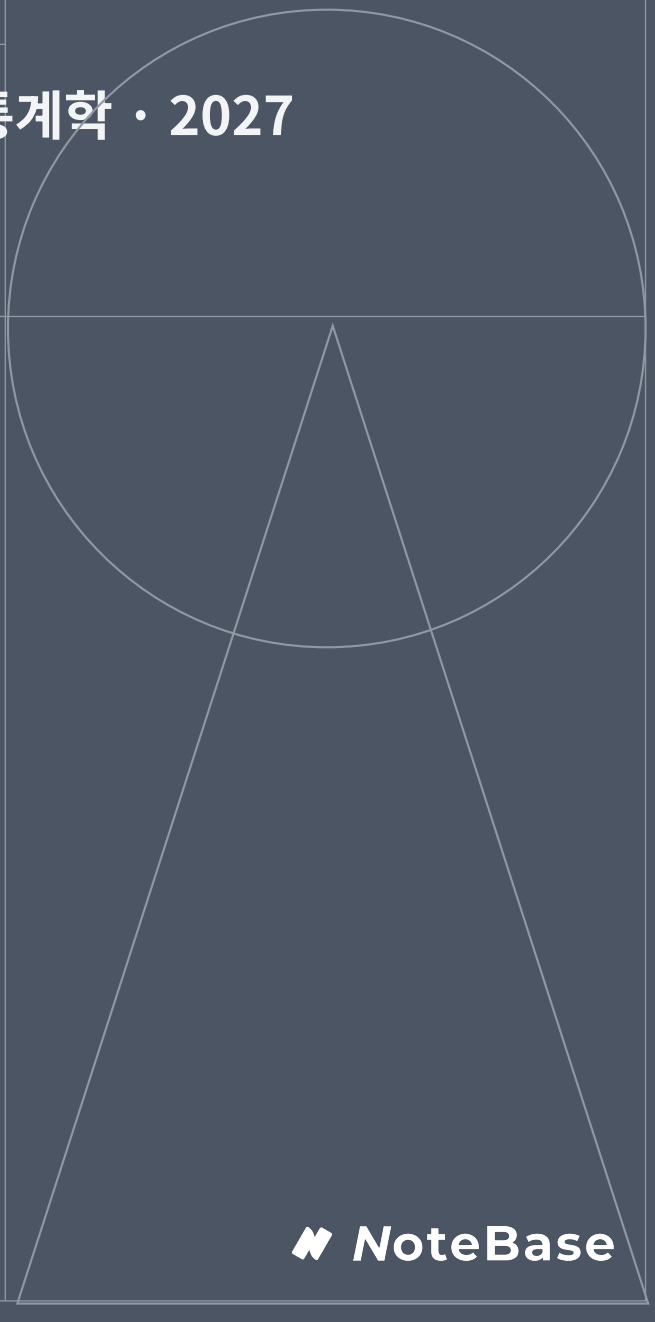
기출OX 해설집

공무원/경찰/공채(경위) · 통계학 · 2027



NoteBase

기출OX 해설집

공무원/경찰/공채(경위) · 통계학 · 2027

공무원/경찰/공채(경위) · 통계학 · 2027

문 1. 모집단으로부터 추출한 표본이 다음과 같을 때, 표본평균은?

3, 4, 5, 6, 7

① 4

② 5

③ 6

④ 7

정답 2번

간단한 해설

관측값의 합은 25이고 표본 크기는 5이다. 따라서 표본평균은 $\bar{x} = 25/5 = 5$이다.

계산 과정

① 공식과 조건: 표본평균은 **표본 관측값의 합을 표본 크기로 나눈 값**이다. 관측값을 $x_1, \ldots, x_n$, 표본 크기를 n이라 하면 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$이다. 여기서는 별도의 가중치가 없는 다섯 관측값이므로 $n = 5$이다.

② 계산: 관측값의 합은 $3 + 4 + 5 + 6 + 7 = 25$이다. 따라서 $\bar{x} = \frac{25}{5} = 5$이다. 표본평균의 단위는 관측값의 단위와 같으며, 문제에 별도의 측정 단위는 주어지지 않았다.

③ 검산: 계산된 평균 5를 기준으로 한 편차의 합은 $(3-5) + (4-5) + (5-5) + (6-5) + (7-5) = 0$이다. 표본평균은 관측값 수인 5로 나누며, 표본분산에서 사용하는 자유도 $n-1$로 나누지 않는다.

따라서 표본평균은 5이며, 이에 해당하는 선택지는 ②이다.

NoteBase

문 2. 다음 중 범주형 자료에 해당하는 것은?

① 코스피(KOSPI) 지수

② 자동차 무게

③ 비행기 속도

④ 혈액형

정답 4번

간단한 해설

혈액형은 대상을 A형·B형·AB형·O형처럼 종류로 구분한 **명목형 범주 자료**이다. 각 범주에 수량의 크기나 순서가 있는 것은 아니다.

개념 구분

범주형 자료는 관측 대상을 종류나 집단으로 구분한 자료이다. 혈액형의 A형·B형·AB형·O형은 서로 다른 범주를 나타내며, 범주 사이에 크기나 순서가 없으므로 명목형 자료이다.

코스피 지수는 주가 수준을 수치로 표현하는 양적 자료이다. 자동차 무게와 비행기 속도도 각각 측정된 양의 크기를 나타내는 양적 자료이다. 이들은 단순히 종류를 구별하는 혈액형과 다르다.

따라서 범주형 자료에 해당하는 것은 ④ 혈액형이다.

NoteBase

문 3. 다음 자료에 대해 산술평균, 중앙값, 최빈값에 대한 설명으로 옳은 것은?

-3, -2, -1, 0, 0, 0, 0, 1, 2, 3

① 산술평균, 중앙값, 최빈값이 모두 같다.

② 중앙값이 최빈값보다 크다.

③ 산술평균이 최빈값보다 작다.

④ 산술평균이 중앙값보다 크다.

정답 1번

간단한 해설

합이 0인 10개 자료의 산술평균은 0이다. 가운데 두 값도 0이고, 가장 자주 나타나는 값도 0이므로 **세 대표값이 모두 같다**.

선지별 해설

① O: 자료의 합은
aligned&(-3+3)+(-2+2)&+(-1+1)+0+0+0+0=0aligned
이고 관측값은 10개이므로 산술평균은 0/10 = 0이다. 정렬된 자료의 다섯째·여섯째 값이 모두 0이므로 중앙값도 (0 + 0)/2 = 0이다. 가장 많이 나타나는 값은 네 번 등장하는 0이다. 따라서 **세 값이 모두 0**이다.

② X: 중앙값은 0이고 최빈값도 0이다. 중앙값이 더 큰 것이 아니라 **두 값이 같다**. 중앙값은 순서상 가운데 위치로, 최빈값은 출현 횟수로 정한다는 기준의 차이에 주의한다.

③ X: 산술평균과 최빈값은 모두 0이다. 산술평균이 더 작지 않다. 음수 관측값이 있어도 같은 크기의 양수들이 상쇄하므로 평균이 음수가 되지 않는다.

④ X: 산술평균과 중앙값은 모두 0이다. 관측값이 짝수 개이므로 중앙값은 한 관측값을 임의로 택하는 것이 아니라 **가운데 두 관측값의 평균**으로 구한다.

따라서 산술평균·중앙값·최빈값이 모두 같다는 ①이 옳다.

NoteBase

문 4. 양적 자료의 모든 값에 3을 더했을 때, 변하지 않는 것은?

① 평균

② 중앙값

③ 분산

④ 최댓값

정답 3번

간단한 해설

모든 값에 3을 더하면 평균도 3 증가하여 **평균으로부터의 편차는 그대로**이다. 따라서 편차의 제곱으로 계산하는 분산은 변하지 않는다.

핵심 풀이

원래 관측값을 x_i, 평균을 $\overline{x}$라고 하면, 새 관측값은 $y_i = x_i + 3$이고 새 평균은 $\overline{y} = \overline{x} + 3$이다. 따라서 평균으로부터의 편차는 $y_i - \overline{y} = (x_i + 3) - (\overline{x} + 3) = x_i - \overline{x}$로 변하지 않는다.

분산은 편차의 제곱을 합하여 관측값 수 또는 표본분산의 경우 관측값 수에서 1을 뺀 값으로 나눈다. 모든 편차와 관측값 수가 그대로이므로, 분산이 정의되는 자료에서 분산은 변하지 않는다.

