의 장기적인 비율이다.

④ O: 같은 표본에서 동일한 방식으로 모평균의 양측 신뢰구간을 만들 때, **신뢰수준을 낮추면 구간의 길이가 짧아진다**. 정규분포를 이용하는 경우든 모분산을 추정하여 t분포를 이용하는 경우든, 신뢰수준이 95%에서 90%로 낮아지면 임계값이 작아져 오차한계가 줄기 때문이다. 주어진 95% 신뢰구간의 길이는 $0.7 - (-0.1) = 0.8$이므로, 90% 신뢰구간의 길이는 0.8보다 작다.

NoteBase

문 10. 어느 회사 신입사원의 입사 시험성적은 평균이 μ인 정규분포를 따른다고 한다. 신입사원 중에서 n명을 임의추출하여 구한 시험성적의 표본평균은 60보다 크고, 가설 $H_0 : \mu = 60$ 대 $H_1 : \mu \neq 60$에 대한 p-값은 0.1이었다. 동일한 표본에 대해 가설 $H_0 : \mu = 60$ 대 $H_1 : \mu > 60$에 대한 p-값은 a이고, 가설 $H_0 : \mu = 60$ 대 $H_1 : \mu < 60$에 대한 p-값은 b일 때, $b - a$의 값은?

① 0.05

② 0.1

③ 0.9

④ 0.95

정답 3번

간단한 해설

표본평균이 60보다 크므로 우측검정의 p-값은 **양측검정 값의 절반**인 0.05이다. 좌측검정의 값은 그 여사건의 확률인 0.95이므로, $b - a = 0.95 - 0.05 = 0.9$이다.

계산 과정

검정통계량의 방향: p-값은 귀무가설이 참일 때 대립가설 방향으로 관측 결과 이상 극단적인 결과가 나올 확률이다. 평균을 검정하는 표준화 통계량을 T, 그 관측값을 t_0라 하자. 표본평균이 60보다 크므로 $t_0 > 0$이다. 귀무가설 $H_0 : \mu = 60$ 아래에서 이 통계량의 분포는 **0을 중심으로 대칭인 연속분포**이다.

우측검정의 값 a: 양측검정은 양쪽 꼬리를 모두 고려한다. 대칭성에 의해 t_0 이상의 우측 꼬리확률과 $-t_0$ 이하의 좌측 꼬리확률이 같으므로, 주어진 양측검정의 p-값 0.1은 우측 꼬리확률의 두 배이다. 따라서 $a = P(T \geq t_0) = \frac{0.1}{2} = 0.05$이다.

좌측검정의 값 b: 대립가설이 $\mu < 60$이면 관측값 t_0 이하의 확률을 구한다. 연속분포에서는 한 점의 확률이 0이므로 좌측과 우측 꼬리확률의 합은 1이다. 따라서 $b = P(T \leq t_0) = 1 - a = 0.95$이다. 관측된 표본평균이 좌측 대립가설과 반대 방향에 있으므로, 이 p-값은 크게 나온다.

차이 계산: $b - a = 0.95 - 0.05 = 0.9$이다. 이는 ③의 0.9와 일치한다.

NoteBase

문 11. 50개의 관측값을 가장 작은 값부터 가장 큰 값까지 크기순으로 나열하여 $x_{(1)}, x_{(2)}, \ldots, x_{(50)}$으로 나타낼 때, 이에 대한 설명으로 옳지 않은 것은?

① 제20 백분위수는 $x_{(20)}$이다.

② 제3 사분위수는 $x_{(38)}$이다.

③ 중앙값은 $\frac{x_{(25)}+x_{(26)}}{2}$이다.

④ 사분위수범위는 $(x_{(38)}-x_{(13)})$이다.

정답 1번

간단한 해설

제20백분위수의 '20'은 **관측값의 순번이 아니라 백분율**을 뜻한다. 이 문항의 계산 규칙에서는 $50 \times 0.2 = 10$이 정수이므로, 제20백분위수는 10번째와 11번째 값의 평균인 $\frac{x_{(10)}+x_{(11)}}{2}$이다.