모든 값에 같은 수를 더하면 크기 순서가 유지된다. 그러므로 중앙값과 최댓값은 각각 3씩 증가하며, 평균도 3 증가한다. 이들은 분산과 달리 자료의 위치가 이동한 영향을 받는다.

따라서 변하지 않는 것은 ③ 분산이다.

NoteBase

문 5. 공평한 동전 2개를 동시에 던질 때, 적어도 한 개가 앞면일 확률은?

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

정답 3번

간단한 해설

두 동전이 모두 뒷면일 확률을 1에서 빼면 된다. $1 - (1/2)^2 = 3/4$이다.

계산 과정

① 사건과 조건: 적어도 한 개가 앞면인 사건의 여사건은 **두 동전이 모두 뒷면인 사건**이다. 공평한 동전의 앞·뒷면 확률은 각각 1/2이며, 두 동전의 결과가 독립인 통상적인 시행을 전제로 한다.

② 여사건 계산: 독립이므로 두 동전이 모두 뒷면일 확률은 $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$이다. 따라서 구하는 확률은 $1 - 1/4 = 3/4$이다. 확률은 단위가 없는 값이다.

③ 직접 검산: 두 동전을 구별하면 가능한 결과는 앞앞·앞뒤·뒤앞·뒤뒤의 네 가지이고 각각 확률이 1/4이다. 이 중 앞앞·앞뒤·뒤앞 세 가지가 조건을 만족하므로 3/4이다. 적어도 한 개에는 **두 개 모두 앞면인 경우도 포함**되며, 정확히 한 개의 확률 1/2과 다르다.

따라서 구하는 확률은 3/4이고, 이에 해당하는 선택지는 ③이다.

NoteBase

문 6. 두 사건 A와 B에 대해 $P(A^c) = 0.7$, $P(B^c) = 0.6$, $P(A \cup B) = 0.6$이다. 이때 $P(A \cap B)$는? (단, A^c는 A의 여집합)

① 0.1

② 0.2

③ 0.3

④ 0.4

정답 1번

간단한 해설

여집합 확률로부터 A와 B의 확률은 각각 0.3과 0.4이다. 합집합에서 **중복된 교집합을 한 번 빼는 덧셈법칙**을 적용하면 교집합 확률은 0.1이다.

계산 과정

여집합의 확률과 원래 사건의 확률의 합은 1이므로 $P(A) = 1 - P(A^c) = 1 - 0.7 = 0.3$,
$P(B) = 1 - P(B^c) = 1 - 0.6 = 0.4$이다.

두 사건의 확률을 더하면 교집합 부분이 두 번 포함된다. 따라서 합집합 확률은
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$이며, 이 공식에는 두 사건이 독립이라는 가정이 필요하지 않다.

구하려는 교집합 확률을 정리하면
aligned&P(A B)=P(A)+P(B)&-P(A B)=0.3+0.4-0.6=0.1aligned
이다. 이는 A와 B가 동시에 발생할 확률이며 단위는 없다.

계산 결과 0.1은 ①과 일치한다.

NoteBase

문 7. 확률변수 X의 평균이 3, 분산이 2일 때, 다음 중 가장 큰 값은?

① $E(3X)$

② $Var(2X)$

③ $E(2X+4)$

④ $Var(X+9)$

정답 3번

간단한 해설

평균은 선형적으로 변하지만, 분산에는 **곱한 상수의 제곱**이 적용되고 더한 상수는 영향을 주지 않는다. 이 규칙으로 비교하면 $E(2X+4)$의 값 10이 가장 크다.

선지별 해설

① 평균의 선형성으로 $E(3X) = 3E(X) = 3 \times 3 = 9$이다.

② 확률변수에 상수를 곱하면 분산에는 그 상수의 제곱이 곱해진다. 따라서 $\mathrm{Var}(2X) = 2^2\mathrm{Var}(X) = 4 \times 2 = 8$이다.

③ 평균의 선형성으로 $E(2X+4) = 2E(X) + 4 = 2 \times 3 + 4 = 10$이다.

④ 상수를 더하면 평균도 같은 만큼 이동하여 평균으로부터의 편차는 변하지 않는다. 따라서 $\mathrm{Var}(X+9) = \mathrm{Var}(X) = 2$이다.

각 값은 순서대로 9, 8, 10, 2이므로 가장 큰 값은 10인 ③이다.

NoteBase

문 8. 두 주식 A와 B의 기대수익률과 위험도(표준편차)가 다음과 같다.
다른 조건이 동일할 때, 투자자가 기대수익률은 높고 위험도가 낮은 주식을 선호한다면 다음 중 가장 적절한 판단은?

주식	A	B
기대수익률(%)	59	14
위험도(표준편차)	78.67	99.62

① A는 B보다 기대수익률이 더 높고, 표준편차는 더 작기 때문에 A를 선호한다.

② B는 A보다 기대수익률이 더 낮기 때문에 B를 선호한다.

③ B는 A보다 표준편차가 더 크기 때문에 투자위험이 낮다.

④ 두 주식 A와 B의 투자 선호도는 동일하다.

정답 1번

간단한 해설

A는 기대수익률이 59%로 더 높고, 표준편차는 78.67로 더 작다. 따라서 문제의 두 선호 기준을 모두 만족하는 A가 유리하다.

NoteBase

핵심 풀이

비교 기준: 문제는 다른 조건이 같을 때 **기대수익률은 높고 표준편차는 낮은 주식**을 선호한다고 정했다.
기대수익률은 평균적으로 예상하는 수익률이고, 표준편차는 수익률의 변동성을 나타내는 위험 지표이다.

자료 적용: 표에서 A의 기대수익률은 59%, B는 14%이므로 $59 > 14$이다. 표준편차는 A가 78.67, B가 99.62이므로 $78.67 < 99.62$이다. A는 수익률과 위험의 두 기준에서 모두 B보다 유리하다.

판단 정리:

① O: A는 B보다 기대수익률이 더 높고 표준편차가 더 작으므로 제시된 선호에 부합한다.

② X: 수익률이 더 낮다는 이유로 B를 선호하는 것은 제시된 선호와 반대이다.

③ X: 표준편차가 더 크면 변동성이 더 커지므로 위험이 더 낮다고 해석할 수도 없다.

④ X: 두 주식은 두 지표 모두 다르므로 선호도가 같지 않다.

이는 **문제에서 정한 조건에 따른 비교**이며 실제 수익을 보장한다는 뜻이 아니다.

따라서 기대수익률이 더 높고 표준편차가 더 작은 A를 선호한다는 ①이 적절하다.

NoteBase

문 9. 두 사건 A와 B에 대해 $P(A)=\frac{1}{3},\ P(B)=\frac{1}{4}$일 때, $\frac{P(B\mid A)}{P(A\mid B)}$의 값은?

① $\frac{1}{2}$
② $\frac{3}{4}$
③ $\frac{7}{8}$
④ $\frac{15}{16}$

정답 2번

간단한 해설

구하려는 비가 정의될 때, 두 조건부확률의 **공통 교집합 확률이 약분**되어 $\frac{P(B\mid A)}{P(A\mid B)}=\frac{P(B)}{P(A)}=\frac{3}{4}$이 된다.

계산 과정

교집합 확률을 $t=P(A\cap B)$로 둔다. 조건부확률은 교집합 확률을 조건이 되는 사건의 확률로 나눈 값이다. 주어진 $P(A)=\frac{1}{3}$, $P(B)=\frac{1}{4}$는 모두 양수이므로 $P(B\mid A)=\frac{t}{1/3}=3t$, $P(A\mid B)=\frac{t}{1/4}=4t$이다.

두 조건부확률 자체는 정의되지만, 그 비를 구하려면 **분모인 조건부확률도 0이 아니어야** 한다. 따라서 $P(A\mid B)=4t>0$, 즉 $t>0$인 경우를 전제로 계산한다. 이 조건이 원문에 명시되어 있지는 않으며, $t=0$이면 두 조건부확률이 모두 0이 되어 요구한 비는 정의되지 않는다.

비가 정의되는 경우에는 공통인 t를 약분할 수 있으므로 $\frac{P(B\mid A)}{P(A\mid B)}=\frac{3t}{4t}=\frac{3}{4}$이다. 따라서 계산값은 ②의 $\frac{3}{4}$에 해당한다.

NoteBase

문 10. 다음 상관계수(r) 중 가장 강한 상관관계를 보여주는 값은?

① $r = -1$

② $r = 0.9$

③ $r = -0.9$

④ $r = 0.1$

정답 1번

간단한 해설

상관관계의 강도는 부호가 아닌 **절댓값**으로 비교한다. $|-1| = 1$이 가장 크므로 완전한 음의 선형 상관이 가장 강하다.

선지별 해설

공통 기준: 상관계수의 부호는 **선형 관계의 방향**을, 절댓값 $|r|$은 **선형 관계의 강도**를 나타낸다. 따라서 음수라는 이유만으로 관계가 약하다고 판단하지 않는다.

같은 기준으로 비교: ①은 $|-1| = 1$, ②는 $|0.9| = 0.9$, ③은 $|-0.9| = 0.9$, ④는 $|0.1| = 0.1$이다. 절댓값 기준 순서는 $1 > 0.9 = 0.9 > 0.1$이다. ②와 ③은 방향은 반대지만 강도는 같다.

해석: $r = -1$은 한 변수가 증가할 때 다른 변수가 선형적으로 감소하는 **완전한 음의 선형 상관**을 뜻한다. 이 비교는 선형 연관성에 대한 것이며, 인과관계를 입증한다는 뜻은 아니다.

따라서 제시된 값 중 가장 강한 상관관계를 나타내는 값은 ① $r = -1$이다.

NoteBase

문 11. 확률변수 X의 확률질량함수가 다음과 같을 때, $P(X \geq 2)$의 값은?
P(X=x)=cases k(x+1), & x=0,1,2,3 \ 0, & 그 외의 경우 cases

① 0.4

② 0.5

③ 0.6

④ 0.7

정답 4번

간단한 해설

전체 확률의 합이 1이므로 $10k = 1$에서 $k = 0.1$이다. 따라서 P(X≥2)=3k+4k=0.7이다.

계산 과정

① 정규화 조건: 확률질량함수는 가능한 모든 값에서 확률이 음이 아니고 그 합이 1이어야 한다. 지지집합은 $\{0, 1, 2, 3\}$이므로 $k(0 + 1) + k(1 + 1) + k(2 + 1) + k(3 + 1) = 1$이다.

② 상수 계산: $k(1 + 2 + 3 + 4) = 10k = 1$이므로 $k = 1/10 = 0.1$이다. 따라서 각 확률은 차례대로 0.1, 0.2, 0.3, 0.4이고 모두 음이 아니며 합은 1이다.

③ 사건 확률: **2 이상에는 2도 포함**되므로 $X = 2$와 $X = 3$의 확률을 더한다.
aligned&P(X≥2)=P(X=2)&+P(X=3)=3k+4k=7k=0.7aligned
이다. 확률은 단위가 없는 값이다.

④ 검산: 여사건은 $X < 2$이므로 $1 - P(X < 2) = 1 - (0.1 + 0.2) = 0.7$로 같은 결과를 얻는다.

따라서 구하는 확률은 0.7이며 선택지 ④이다.

NoteBase

문 12. 확률변수 X의 확률분포가 다음과 같을 때, X의 분산 $Var(X)$은?

x	0	1	2
$P(X=x)$	0.2	0.5	0.3

① 0.29

② 0.39

③ 0.49

④ 0.59

정답 3번

간단한 해설

표로부터 $E(X) = 1.1$, $E(X^2) = 1.7$이다. **제곱의 기댓값에서 기댓값의 제곱을 빼면** 분산은 0.49이다.

계산 과정

표의 위 행에는 X의 가능한 값 0, 1, 2가 있고, 아래 행의 대응 확률은 0.2, 0.5, 0.3이다. 모든 확률이 음이 아니며 합이 1이므로 확률분포 조건을 만족한다.

기댓값은 각 값에 해당 확률을 곱하여 더한 가중평균이다. 따라서 $E(X) = 0 \times 0.2 + 1 \times 0.5 + 2 \times 0.3 = 1.1$이다.

제곱의 기댓값은 값을 먼저 제곱한 뒤 확률로 가중한다. $E(X^2) = 0^2 \times 0.2 + 1^2 \times 0.5 + 2^2 \times 0.3 = 1.7$이다.

분산은 평균으로부터의 제곱 편차의 기댓값이며, $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$로 계산할 수 있다. 따라서 $\mathrm{Var}(X) = 1.7 - 1.1^2 = 1.7 - 1.21 = 0.49$이다. 별도의 측정 단위는 제시되어 있지 않다.

구한 분산 0.49는 ③과 일치한다.

NoteBase

문 13. 확률변수 X가 이항분포 $Bin(20, 1-p)$를 따른다고 하자. 확률변수 X의 기댓값과 분산이 각각 $E(X)=10$, $Var(X)=5$일 때, p의 값은?

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

정답 3번

간단한 해설

성공확률은 $1-p$이므로 $20(1-p)=10$이다. 따라서 $p=1/2$이고 분산도 $20(1/2)(1/2)=5$로 일치한다.

계산 과정

① 모수 구별: 이항분포 $Bin(n, q)$의 기댓값은 nq, 분산은 $nq(1-q)$이다. 이 문제에서는 시행 횟수가 $n=20$, **성공확률이** $q=1-p$이다. 따라서 $E(X)=20(1-p)$, $\mathrm{Var}(X)=20(1-p)p$이다.

② 기댓값 이용: $20(1-p)=10$이므로 $1-p=1/2$이고, $p=1-1/2=1/2$이다. 구하는 모수는 확률이므로 단위가 없으며 $0\leq p\leq 1$을 만족한다.

③ 분산으로 검산: $p=1/2$를 대입하면 $20(1-1/2)(1/2)=20/4=5$로 주어진 분산과 일치한다. 또는 $\mathrm{Var}(X)/E(X)=p=5/10=1/2$로 직접 확인할 수 있다.

따라서 p=12이며 이에 해당하는 선택지는 ③이다.

NoteBase

문 14. 서로 독립인 두 확률변수 X와 Y의 분포가 각각 $X \sim N(2, 3)$, $Y \sim N(5, 4)$일 때, $X - Y$의 분포는?

① $N(3, 7)$

② $N(7, 3)$

③ $N(-3, 1)$

④ $N(-3, 7)$

정답 4번

간단한 해설

독립인 정규확률변수의 차도 정규분포를 따른다. 평균은 빼지만 **분산은 더하므로** 평균 -3, 분산 7인 정규분포이다.

계산 과정

정규분포 표기를 $N(\mu, \sigma^2)$로 읽으면, X의 평균과 분산은 2와 3이고 Y의 평균과 분산은 5와 4이다. 두 변수가 독립이므로 이들의 선형결합인 차도 정규분포를 따른다.

평균은 선형성이 성립하므로 $E(X - Y) = E(X) - E(Y) = 2 - 5 = -3$이다.

차의 분산은 $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y)$이다. 독립이므로 공분산은 0이고, Y에 곱한 -1이 제곱되어 분산은 더해진다. 따라서 $\mathrm{Var}(X - Y) = 3 + 4 = 7$이다.

평균 -3, 분산 7인 정규분포이므로 $X - Y \sim N(-3, 7)$이며 ④와 일치한다.

NoteBase

문 15. 두 변수 (x, y)의 자료가 다음과 같이 주어졌을 때, 표본상관계수 r은?

x	-2	-1	1	2
y	2	1	-1	-2

① −1

② 0

③ 0.5

④ 1

정답 1번

간단한 해설

모든 관측쌍이 $y = -x$를 만족하므로 완전한 음의 선형 상관이다. 공식으로도 $r = -10/\sqrt{10 \times 10} = -1$이다.

계산 과정

① 공식과 자료: 표본상관계수는 $r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$이다. 표의 관측쌍은 $(-2, 2), (-1, 1), (1, -1), (2, -2)$이며, 두 변수 모두 값이 변하므로 분모가 0이 아니다.

② 평균과 편차합: $\bar{x} = (-2 - 1 + 1 + 2)/4 = 0$, $\bar{y} = (2 + 1 - 1 - 2)/4 = 0$이다. 따라서 교차곱합은 $(-2)(2) + (-1)(1) + (1)(-1) + (2)(-2) = -10$이다. 두 변수의 제곱편차합은 각각 $4 + 1 + 1 + 4 = 10$이다.

③ 대입: $r = \frac{-10}{\sqrt{10 \times 10}} = \frac{-10}{10} = -1$이다. 상관계수는 단위가 없는 값이다. 표본 공분산과 표본 표준편차를 사용해도 공통된 자유도 인자가 상쇄되어 같은 값이 나온다.