선지별 해설

$x_{(k)}$는 크기순으로 정렬했을 때 k번째 관측값이다. 이 문항에서는 제p백분위수를 구할 때 위치 $r=\frac{50p}{100}$를 계산하여, r이 정수이면 r번째와 그다음 관측값의 평균을 취하고, 정수가 아니면 r을 올림한 순번의 관측값을 취하는 규칙을 적용한다.

① X: 제20백분위수의 위치는 $\frac{50\times 20}{100}=10$이다. 위치가 정수이므로 **10번째와 11번째 관측값의 평균**을 취한다. 따라서 제20백분위수는 $x_{(20)}$이 아니라 $\frac{x_{(10)}+x_{(11)}}{2}$이다.

② O: 제3사분위수는 **제75백분위수**이다. 위치가 $\frac{50\times 75}{100}=37.5$이므로 올림한 38번째 관측값 $x_{(38)}$이다.

③ O: 중앙값은 자료를 크기순으로 나열했을 때 가운데에 놓이는 값이다. 자료 수가 50개로 짝수이므로 **가운데 두 관측값의 평균**을 취한다. 가운데 순번은 25번째와 26번째이므로 중앙값은 $\frac{x_{(25)}+x_{(26)}}{2}$이다.

④ O: 사분위수범위는 **제3사분위수에서 제1사분위수를 뺀 값**으로, 가운데 50% 자료가 놓인 구간의 폭을 나타낸다. 제1사분위수는 제25백분위수이며, 위치가 $\frac{50\times 25}{100}=12.5$이므로 $x_{(13)}$이다. 제3사분위수는 $x_{(38)}$이므로 사분위수범위는 $x_{(38)}-x_{(13)}$이다.

NoteBase

문 12. 10개의 자료 $(x_1, y_1), (x_2, y_2), \ldots, (x_{10}, y_{10})$에 대하여 단순선형 회귀모형 $y_i = \alpha + \beta x_i + \varepsilon_i$ $(i = 1, 2, \ldots, 10)$를 적합하고자 한다.
$\bar{x} = 20,\ \bar{y} = 80,\ \sum_{i=1}^{10} (x_i - \bar{x})^2 = 20,\ \sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 60$
일 때, 최소제곱법을 이용하여 추정한 회귀계수 $\hat{\alpha}$와 $\hat{\beta}$의 값은? (단, $\bar{x}, \bar{y}$는 각각 x, y의 표본평균이고, ε_i는 $N(0, \sigma^2)$를 따르고 서로 독립이다.)

① $\hat{\alpha} = 30,\ \hat{\beta} = 2$
② $\hat{\alpha} = 30,\ \hat{\beta} = 3$
③ $\hat{\alpha} = 20,\ \hat{\beta} = 2$
④ $\hat{\alpha} = 20,\ \hat{\beta} = 3$

정답 4번

간단한 해설

최소제곱법에서 **기울기**는 두 변수의 편차곱합을 x의 편차제곱합으로 나눈 값이므로 $\hat{\beta} = \frac{60}{20} = 3$이다.
회귀직선은 **표본평균점**을 지나므로 절편은 $\hat{\alpha} = 80 - 3 \cdot 20 = 20$이다.

계산 과정

최소제곱법은 관측값과 회귀직선의 예측값 사이의 차이인 잔차의 제곱합을 최소화하여 절편 α와 기울기 β를 추정하는 방법이다. 이 문제는 절편이 있는 단순선형 회귀모형이므로, 먼저 기울기를 구한 뒤 두 표본평균으로 절편을 구한다.

기울기 계산: x의 편차제곱합을 S_{xx}, x와 y의 편차곱합을 S_{xy}라고 하자. 주어진 값은 $S_{xx} = 20, S_{xy} = 60$이다. $S_{xx} > 0$일 때 최소제곱 기울기는 $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{60}{20} = 3$이다. 이는 **x가 1 증가할 때 예측되는 y가 3 증가**한다는 뜻이다.