④ 관계로 검산: 네 관측쌍이 모두 $y = -x$를 만족한다. 기울기가 음수인 한 직선 위에 있으므로 **완전한 음의 선형 상관**이며, 양의 완전 상관을 뜻하는 1과 구별한다.

따라서 표본상관계수는 −1로 선택지 ①이다.

NoteBase

문 16. 확률변수 X의 평균과 분산이 각각 2와 8, 확률변수 Y의 평균과 분산이 각각 3과 18, 그리고 두 확률변수의 공분산이 -6일 때, X와 Y의 상관계수는?

① -1

② -0.5

③ 0.5

④ 1

정답 2번

간단한 해설

상관계수는 **공분산을 두 표준편차의 곱으로 나눈 값**이다. $\frac{-6}{\sqrt{8}\sqrt{18}} = -0.5$이며 공분산의 음수 부호가 유지된다.

계산 과정

상관계수는 공분산을 두 표준편차의 곱으로 나눈 값이다. $\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$이며, 두 분산이 모두 양수이므로 분모가 0이 아니고 공식이 적용된다.

분산은 각각 8과 18이므로 표준편차는 $\sigma_X = \sqrt{8}$, $\sigma_Y = \sqrt{18}$이다. 이들의 곱은 $\sqrt{8}\sqrt{18} = \sqrt{144} = 12$이다. 공분산이 이미 주어져 있으므로 평균 2와 3을 다시 대입할 필요는 없다.

따라서 $\rho_{XY} = \frac{-6}{12} = -0.5$이다. 표준편차의 곱은 양수이므로 상관계수의 부호는 공분산과 같은 음수이며, 상관계수에는 측정 단위가 없다.

상관계수 -0.5는 ②와 일치한다.

NoteBase

문 17. 확률변수 X가 구간 [2,8]에서 균일분포 U(2,8)을 따를 때, P(3<X<6)의 값은?

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

정답 3번

간단한 해설

균일분포의 구간 확률은 길이의 비율이다. 따라서 $P(3 < X < 6) = (6 - 3)/(8 - 2) = 1/2$이다.

계산 과정

① 분포와 공식: 연속균일분포 $U(a, b)$는 구간 안에서 밀도가 $1/(b - a)$로 일정하다. 따라서 구간 안의 $c < d$에 대해 $P(c < X < d) = (d - c)/(b - a)$이다. 여기서는 $a = 2, b = 8, c = 3, d = 6$이며, 관심 구간 전체가 지지구간 안에 있다.

② 대입: 전체 구간의 길이는 $8 - 2 = 6$, 관심 구간의 길이는 $6 - 3 = 3$이다. 따라서 P(3<X<6)=(3)/(6)=12이다. 길이 단위가 약분되므로 확률은 단위가 없다.

③ 검산과 경계: 밀도 1/6을 관심 구간에서 적분해도 ∫ _3^616 dx=3/6=1/2이다. 연속분포에서는 한 점의 확률이 0이므로 끝점 포함 여부가 이 확률을 바꾸지 않는다. **정수의 개수가 아니라 구간 길이**를 비교해야 한다.

따라서 구하는 확률은 12로 선택지 ③이다.

NoteBase

문 18. 두 확률변수 X와 Y에 대해 $Var(3X+2Y)=16$이고, $Var(3X-2Y)=9$일 때, $3X$와 $2Y$의 공분산 $Cov(3X,2Y)$의 값은?

① $\frac{3}{4}$
② $\frac{5}{4}$
③ $\frac{3}{2}$
④ $\frac{7}{4}$

정답 4번

간단한 해설

두 확률변수의 **합과 차의 분산 차이는 공분산의 4배**이다. 따라서 $\mathrm{Cov}(3X,2Y)=\frac{16-9}{4}=\frac{7}{4}$이다.

계산 과정

구하려는 공분산의 두 대상을 $U=3X$, $V=2Y$로 놓는다. 주어진 조건은 $\mathrm{Var}(U+V)=16$, $\mathrm{Var}(U-V)=9$이다. 독립이라는 조건은 없으므로 공분산 항을 유지한다.

합과 차의 분산은 각각 $\mathrm{Var}(U+V)=\mathrm{Var}(U)+\mathrm{Var}(V)+2\mathrm{Cov}(U,V)$,
$\mathrm{Var}(U-V)=\mathrm{Var}(U)+\mathrm{Var}(V)-2\mathrm{Cov}(U,V)$이다.

두 식을 빼면 개별 분산 항이 없어져 $16-9=4\mathrm{Cov}(U,V)$가 된다. 따라서
$\mathrm{Cov}(3X,2Y)=\mathrm{Cov}(U,V)=\frac{16-9}{4}=\frac{7}{4}$이다. 이미 3X와 2Y의 공분산을 구했으므로 다시 6으로 나누지 않는다.

구한 값 $\frac{7}{4}$는 ④와 일치한다.

NoteBase

문 19. 다음 중 이산형 확률분포에 해당하지 않는 것은?

① 이항분포

② 포아송분포

③ 기하분포

④ 감마분포

정답 4번

간단한 해설

이항·포아송·기하분포는 정숫값을 갖는 이산형 분포이다. **감마분포는 양의 실수 구간에 밀도를 갖는 연속형 분포**이다.

핵심 풀이

구분 기준: 이산형 분포는 정수처럼 **셀 수 있는 값에 확률질량을 부여**한다. 연속형 분포는 구간에서 확률밀도를 적분하여 확률을 구하며, 특정 한 점의 확률은 0이다.

이산형 세 분포: 이항분포는 정해진 횟수의 독립적인 베르누이 시행에서 성공 횟수를 세므로 0, 1, …, n의 값을 갖는다. 포아송분포는 발생 횟수를 나타내며 0, 1, 2, …의 값을 갖는다. 기하분포는 첫 성공까지의 시행 횟수 또는 첫 성공 전 실패 횟수를 세므로 정의에 따라 1 또는 0부터 시작하는 정숫값을 갖는다.

연속형 분포: 감마분포는 양의 실수 구간에 밀도를 갖는 **연속형 분포**이다. 기하분포가 첫 성공까지의 횟수를 세는 것과 달리, 감마분포는 연속적인 대기시간 등을 모형화하는 데 사용된다. 따라서 분포가 나타내는 값이 횟수인지 연속량인지 구별한다.

따라서 이산형 확률분포에 해당하지 않는 것은 ④ 감마분포이다.

NoteBase

문 20. 공평한 동전을 3회 던져 앞면이 나온 횟수를 X라 할 때, 확률변수 X는 이항분포 $Bin(3, 1/2)$를 따르고 다음과 같은 확률함수를 갖는다. 이때 앞면이 1회 나올 확률은?

$f(x) = \binom{3}{x}\left(\frac{1}{2}\right)^3, \quad x = 0, 1, 2, 3.$

① $\frac{1}{4}$

② $\frac{3}{8}$

③ $\frac{1}{2}$

④ $\frac{5}{8}$

정답 2번

간단한 해설

앞면이 한 번 나오는 위치는 **세 가지**이고 각 배열의 확률은 1/8이다. 따라서 구할 확률은 $3 \times \frac{1}{8} = \frac{3}{8}$이다.

계산 과정

X는 세 번의 시행에서 앞면이 나온 횟수이므로 구할 확률은 $P(X = 1)$이다. 주어진 이항분포에서는 시행 횟수가 3, 각 시행의 앞면 확률이 1/2이고 시행들은 독립이다.

앞면이 한 번 나오는 배열은 앞·뒤·뒤, 뒤·앞·뒤, 뒤·뒤·앞의 세 가지이다. 각 배열의 확률은 $\left(\frac{1}{2}\right)^3 = \frac{1}{8}$이다.

세 배열은 서로 배반이므로 확률을 더한다. 주어진 확률함수에 1을 넣어도 같은 결과를 얻으며,
$P(X = 1) = \binom{3}{1}\left(\frac{1}{2}\right)^3 = 3 \times \frac{1}{8} = \frac{3}{8}$이다.

앞면이 정확히 한 번 나올 확률 $\frac{3}{8}$은 ②와 일치한다.

NoteBase

문 21. 평균이 50이고 표준편차가 12인 모집단에서 크기 36인 확률표본을 추출할 때, 표본평균 $\overline{X}$의 평균과 표준편차는?

① 평균 50, 표준편차 12

② 평균 50, 표준편차 2

③ 평균 36, 표준편차 2

④ 평균 36, 표준편차 12

정답 2번

간단한 해설

표본평균의 평균은 모평균 50과 같다. 표본평균의 표준편차는 $12/\sqrt{36} = 2$이다.

계산 과정

① 공식과 조건: 같은 모집단에서 독립적으로 추출한 크기 n의 확률표본에 대해 E(X)=μ, Var(X)=σ^2/n이다. 따라서 표본평균의 표준편차, 즉 **표준오차**는 σ/ n이다. 이 문제는 별도의 유한모집단 비복원추출 조건이 없는 통상적인 확률표본을 전제로 한다.

② 대입: $\mu = 50, \sigma = 12, n = 36$이므로 E(X)=50, SD(X)=12/√(36)=12/6=2이다. 평균과 표준오차의 단위는 원자료의 단위와 같으며, 구체적 단위는 제시되지 않았다.

③ 검산과 구별: 표본평균의 분산은 $12^2/36 = 144/36 = 4$이고 그 제곱근은 2이다. 표본크기 36은 평균이 아니며, 모집단의 표준편차 12를 표본평균의 표준편차로 그대로 쓰면 안 된다. 이 평균·분산 공식 자체는 모집단의 정규성을 요구하지 않는다.

따라서 표본평균의 평균은 50, 표준편차는 2이므로 선택지 ②이다.

NoteBase

문 22. 연속형 확률변수 X가 다음과 같은 확률밀도함수를 가질 때, 이 함수가 확률밀도함수를 만족하기 위한 상수 c의 값은?
f(x)=casescx^2,&0<x<3\\0,&그 외의 경우cases

① $\frac{1}{3}$
② $\frac{1}{6}$
③ $\frac{1}{9}$
④ $\frac{1}{12}$

정답 3번

간단한 해설

확률밀도함수는 **전체 적분값이 1**이어야 한다. $\int_0^3 cx^2 dx = 9c = 1$이므로 $c = \frac{1}{9}$이다.

계산 과정

확률밀도함수는 모든 x에서 음이 아니고 전체 실수 구간의 적분값이 1이어야 한다. 주어진 함수는 $0 < x < 3$ 밖에서 0이므로 이 구간의 적분만 계산한다.

정규화 조건은 $\int_0^3 cx^2 dx = 1$이다. 적분하면 $\int_0^3 cx^2 dx = c\left[\frac{x^3}{3}\right]_0^3 = c\frac{27}{3} = 9c$이다.

따라서 $9c = 1$에서 $c = \frac{1}{9}$이다. 이 값은 양수이므로 cx^2도 해당 구간에서 음이 아니며, 적분값 조건과 비음수 조건을 모두 만족한다.

상수 $c = \frac{1}{9}$는 ③과 일치한다.

NoteBase

문 23. 분포에 대한 다음 설명 중 옳은 것은?

① 포아송분포는 평균과 분산이 같다.

② 초기하분포는 연속형 분포이다.

③ 지수분포는 평균과 분산이 같다.

④ 카이제곱 분포의 분산은 평균의 제곱이다.

정답 1번

간단한 해설

포아송분포는 평균과 분산이 모두 λ이다. 지수분포는 분산이 평균의 제곱, 카이제곱분포는 분산이 평균의 2배이며, 초기하분포는 이산형이다.

선지별 해설

① O: $X \sim \text{Poisson}(\lambda)$이면 $E(X) = \lambda$, $\text{Var}(X) = \lambda$이다. 따라서 **평균과 분산의 값이 같다**. 표준편차는 $\sqrt{\lambda}$이므로 분산과 혼동하지 않는다.

② X: 초기하분포는 유한 모집단에서 비복원추출한 표본에 특정 범주의 대상이 몇 개 포함되는지를 나타낸다. 가능한 값이 정수이므로 **이산형 분포**이다. 연속적인 측정값을 나타내는 분포가 아니다.

③ X: 지수분포의 율 모수를 $\lambda > 0$라 하면 평균은 $1/\lambda$, 분산은 $1/\lambda^2$이다. 즉 **분산은 평균의 제곱**이다. $\lambda = 1$에서는 우연히 두 수치가 같지만, 지수분포 전체의 일반적 성질로 평균과 분산이 같다고 할 수는 없다.

④ X: 자유도가 k인 카이제곱분포는 평균이 k, 분산이 $2k$이다. 따라서 일반적인 관계는 **분산이 평균의 2배**라는 것이다. $k = 2$에서 평균의 제곱과 수치가 일치하더라도 일반 공식은 k^2가 아니다.

따라서 분포의 일반적인 성질로 옳은 설명은 ①이다.

NoteBase

문 24. 평균이 40이고 분산이 100이며 정규분포를 따르지 않는 모집단에서 표본크기 100인 확률표본을 추출한다. 이때 표본평균 $\overline{X}$의 분포에 관한 설명으로 가장 적절한 것은?

① $\overline{X}$는 평균 4, 분산 1인 정규분포로 근사할 수 있다.

② $\overline{X}$는 평균 40, 분산 100인 정규분포로 근사할 수 있다.

③ $\overline{X}$는 평균 40, 분산 1인 정규분포로 근사할 수 있다.

④ $\overline{X}$는 평균 4, 분산 100인 정규분포로 근사할 수 있다.

정답 3번

간단한 해설

표본평균의 평균은 40이고, **분산은 모집단 분산을 표본크기로 나눈 값**이므로 $\frac{100}{100} = 1$이다. 큰 표본에서 중심극한정리에 따른 정규근사를 적용한다.

계산 과정

확률표본을 같은 모집단에서 독립적으로 추출한 표본으로 읽는다. 모집단 평균은 $\mu = 40$, 분산은 $\sigma^2 = 100$, 표본크기는 $n = 100$이다.

표본평균의 평균은 모집단 평균과 같으므로 $E(\overline{X}) = \mu = 40$이다. 표본크기가 커져도 평균을 4로 나누어 줄이는 것은 아니다.

독립 표본의 평균을 내면 분산은 표본크기로 나뉜다. 따라서 $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{100}{100} = 1$이다. 개별 관측값의 분산 100과 표본평균의 분산 1을 구별한다.

모집단이 정규분포가 아니더라도 유한한 분산을 갖는 독립·동일분포 표본에서는 중심극한정리에 따라 표본평균의 분포를 큰 표본에서 정규분포로 근사한다. 이 문항은 표본크기 100을 큰 표본으로 보는 통상적 근사를 적용한다. 다만 정확히 정규분포가 되는 것은 아니며 근사의 정확도는 모집단 모양에도 영향을 받는다.

따라서 가장 적절한 설명은 표본평균을 평균 40, 분산 1인 정규분포로 근사한다는 ③이다.

NoteBase

문 25. 표준편차가 10인 정규분포에서 크기 100인 표본을 추출한 결과 표본평균이 52였다. 이때 모평균 μ의 95% 신뢰구간은? (단, $Z_{0.025} = 1.96$)

① (50.04, 53.96)

② (50.36, 53.64)

③ (42.20, 61.80)

④ (51.02, 52.98)

정답 1번

간단한 해설

표준오차는 $10/\sqrt{100} = 1$이다. 따라서 95% 신뢰구간은 52 ± 1.96, 즉 $(50.04, 53.96)$이다.

계산 과정

① 공식과 조건: 정규모집단에서 독립적인 확률표본을 추출하고 모표준편차 σ를 알 때, 모평균의 양측 95% 신뢰구간은 x± z_0.025(σ)/(n)이다. 여기서 x=52,σ=10,n=100,z_0.025=1.96이다.

② 표준오차와 오차한계: 표준오차는 $10/\sqrt{100} = 10/10 = 1$이다. 따라서 **오차한계**는 $1.96 \times 1 = 1.96$이다. 모표준편차 10을 그대로 오차한계 계산에 넣지 않고 표본평균의 표준오차를 사용한다.

③ 양 끝점: 하한은 $52 - 1.96 = 50.04$, 상한은 $52 + 1.96 = 53.96$이다. 따라서 신뢰구간은 $(50.04, 53.96)$이며, 두 끝점의 단위는 원자료와 같다. 구체적인 측정 단위는 주어지지 않았다.

④ 검산: 두 끝점의 중심은 $(50.04 + 53.96)/2 = 52$, 반너비는 $(53.96 - 50.04)/2 = 1.96$으로 일치한다. 95%는 같은 방식으로 표본추출과 구간 구성을 반복할 때 구간들이 모평균을 포함하는 장기적인 비율을 뜻한다.

따라서 계산된 구간과 일치하는 선택지는 ①이다.