절편 계산: 절편이 있는 최소제곱 회귀직선은 **표본평균점** $(\bar{x}, \bar{y})$를 지난다. 따라서 $\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$이고, $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$이다. 여기에 $\bar{x} = 20, \bar{y} = 80, \hat{\beta} = 3$을 대입하면 $\hat{\alpha} = 80 - 3 \cdot 20 = 20$이다.

추정한 회귀직선은 $\hat{y} = 20 + 3x$이며, $\hat{\alpha} = 20, \hat{\beta} = 3$은 ④의 값과 일치한다.

NoteBase

문 13. 인자의 수준수가 k이고 각 수준에서의 반복수가 n인 일원배치법에서 얻은 분산분석표의 일부가 다음과 같을 때, 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	$F-$값	$p-$값
처리	60		20	(㉠)	< 0.001
오차	28	28			
계	88				

① 인자의 수준수 k는 4이다.

② 반복수 n은 8이다.

③ ㉠의 값은 20이다.

④ 인자 수준에 따라 모평균이 같은지를 유의수준 5%에서 검정할 때 모평균은 모두 같다고 할 수 있다.

정답 4번

간단한 해설

표의 p-값이 0.001보다 작아 유의수준 0.05보다 작으므로, '모평균이 모두 같다'는 귀무가설을 **기각한다**. 따라서 모평균이 모두 같다고 할 수 없으며, **적어도 한 쌍의 모평균은 다르다**고 판단한다.

NoteBase

선지별 해설

① O: 일원배치 분산분석에서 처리 자유도는 $k-1$이고, 평균제곱은 제곱합을 자유도로 나눈 값이다. 표에서 처리 제곱합이 60, 처리 평균제곱이 20이므로 처리 자유도는 $\frac{60}{20}=3$이다. 따라서 $k-1=3$이므로 **수준수는 4**이다.

② O: 각 수준의 반복수가 n으로 같으면 전체 관측값 수는 kn이고, 각 수준에서 평균을 하나씩 추정하므로 오차 자유도는 $kn-k=k(n-1)$이다. 표의 오차 자유도가 28이고 수준수가 4이므로 $4(n-1)=28$에서 **각 수준의 반복수는 8**이다.

③ O: F-값은 **처리 평균제곱을 오차 평균제곱으로 나눈 값**이다. 오차 제곱합과 오차 자유도가 모두 28이므로 오차 평균제곱은 $\frac{28}{28}=1$이다. 따라서 $F=\frac{20}{1}=20$이므로 ㉠은 20이다.

④ X: 검정의 귀무가설은 $H_0:\mu_1=\mu_2=\mu_3=\mu_4$이고, 대립가설은 모평균이 모두 같지는 않다는 것이다. 표의 p-값이 0.001보다 작으므로 유의수준 5%에 해당하는 0.05보다 작아 귀무가설을 기각한다. 따라서 '모평균은 모두 같다'가 아니라 **적어도 한 쌍의 모평균은 다르다**고 판단한다. 이 검정만으로 모든 쌍의 모평균이 서로 다르다고 결론 내리는 것은 아니다.

NoteBase

문 14. 전체 고등학생이 치른 수학 시험점수는 평균이 60점, 표준편차가 12점인 정규분포를 따른다고 한다. 전체 고등학생 중에서 임의로 뽑은 한 명의 수학 시험점수를 X라 하고, 임의로 뽑은 36명의 수학 시험점수의 표본평균을 Y라 할 때, 이에 대한 설명으로 옳지 않은 것은?

① $E(X) = E(Y)$

② $Var(X) = 2 \times Var(Y)$

③ $P(X > 60) = P(Y > 60)$

④ $P(X < 72) = P(Y < 62)$

정답 2번

간단한 해설

표본평균의 분산은 **모분산을 표본크기 36으로 나눈 값**이다. 따라서 $\mathrm{Var}(X) = 36\mathrm{Var}(Y)$이며, 2배라는 설명은 틀리다.

선지별 해설

임의표본의 표본평균은 평균이 모평균과 같고, 분산은 모분산을 표본크기로 나눈 값이다. 모집단이 정규분포를 따르면 표본평균도 정규분포를 따른다.