NoteBase

문 26. 다음 표는 초등학생 140명을 대상으로 7가지 무지개 색 중 선호하는 색에 대한 빈도표이다. “7가지 색의 선호도가 같다”는 귀무가설 하에 적합도 검정에 필요한 각 색의 추정된 기대돗수의 값은?

색	빨강	주황	노랑	초록	파랑	남색	보라
관측돗수	18	18	22	18	20	22	22

① 18

② 20

③ 22

④ 24

정답 2번

간단한 해설

귀무가설 아래에서 **각 색의 선택 확률은 모두 1/7**이다. 따라서 각 기대도수는 $140 \times \frac{1}{7} = 20$명이다.

계산 과정

표에서 빨강·주황·노랑·초록·파랑·남색·보라의 관측도수는 각각 18, 18, 22, 18, 20, 22, 22명이다. 합계는 140명이며, 이는 조사된 전체 학생 수와 일치한다.

귀무가설은 일곱 색의 선택 확률이 모두 같다는 것이다. 따라서 각 색의 귀무가설상 확률은 $p_i = \frac{1}{7}$이며, 실제 관측비율을 그대로 기대확률로 쓰는 것이 아니다.

각 범주의 기대도수는 전체 표본크기에 귀무가설상 확률을 곱한 값이다. 따라서 모든 색에 대해 $E_i = np_i = 140 \times \frac{1}{7} = 20$명이다. 일곱 기대도수의 합도 140명이다.

각 색의 기대도수 20은 ②와 일치한다.

NoteBase

문 27. 전국 고등학교 3학년 수험생들의 수학능력시험에서 지역별(6개 지역: 서울·경기도, 충청도, 전라도, 강원도, 경상도, 제주도)로 성적(수능점수 평균)에 차이가 존재하는지 검정하고자 한다. 이를 위한 방법으로 옳은 것은?

① 이표본 t-검정

② 분산분석

③ 로지스틱 회귀분석

④ 적합도 검정

정답 2번

간단한 해설

지역이라는 하나의 요인으로 구분된 여섯 집단의 점수 평균을 비교하므로 **일원분산분석**을 사용한다.

핵심 풀이

분석 대상: 반응변수는 수능점수라는 양적 변수이고, 집단을 나누는 요인은 지역 하나이며 수준은 여섯 개이다. 따라서 **여러 집단의 모평균이 같은지**를 한꺼번에 검정하는 일원분산분석이 맞다.

검정 원리: 귀무가설은 $H_0 : \mu_1 = \mu_2 = \cdots = \mu_6$이다. 집단 간 변동을 집단 내 변동과 비교하는 F 통계량을 사용한다. 귀무가설을 기각하면 적어도 한 집단의 모평균이 다르다는 뜻이며, 모든 집단이 서로 다르다는 뜻은 아니다. 어느 집단끼리 다른지는 별도 비교가 필요하다.

다른 방법과 구별: 이표본 t-검정은 두 집단의 평균 비교가 기본 목적이다. 로지스틱 회귀분석은 범주형 반응의 확률을 모형화하므로 연속형 점수의 여섯 집단 평균 비교에 직접 대응하지 않는다. 적합도 검정은 관측된 분포가 가정한 분포와 맞는지 확인하는 것이므로 이 문제의 목적과 다르다.

적용 조건: 고전적인 일원분산분석은 관측의 독립성, 각 집단 오차의 정규성 및 등분산성을 전제로 한다. 실제 분석에서는 이 조건과 표본설계를 점검해야 하지만, 문제에서 묻는 기본 검정 방법은 분산분석이다.

따라서 적절한 방법은 ② 분산분석이다.

NoteBase

문 28. 투입된 비료의 양(단위: kg)과 농작물 수확량(단위: kg)과의 관계를 확인하기 위해 아래 단순선형회귀모형

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \cdots, n$

을 적합한 결과가 다음과 같다. (ㄱ)의 값은? (단, ε_i들은 서로 독립이고 ε_i들이 각각 정규분포 $N(0, \sigma^2)$을 따른다.)

변수	추정값	표준오차	t-값
절편(β_0)	96	(ㄱ)	6
기울기(β_1)	8	2	

① 4

② 8

③ 16

④ 32

정답 3번

간단한 해설

절편의 t값은 **절편 추정값을 그 표준오차로 나눈 값**이다. 따라서 빈칸의 표준오차는 $\frac{96}{6} = 16$이다.

계산 과정

표에서 (ㄱ)은 절편 행의 표준오차이다. 같은 행의 절편 추정값은 96, t값은 6이다. 기울기 행의 추정값 8과 표준오차 2를 절편의 계산에 섞지 않는다.

회귀계수 표의 통상적인 t값은 해당 계수가 0이라는 귀무가설을 검정하는 값이다. 절편에 대해서는 t=β_0-0SE(β_0)를 쓴다. 제시된 독립·정규·등분산 오차 조건은 이러한 회귀 추론의 모형 조건이다.

절편 행의 값을 넣으면 6=96SE(β_0)이므로 SE(β_0)=(96)/(6)=16이다. 절편과 그 표준오차의 단위는 반응변수인 수확량과 같은 kg이고, t값은 단위가 없다.

따라서 (ㄱ)의 값은 16으로 ③과 일치한다.

NoteBase

문 29. 두 사건 A와 B에 대해 P(A)=0.4, P(B)=0.5, P(A ㅣ B)=0.6일 때, 다음 중 옳은 것은?

① A와 B는 독립이다.

② P(A ∩ B)=0.24

③ P(A ∪ B)=0.74

④ P(B ㅣ A)=0.75

정답 4번

간단한 해설

교집합 확률은 $P(A \cap B) = 0.6 \times 0.5 = 0.3$이다. 이를 $P(A) = 0.4$로 나누면 $P(B \mid A) = 0.75$이다.

선지별 해설

① X: $P(B) > 0$일 때 독립이라면 $P(A \mid B) = P(A)$이어야 한다. 여기서는 0.6"22600.4이므로 **독립이 아니다**. 실제 교집합 확률 0.3도 독립일 때의 곱 $0.4 \times 0.5 = 0.2$와 다르다.

② X: 조건부확률의 정의에서 $P(A \cap B) = P(A \mid B)P(B)$이다. 따라서 $0.6 \times 0.5 = 0.3$이다. 조건 사건은 B이므로 $P(A)$를 곱해 0.24를 만드는 것은 잘못이다.

③ X: 합집합은 중복을 한 번 빼야 하므로 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$이다. 따라서 $0.4 + 0.5 - 0.3 = 0.6$이지 0.74가 아니다.

④ O: 조건이 A로 바뀌었으므로 교집합 확률을 $P(A) = 0.4$로 나눈다. $P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.3}{0.4} = 0.75$이다. **조건부확률은 조건 방향에 따라 달라질 수 있다**.

따라서 주어진 조건과 일치하는 설명은 ④이다.

NoteBase

문 30. 회귀분석에서 모형이 자료를 얼마나 잘 설명하는지에 대한 측도로 가장 적합한 것은?

① 절편

② 기울기

③ 결정계수

④ 표준오차

정답 3번

간단한 해설

결정계수 R^2는 종속변수의 전체 변동 중 회귀모형이 설명하는 비율을 나타내는 대표적인 측도이다.

핵심 풀이

핵심 측도: 결정계수 R^2는 통상적인 절편 포함 최소제곱 회귀에서 **종속변수의 전체 변동 중 모형이 설명하는 비율**이다. 전체 제곱합을 SST, 잔차제곱합을 SSE라 하면 $R^2 = 1 - SSE/SST$이다. 종속변수의 변동이 있어 $SST > 0$인 경우를 전제로 한다.

다른 개념과 구별: 절편은 설명변수가 0일 때의 모형 예측값과 관련되고, 기울기는 설명변수 변화에 따른 종속변수의 평균적인 변화를 나타낸다. 표준오차는 추정치의 불확실성이나 잔차의 산포를 나타내는 데 쓰인다. 이들은 모형이 설명하는 전체 변동의 비율 그 자체가 아니다.

해석의 한계: 같은 종속변수와 자료에 대해 결정계수가 클수록 표본 내 설명 비율이 높다. 그러나 높은 결정계수만으로 **인과관계나 새로운 자료에 대한 예측 성능**이 보장되지는 않는다. 모형의 가정과 잔차, 과적합 여부 등은 따로 검토해야 한다.

따라서 제시된 선택지 중 가장 적합한 측도는 ③ 결정계수이다.

NoteBase

문 31. 두 변수의 관계에 대한 다음 설명 중 옳지 않은 것은?

① 두 변수가 서로 독립이면 공분산은 0이다.

② 상관계수는 −1과 1사이의 값을 갖는다.

③ 공분산이 0이면 상관계수도 0이다.

④ 두 변수의 상관계수가 0이면 서로 독립이다.

정답 4번

간단한 해설

상관계수 0은 **선형 상관이 없다는 뜻일 뿐 독립을 보장하지 않는다**. 비선형 의존관계가 있어도 공분산과 상관계수는 0일 수 있다.

선지별 해설

이 문항은 공분산과 상관계수가 정의되는 유한한 분산, 양의 표준편차를 갖는 변수를 전제로 해석한다.

① O: 독립이면 $E(XY) = E(X)E(Y)$이다. 따라서 $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$이다.

② O: 상관계수는 공분산을 표준편차의 곱으로 표준화한 값이며 $-1 \le \rho \le 1$이다. -1과 1도 포함한다.

③ O: $\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$에서 분모가 양수이므로 공분산이 0이면 상관계수도 0이다. 표준편차가 0인 상수변수에서는 상관계수가 정의되지 않는다는 전제에 유의한다.

④ X: 상관계수 0은 선형 상관이 없다는 뜻이지 모든 의존관계가 없다는 뜻은 아니다. 예를 들어 X가 -1, 0, 1을 같은 확률로 갖고 $Y = X^2$이면 $E(X) = E(XY) = 0$이어서 공분산과 상관계수는 0이다. 하지만 X를 알면 Y가 결정되고, X=0일 때 Y=0이 반드시 성립하므로 서로 독립이 아니다.

따라서 무상관을 독립으로 단정한 ④가 옳지 않다.

NoteBase

문 32. 다음 중 연속형 자료를 요약하기 위한 시각화 방법으로 적당하지 않은 것은?

① 히스토그램(histogram)

② 상자 그림(box plot)

③ 줄기잎 그림(stem and leaf plot)

④ 막대그래프(bar graph)

정답 4번

간단한 해설

히스토그램·상자 그림·줄기잎 그림은 수치 자료의 분포를 요약할 수 있다. **막대그래프는 주로 범주별 도수나 비율을 비교**하므로 이 문항의 대상과 구별된다.

핵심 풀이

구분 기준: 연속형 자료를 요약할 때는 수치의 위치·간격·분포를 나타내는 방법을 사용한다. 막대그래프는 기본적으로 **범주별 도수나 비율을 비교**하는 방법이므로, 연속형 자료의 분포를 직접 요약하는 대표 방법과 구별한다.

적절한 세 방법: 히스토그램은 수치 자료를 계급구간으로 나누어 분포 형태를 나타낸다.

상자 그림은 중앙값·사분위수·퍼짐과 잠재적인 이상치를 요약한다.

줄기잎 그림은 관측값을 자릿수에 따라 나누어 배치하여 개별 값과 분포를 함께 파악할 수 있게 한다.

히스토그램과 막대그래프: 히스토그램의 가로축 구간은 연속된 수치 범위를 뜻하고, 막대그래프의 막대들은 서로 다른 범주를 뜻한다. 특히 히스토그램의 계급 폭이 다르면 단순 도수 높이가 아니라 **막대 넓이가 도수 또는 상대도수에 비례**하도록 해야 한다. 연속값을 범주화하거나 집단별 평균을 막대로 표시할 수는 있지만, 이는 원래 연속형 분포를 직접 요약하는 문제와 다르다.

따라서 이 문항에서 적당하지 않은 방법은 ④ 막대그래프이다.

NoteBase

문 33. 확률변수 X가 정규분포 $N(\mu, \sigma^2)$를 따를 때, 다음 중 옳지 않은 것은?

① 확률변수 $Z = \frac{X-\mu}{\sigma}$는 표준정규분포를 따른다.

② $P(X > \mu) = 0.5$

③ 중앙값은 μ이다.

④ 제3사분위수 $Q_3 = \mu + 2\sigma$이다.

정답 4번

간단한 해설

제3사분위수는 **누적확률이 0.75인 값**이다. 정규분포에서는 $Q_3 \approx \mu + 0.6745\sigma$이며, 평균에 표준편차의 두 배를 더한 값이 아니다.

선지별 해설

통상적인 정규분포는 표준편차 $\sigma > 0$이며, 평균 μ를 중심으로 대칭인 연속분포이다.

① O: 정규확률변수에서 평균을 빼고 표준편차로 나누면 $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$이다. 평균은 0, 분산은 1이 되어 표준정규분포를 따른다.

② O: 평균을 중심으로 좌우 확률이 같고 한 점의 확률은 0이다. 따라서 $P(X > \mu) = 0.5$이다.

③ O: 중앙값은 누적확률이 0.5가 되는 지점이다. 정규분포에서는 $P(X \le \mu) = 0.5$이므로 중앙값은 μ이다.

④ X: 제3사분위수는 누적확률 0.75에 대응한다. 표준정규분포의 75백분위수는 약 0.6745이므로 $Q_3 \approx \mu + 0.6745\sigma$이다. $\mu + 2\sigma$의 누적확률은 약 0.9772로 0.75가 아니다.

따라서 제3사분위수를 평균에 표준편차의 두 배를 더한 값으로 설명한 ④가 옳지 않다.

NoteBase

문 34. 두 확률변수 X와 Y의 결합확률질량함수가 다음과 같다. 다음 설명 중 옳지 않은 것은?

	$Y=-1$	$Y=0$	$Y=1$
$X=0$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$X=1$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$X=2$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

① E(XY)=1이다.

② X와 Y는 서로 독립이다.

③ E(X)=1이다.

④ E(Y)=0이다.

정답 1번

간단한 해설

표의 각 행·열 주변확률은 모두 1/3이고 두 변수는 독립이다. $E(X)=1$, $E(Y)=0$이므로 $E(XY)=0$이다.

NoteBase

선지별 해설

① X: 기댓값은 각 곱 xy에 해당 결합확률을 곱하여 합한다. 모든 칸의 확률이 1/9이므로 E(XY)=Σ_x=0^2x(-1)+x(0)+x(1)9=0이다. **각 행의 음수·양수 기여분이 상쇄**되므로 1이 아니다.

② O: 각 행의 합과 각 열의 합은 모두 $3 \times (1/9) = 1/3$이다. 따라서 모든 가능한 (x, y)에 대해 aligned&P(X=x,Y=y)=1/9&=(1/3)(1/3)=P(X=x)P(Y=y)aligned 가 성립하여 **서로 독립**이다. 단지 공분산이 0이라는 이유만으로 독립을 결론 내리는 것이 아니다.

③ O: X는 0, 1, 2를 각각 확률 1/3로 갖는다. 따라서 E(X)=0×13+1×13+2×13=1이다.

④ O: Y는 −1, 0, 1을 각각 확률 1/3로 갖는다. 따라서 E(Y)=(-1)×13+0×13+1×13=0이다. 독립성으로 $E(XY) = E(X)E(Y) = 1 \times 0 = 0$도 확인할 수 있다.

따라서 옳지 않은 설명은 ①이다.

NoteBase

문 35. 절편을 포함한 단순선형회귀분석에서 잔차(e_i)의 성질 중 항상 성립하는 것은?

① 모든 잔차($e_i, i = 1, \cdots, n$)는 0이다.

② 모든 잔차의 합은 0이다.

③ 모든 잔차의 제곱합은 0이다.

④ 모든 잔차의 제곱합은 1이다.

정답 2번

간단한 해설

절편을 포함한 최소제곱 회귀에서는 절편에 대한 정상방정식으로부터 **잔차의 합이 0**이다. 개별 잔차나 잔차제곱합이 모두 0이라는 뜻은 아니다.

선지별 해설

통상적인 최소제곱법으로 절편을 포함한 단순선형회귀모형을 적합한 경우를 다룬다. 잔차는 관측값과 적합값의 차 e_i=y_i-"0177_i이다.

① X: 각 관측값이 회귀직선 위에 정확히 놓일 때에만 모든 잔차가 0이다. 일반적인 자료에는 회귀직선에서 벗어난 관측값이 있으므로 개별 잔차가 모두 0일 필요는 없다.

② O: 최소제곱법은 $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$를 최소화한다. 이를 절편에 대해 미분한 정상방정식에서 $\sum_{i=1}^{n} e_i = 0$을 얻는다. 개별 잔차가 0이 아니라도 양수와 음수가 상쇄되어 합은 0이다.