① O: 한 명의 점수의 기댓값은 $E(X) = 60$이고, 36명 점수의 표본평균도 **모평균을 중심으로 분포**하므로 $E(Y) = 60$이다. 따라서 $E(X) = E(Y)$이다.

② X: 한 명의 점수의 분산은 $\mathrm{Var}(X) = 12^2 = 144$이다. 표본평균의 분산은 **모분산의 1/36**이므로 $\mathrm{Var}(Y) = \frac{12^2}{36} = 4$이다. 따라서 올바른 관계는 $\mathrm{Var}(X) = 36\mathrm{Var}(Y)$이다. 표본평균의 표준편차는 $\sqrt{4} = 2$점이며, 이 값은 두 분산의 배수가 아니다.

③ O: 두 정규분포는 모두 평균이 60점이고 **평균을 기준으로 좌우 대칭**이다. 따라서 표준편차가 달라도 평균보다 클 확률은 각각 0.5이므로 $P(X > 60) = P(Y > 60)$이다.

④ O: 한 명의 점수는 표준편차가 12점이고, 표본평균은 표준편차가 2점이다. 72점을 X의 분포에서 표준화하면 $\frac{72-60}{12} = 1$이고, 62점을 Y의 분포에서 표준화하면 $\frac{62-60}{2} = 1$이다. 두 기준값이 모두 **각 평균보다 표준편차 한 개만큼 큰 값**이므로, 표준정규확률변수 Z에 대하여 두 확률은 모두 $P(Z < 1)$이다. 따라서 $P(X < 72) = P(Y < 62)$이다.

NoteBase

문 15. 정규분포 $N(\mu, 2^2)$을 따르는 모집단에서 임의추출한 9개 표본의 표본평균이 12일 때, 가설 $H_0 : \mu = 10$ 대 $H_1 : \mu \neq 10$에 대한 p-값과 같은 것은? (단, Z는 표준정규분포를 따르는 확률변수이다)

① $2 \times P(Z \geq 3)$

② $P(Z \geq 3)$

③ $P(Z \leq 3)$

④ $2 \times P(Z \leq 3) - 1$

정답 1번

간단한 해설

표본평균 12는 귀무가설의 평균 10보다 표준오차의 3배만큼 크다. 대립가설이 평균 10과 다른 경우를 모두 포함하는 **양측검정**이므로, p-값은 양쪽 꼬리확률을 합한 $2P(Z \geq 3)$이다.

계산 과정

표본평균의 분포: 모집단이 정규분포이고 모표준편차가 2로 주어졌으므로, 표본크기 9인 표본평균 $\bar{X}$도 정규분포를 따른다. **표본평균의 표준편차인 표준오차**는 (σ)/(n)=(2)/(9)=(2)/(3)이다. 모표준편차를 알고 있으므로 표준정규분포를 이용한다.

검정통계량: 귀무가설 $H_0 : \mu = 10$이 참일 때 $Z = \frac{\bar{X} - 10}{2/3}$은 표준정규분포를 따른다. 관측된 표본평균 12를 대입하면 $z = \frac{12 - 10}{2/3} = 3$이다. 이는 관측된 표본평균이 귀무가설의 평균보다 표준오차의 3배만큼 위에 있다는 뜻이다.

양측 꼬리확률: p-값은 귀무가설이 참일 때 관측 결과만큼 또는 그보다 대립가설을 지지하는 극단적인 결과가 나올 확률이다. 대립가설이 $H_1 : \mu \neq 10$이므로 **평균보다 큰 방향과 작은 방향을 모두** 고려하여 p=P(Z≤-3)+P(Z≥3)으로 구한다. 표준정규분포는 0을 중심으로 대칭이므로 두 꼬리확률은 같다. 따라서 p=2P(Z≥3)이며, 이는 ①의 식과 같다.

NoteBase

문 16. 어느 공장에서 생산되는 제품의 불량률 p에 대한 가설 $H_0 : p = 0.04$ 대 $H_1 : p > 0.04$를 검정하려고 한다. 이 공장에서 임의추출한 384개 제품 중 불량품의 개수를 X라고 하자. 기각역으로 $R : \frac{X}{384} > 0.06$을 사용할 때, 제1종 오류를 범할 확률과 같은 것은? (단, Z는 표준정규분포를 따르는 확률변수이다)

① $P(Z > 1.5)$

② $P(Z > 2)$

③ $P(Z > 2.5)$

④ $P(Z > 3)$

정답 2번

간단한 해설

제1종 오류는 **귀무가설이 참인데 이를 기각하는 오류**이므로, 불량률이 0.04일 때 표본불량률이 0.06을 넘을 확률을 구한다. 귀무가설 아래 표준오차는 0.01이므로 기각 경계는 평균보다 표준오차의 2배만큼 높으며, 정규근사하면 그 확률은 $P(Z > 2)$이다.

계산 과정

제1종 오류의 확률: 제1종 오류는 귀무가설이 참인데도 기각하는 오류이다. 따라서 **실제 불량률을 0.04로 두고**, 주어진 기각역에 들어갈 확률을 계산한다. 표본불량률을 $\hat{p} = \frac{X}{384}$라 하면 구할 확률은 $\alpha = P_{p=0.04}(\hat{p} > 0.06)$이다.

표본불량률의 표준오차: 이항모형에서 귀무가설 아래 불량품 수는 $X \sim B(384, 0.04)$이고, 표본불량률의 평균은 0.04이다. 표준오차는 $\sqrt{\frac{0.04 \cdot 0.96}{384}} = 0.01$이다. 이는 불량률이 4%인 공장에서 384개씩 임의추출할 때, 표본불량률이 평균 주위로 변동하는 크기가 1%포인트임을 뜻한다.

정규근사: 귀무가설 아래 기대 불량품 수는 $384 \times 0.04 = 15.36$개, 기대 정상제품 수는 $384 \times 0.96 = 368.64$개로 모두 충분히 크므로 표본불량률을 정규분포로 근사한다. 이에 따라 $\frac{\hat{p} - 0.04}{0.01}$은 근사적으로 표준정규분포를 따른다.

기각 경계의 표준화: 경계 0.06을 표준화하면 $\frac{0.06 - 0.04}{0.01} = 2$이다. 대립가설이 불량률 증가를 뜻하고 기각역도 표본불량률이 0.06보다 큰 영역이므로, **표준화한 경계의 오른쪽 꼬리확률**을 구한다. 따라서 $\alpha \approx P(Z > 2)$이며, 이는 ②의 확률식에 해당한다.

NoteBase

문 17. 1부터 13까지의 자연수 중 하나와 4개의 무늬 ♣, ♠, ♥, ♦ 중 하나를 조합하여 각각 적은 같은 크기의 52장의 카드가 있다. 이 52장의 카드에서 임의로 한 장의 카드를 뽑을 때, 이 카드에 ♣ 무늬가 적힌 사건을 A, 숫자 1이 적힌 사건을 B라고 하자. 이에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. $P(A) = \frac{1}{4}$

ㄴ. $P(B) = \frac{1}{13}$

ㄷ. 사건 A와 B는 서로 독립이다.

ㄹ. $P(A^C \cup B) = \frac{9}{13}$ (단, A^C은 A의 여사건이다)

① ㄱ, ㄹ

② ㄴ, ㄷ

③ ㄱ, ㄴ, ㄷ

④ ㄱ, ㄴ, ㄷ, ㄹ

정답 3번

간단한 해설

ㄹ에서 $A^C \cup B$는 **♣가 아닌 카드 39장과 ♣ 무늬의 숫자 1 카드 1장**을 포함한다. 따라서 그 확률은 $\frac{40}{52} = \frac{10}{13}$이다.

NoteBase

보기별 해설

52장의 카드가 뽑힐 가능성은 모두 같으므로, 각 사건의 확률은 그 사건에 해당하는 카드 수를 52로 나누어 구한다.

ㄱ. O: ♣ 무늬에는 1부터 13까지의 숫자가 각각 하나씩 있으므로 해당 카드는 **13장**이다. 따라서 $P(A) = \frac{13}{52} = \frac{1}{4}$이다.

ㄴ. O: 숫자 1은 네 무늬마다 하나씩 있으므로 해당 카드는 **4장**이다. 따라서 $P(B) = \frac{4}{52} = \frac{1}{13}$이다.

ㄷ. O: 두 사건이 독립일 조건은 **교집합의 확률이 각 확률의 곱과 같은 것**이다. ♣ 무늬이면서 숫자 1인 카드는 1장이므로 $P(A \cap B) = \frac{1}{52}$이다. 한편 $P(A)P(B) = \frac{1}{4} \cdot \frac{1}{13} = \frac{1}{52}$이므로 두 사건은 서로 독립이다.

ㄹ. X: A^C는 ♣가 아닌 카드 전체이므로 39장이 해당한다. 여기에 숫자 1인 카드들을 합치면, ♠ · ♥ · ♦ 무늬의 숫자 1 카드는 이미 포함되어 있으므로 **♣ 무늬의 숫자 1 카드만 추가**된다. 따라서 합집합에 해당하는 카드는 39 + 1 = 40장이고, $P(A^C \cup B) = \frac{40}{52} = \frac{10}{13}$이다. 제시된 확률 $\frac{9}{13}$은 $\frac{10}{13}$으로 고쳐야 한다.

옳은 보기는 ㄱ, ㄴ, ㄷ이므로 이 조합을 제시한 ③에 해당한다.

NoteBase

문 18. 성공의 확률이 p인 베르누이 시행을 독립적으로 100회 반복할 때 성공 횟수를 확률변수 X라 하자. 이에 대한 설명으로 옳지 않은 것은? (단, $0 < p < 1$)

① $p = \frac{1}{2}$이면 $P(X \le 49) = P(X \ge 51)$이다.

② X의 분산은 $p = \frac{1}{2}$일 때 최댓값을 가진다.

③ $p < \frac{1}{2}$이면 $P(X \le 49) < P(X \ge 51)$이다.

④ $p > \frac{1}{2}$이면 $P(X = 49) < P(X = 51)$이다.

정답 3번

간단한 해설

③의 **부등호 방향이 반대**이다. $p < \frac{1}{2}$이면 50을 기준으로 대칭인 두 성공 횟수 중 작은 쪽의 확률이 더 크므로, P(X≤49)>P(X≥51)이다.

선지별 해설

독립적인 베르누이 시행을 100회 반복한 성공 횟수이므로 X는 이항분포 $B(100, p)$를 따른다. 각 성공 횟수의 확률을 $f(k) = P(X = k)$라 쓰면, $f(k) = \binom{100}{k} p^k (1-p)^{100-k}$이다. 여기서 $k = 0, 1, \ldots, 100$이다.

① O: $p = \frac{1}{2}$이면 성공과 실패의 확률이 같고, $\binom{100}{k} = \binom{100}{100-k}$이므로 $f(k) = f(100-k)$이다. 즉, **50을 중심으로 확률이 대칭**이다. 성공 횟수 0~49는 각각 100~51과 짝을 이루므로, 이 확률들을 더하면 P(X≤49)=P(X≥51)이다.

② O: 독립인 베르누이 확률변수 하나의 분산은 $p(1-p)$이고, 독립인 확률변수의 합에서는 분산을 더하므로 $\mathrm{Var}(X) = 100p(1-p)$이다. 이를 정리하면 $100p(1-p) = 25 - 100\left(p - \frac{1}{2}\right)^2$이다. 빼는 제곱항이 0 이상이므로 분산은 25 이하이며, **제곱항이 0이 되는** $p = \frac{1}{2}$에서 최댓값 25를 갖는다.

③ X: $p < \frac{1}{2}$이면 $1 - p > p$이다. 50보다 작은 성공 횟수 k와 대칭인 $100 - k$를 비교하면 $\frac{f(k)}{f(100-k)} = \left(\frac{1-p}{p}\right)^{100-2k}$이다. $k = 0, 1, \ldots, 49$에서는 밑이 1보다 크고 지수가 양수이므로 이 비는 1보다 크다. 따라서 각 짝에서 **작은 성공 횟수의 확률이 더 크다**. 0~49의 확률과 이에 대응하는 100~51의 확률을 각각 더하면 P(X≤49)>P(X≥51)이므로, 선지의 부등호를 반대로 고쳐야 한다.

④ O: 성공 횟수 49와 51의 이항계수는 같으므로 확률의 비는 $\frac{f(51)}{f(49)} = \left(\frac{p}{1-p}\right)^2$이다. $p > \frac{1}{2}$이면 **성공확률이 실패확률보다 크므로** 이 비가 1보다 크다. 따라서 $P(X = 49) < P(X = 51)$이다.

NoteBase

문 19. 분산 σ^2이 알려진 정규분포 $N(\mu, \sigma^2)$을 따르는 모집단에서 임의추출한 n개 표본을 이용하여 유의수준 α에서 가설 $H_0 : \mu = \mu_0$ 대 $H_1 : \mu \neq \mu_0$을 검정하려고 한다. 검정통계량이 $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, $0 < \alpha < 1$, μ_0은 상수이고 $\overline{X}$는 표본평균이다)

ㄱ. $|Z| \geq z_\alpha$이면 H_0을 기각한다. (단, z_α는 표준정규분포의 제$100 \times (1 - \alpha)$ 백분위수이다)
ㄴ. μ_0이 μ에 대한 $100 \times (1 - \alpha)$% 신뢰구간에 속하지 않으면 p-값이 α보다 작다.
ㄷ. $Z^2 \geq \chi^2_\alpha(1)$이면 H_0을 기각한다. (단, $\chi^2_\alpha(k)$는 자유도가 k인 카이제곱분포의 제$100 \times (1 - \alpha)$ 백분위수이다)

① ㄴ
② ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

정답 3번

간단한 해설

양측검정은 **양쪽 꼬리에 유의수준을 절반씩** 배분하므로 임계값은 z_α가 아니라 $z_{\alpha/2}$이다. 표준정규 검정통계량을 제곱하면 자유도 1인 카이제곱분포를 따르므로, 제곱값의 상위 꼬리를 이용해서도 같은 검정을 할 수 있다.

NoteBase

보기별 해설

모집단이 정규분포를 따르고 모분산을 알고 있으므로, 귀무가설 H_0 아래에서 검정통계량 Z는 표준정규분포를 따른다. 대립가설이 $\mu \neq \mu_0$이므로 표본평균이 가설의 모평균보다 충분히 작거나 큰 경우를 모두 기각하는 양측검정이다.

ㄱ. X: 유의수준 α는 귀무가설이 참일 때 이를 기각할 확률이다. 양측검정에서는 **좌우 꼬리에 각각** $\alpha/2$의 확률을 배분하므로, 기각역은 $|Z| \geq z_{\alpha/2}$이다. z_α는 오른쪽 꼬리확률이 α인 값이므로 양측검정의 임계값으로 그대로 사용할 수 없다.

ㄴ. O: 모평균의 $100(1-\alpha)\%$ 신뢰구간은 $\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$이다. μ_0이 이 구간 밖에 있으면 $|\overline{X} - \mu_0| > z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ 이므로 $|Z| > z_{\alpha/2}$이다. 양측검정의 p-값은 $p = 2\{1 - \Phi(|Z|)\}$이며, Φ는 표준정규분포의 누적분포함수이다. 따라서 **신뢰구간 밖**이라는 조건은 $p < \alpha$로 이어진다.

ㄷ. O: 귀무가설 아래에서 **표준정규 확률변수의 제곱**은 자유도 1인 카이제곱분포를 따르므로 $Z^2 \sim \chi^2(1)$이다. 제곱값이 크다는 것은 원래의 Z가 양수 또는 음수 방향으로 0에서 멀다는 뜻이다. 따라서 제곱값의 상위 꼬리확률이 α가 되도록 $Z^2 \geq \chi^2_\alpha(1)$일 때 기각한다. 임계값의 관계는 $\chi^2_\alpha(1) = z^2_{\alpha/2}$이므로 이 기각역은 양측 Z검정의 기각역과 같다.

옳은 보기는 ㄴ, ㄷ이며, 이 조합은 ③에 해당한다.

NoteBase

문 20. 한 요인의 3개 수준 A, B, C에서 10회씩 반복 측정한 자료에 대해 요인의 수준에 따라 반응변수 Y의 평균에 차이가 있는지 알아보기 위하여 다음 선형회귀모형을 적합하고자 한다.

$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (i = 1, 2, \ldots, 30)$

위 모형에서 i번째 자료의 요인의 수준이 A이면 $(x_{1i}, x_{2i}) = (0, 0)$, 요인의 수준이 B이면 $(x_{1i}, x_{2i}) = (1, 0)$, 요인의 수준이 C이면 $(x_{1i}, x_{2i}) = (0, 1)$이다. 모수 β_1, β_2의 가설검정 결과에 대한 해석으로 옳은 것은? (단, ε_i는 $N(0, \sigma^2)$을 따르고 서로 독립이다)

① 가설 $H_0 : \beta_1 = 0$ 대 $H_1 : \beta_1 \neq 0$에서 H_0을 기각하지 못한다면 수준 A와 수준 B의 반응변수의 평균은 유의한 차이가 있다.

② 가설 $H_0 : \beta_1 = 0$ 대 $H_1 : \beta_1 \neq 0$에서 H_0을 기각한다면 수준 B와 수준 C의 반응변수의 평균은 유의한 차이가 있다.

③ 가설 $H_0 : \beta_2 = 0$ 대 $H_1 : \beta_2 \neq 0$에서 H_0을 기각하지 못한다면 수준 B와 수준 C의 반응변수의 평균은 유의한 차이가 없다.

④ 가설 $H_0 : \beta_2 = 0$ 대 $H_1 : \beta_2 \neq 0$에서 H_0을 기각한다면 수준 A와 수준 C의 반응변수의 평균은 유의한 차이가 있다.

정답 4번

간단한 해설

이 모형에서 β_2는 **수준 C와 수준 A의 평균 차이**이다. 따라서 $H_0 : \beta_2 = 0$을 기각하면 두 수준의 평균에 통계적으로 유의한 차이가 있다고 판단한다.

NoteBase

선지별 해설

각 수준의 반응변수 평균을 μ_A, μ_B, μ_C라 하자. 오차의 평균이 0이므로 주어진 지시변수 값을 모형에 대입하면 $\mu_A = \alpha, \mu_B = \alpha + \beta_1, \mu_C = \alpha + \beta_2$이다. 따라서 $\beta_1 = \mu_B - \mu_A, \beta_2 = \mu_C - \mu_A$이며, A가 비교의 기준 수준이다.

① X: $H_0 : \beta_1 = 0$은 A와 B의 평균이 같다는 가설이다. 이를 기각하지 못하면 두 평균에 **유의한 차이가 있다고 판단할 근거가 부족하다**. 따라서 '유의한 차이가 있다'는 '유의한 차이가 있다고 할 수 없다'로 고쳐야 한다. 기각하지 못했다는 사실이 두 평균이 같음을 입증하는 것은 아니다.

② X: $H_0 : \beta_1 = 0$을 기각하면 **A와 B의 평균에 유의한 차이**가 있다고 판단한다. B와 C의 평균 차이는 $\mu_C - \mu_B = \beta_2 - \beta_1$이므로, 두 수준을 비교하려면 $H_0 : \beta_2 - \beta_1 = 0$을 검정해야 한다.

③ X: $H_0 : \beta_2 = 0$은 **B와 C가 아니라 A와 C를 비교하는 가설**이다. 이를 기각하지 못하면 A와 C의 평균에 유의한 차이가 있다고 할 수 없으며, B와 C의 차이에 대해서는 이 검정만으로 판단할 수 없다. B와 C의 평균 차이는 $H_0 : \beta_2 - \beta_1 = 0$으로 검정해야 한다.

④ O: $\beta_2 = \mu_C - \mu_A$이므로 $H_0 : \beta_2 = 0$은 A와 C의 평균이 같다는 가설이고, $H_1 : \beta_2 \neq 0$은 두 평균이 다르다는 가설이다. 귀무가설을 기각하면 **A와 C의 평균에 유의한 차이가 있다**고 해석한다.

NoteBase