③ X: 잔차제곱합 $\sum_{i=1}^{n} e_i^2$은 음이 아니며 모든 잔차가 0인 완전 적합일 때에만 0이다. 최소화한다는 사실이 최솟값 0을 보장하지는 않는다.

④ X: 잔차제곱합은 관측 자료와 단위에 따라 달라지므로 항상 1로 고정되지 않는다. 최소제곱 추정은 잔차제곱합을 1로 정규화하는 절차가 아니다.

따라서 항상 성립하는 성질은 잔차의 합이 0이라는 ②이다.

NoteBase

문 36. 확률변수 X가 자유도가 5인 t-분포, $t(5)$를 따를 때, 확률변수 X^2은 F-분포, $X^2 \sim F(a, b)$를 따른다. 이때 $a + b$의 값은?

① 6

② 10

③ 16

④ 25

정답 1번

간단한 해설

$t(\nu)$의 제곱은 $F(1, \nu)$를 따른다. 따라서 $X^2 \sim F(1, 5)$이고 $a + b = 6$이다.

계산 과정

① 분포 관계: 자유도 ν인 t분포 확률변수 T에 대해 **제곱은 분자 자유도 1, 분모 자유도 ν인 F분포**를 따른다. 즉 $T^2 \sim F(1, \nu)$이다.

② 관계의 근거: 서로 독립인 $Z \sim N(0, 1)$, $V \sim \chi^2(\nu)$로 $T = Z/\sqrt{V/\nu}$를 나타낼 수 있다. 제곱하면 $T^2 = (Z^2/1)/(V/\nu)$이고 $Z^2 \sim \chi^2(1)$이므로 두 독립 카이제곱을 각각의 자유도로 나눈 비인 F분포가 된다.

③ 대입과 계산: $\nu = 5$이므로 $X^2 \sim F(1, 5)$이다. 따라서 $a = 1, b = 5$이고 $a + b = 1 + 5 = 6$이다. 자유도는 단위가 없는 수이며, 확률변수를 제곱한다고 자유도를 25로 바꾸지 않는다.

따라서 구하는 값은 6으로 선택지 ①이다.

NoteBase

문 37. 커피콩의 볶는 온도가 커피의 카페인 함량에 미치는 영향을 알아보기 위해 세 온도(20℃, 40℃, 60℃)에서 각각 5회 반복실험을 실시하여 다음과 같은 분산분석표를 얻었다. (ㄱ)에 들어갈 값은?

요인	제곱합	자유도	평균제곱	F-값
처리	100		(ㄱ)	25
오차	24			
합계	124	14		

① 4

② 15

③ 25

④ 50

정답 4번

간단한 해설

온도 수준이 3개이므로 **처리 자유도는 2**이다. 처리 평균제곱은 제곱합을 자유도로 나눈 $\frac{100}{2} = 50$이다.

계산 과정

온도 수준은 20℃·40℃·60℃의 3개이고 각 수준에서 5회 반복하므로 전체 관측수는 15이다. 표의 (ㄱ)은 처리 행의 평균제곱이며, 처리 제곱합은 100이다.

일원분산분석에서 처리 자유도는 집단 수에서 1을 뺀 $3 - 1 = 2$이다. 평균제곱은 해당 제곱합을 해당 자유도로 나눈 값이므로 MS_처리=(100)/(2)=50이다.

오차 자유도는 $15 - 3 = 12$이고 오차 평균제곱은 MS_오차=(24)/(12)=2이다. 따라서 F=MS_처리MS_오차=(50)/(2)=25로 표의 F값과 일치한다. 합계 자유도도 $15 - 1 = 14$로 표와 같다. 평균제곱 50을 F값 25와 혼동하지 않는다.

따라서 (ㄱ)의 값은 50으로 ④와 일치한다.

NoteBase

문 38. 절편을 포함한 단순선형회귀분석에서 결정계수가 $R^2 = 0.64$이고 추정된 회귀모형의 기울기가 음수일 때, 표본상관계수 r은?

① −0.8

② −0.64

③ 0.64

④ 0.8

정답 1번

간단한 해설

$R^2 = r^2$이므로 $|r| = \sqrt{0.64} = 0.8$이다. 기울기가 음수이므로 $r = -0.8$이다.

계산 과정

① 적용 관계: 절편을 포함한 단순 최소제곱 선형회귀에서는 **결정계수가 표본상관계수의 제곱**과 같아 $R^2 = r^2$이다. 따라서 $|r| = \sqrt{0.64} = 0.8$이다. 이 관계를 다중회귀나 절편 없는 회귀에 그대로 일반화하지 않는다.

② 부호 결정: 추정 기울기는 β_1=r(s_y/s_x)이고 표본 표준편차의 비 s_y/s_x는 양수이므로, 기울기와 상관계수의 부호가 같다. 문제에서 기울기가 음수라고 했으므로 $r = -0.8$이다.

③ 검산과 구별: $(-0.8)^2 = 0.64$로 주어진 결정계수와 일치한다. 결정계수 0.64 자체에 음수 부호만 붙인 −0.64가 아니라 **제곱근을 취한 뒤 부호를 정해야** 한다. 상관계수는 단위가 없다.

따라서 표본상관계수는 −0.8로 선택지 ①이다.

NoteBase

문 39. 대학 졸업생의 초기 연봉(단위: 만원)은 표준편차가 10인 정규분포를 따른다. 졸업생 100명을 조사하여 초기 연봉의 평균에 대한 95% 신뢰구간을 추정하였을 때, 신뢰구간의 길이는? (단, $Z_{0.05} = 1.65, Z_{0.025} = 1.96$)

① 1.65

② 1.96

③ 3.3

④ 3.92

정답 4번

간단한 해설

표준오차는 $\frac{10}{\sqrt{100}} = 1$만원이다. 양측 95% 신뢰구간의 길이는 **오차한계의 두 배**이므로 $2 \times 1.96 \times 1 = 3.92$만원이다.

계산 과정

모집단이 정규분포이고 모표준편차 $\sigma = 10$만원을 알고 있으므로, 독립 확률표본 100명에 대한 모평균의 양측 95% 신뢰구간은 $\overline{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}}$이다. 전체 꼬리확률 0.05를 양쪽에 0.025씩 배분하여 임계값 1.96을 쓴다.

표본평균의 표준오차는 $\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1$만원이다. 따라서 평균 추정값에서 한쪽 끝까지의 거리인 오차한계는 $1.96 \times 1 = 1.96$만원이다.

신뢰구간의 길이는 상한에서 하한을 뺀 값이므로 오차한계의 두 배이다. 따라서 $2 \times 1.96 = 3.92$만원이다. 표본평균은 구간의 위치를 정하지만 구간 길이 계산에는 필요하지 않다.

95% 신뢰구간의 전체 길이 3.92만원은 ④와 일치한다.

NoteBase

문 40. 확률변수 X가 성공확률이 p, 시행횟수가 n인 이항분포 $Bin(n, p)$를 따를 때, 확률변수 $Y = n - X$의 분포는?

① $Bin(n, p)$

② $Bin(n - X, p)$

③ $Bin(n, 1 - p)$

④ $Bin(n - X, 1 - p)$

정답 3번

간단한 해설

$n - X$는 전체 n번 시행 중 실패 횟수이다. 실패확률은 $1 - p$이므로 $Y \sim Bin(n, 1 - p)$이다.

계산 과정

① 변수의 의미: X는 성공확률이 p인 독립적인 n번의 시행에서 성공 횟수이다. 따라서 $Y = n - X$는 같은 시행에서의 **실패 횟수**이다. 시행 횟수는 여전히 고정된 n이고, 실패확률은 $1 - p$이다.

② 확률질량함수 변환: $y = 0, 1, \ldots, n$에 대해 $P(Y = y) = P(X = n - y)$이다. 이항분포 공식을 대입하면 $P(Y = y) = \binom{n}{n-y} p^{n-y}(1 - p)^{y} = \binom{n}{y}(1 - p)^{y} p^{n-y}$이다. 이는 **모수가 $n, 1 - p$인 이항분포**의 확률질량함수이다.

③ 검산과 구별: $E(Y) = n - E(X) = n - np = n(1 - p)$이고 $\mathrm{Var}(Y) = \mathrm{Var}(X) = np(1 - p)$이다. 이는 $Bin(n, 1 - p)$의 평균·분산과 같다. $n - X$는 관측되는 실패 횟수이지 새 분포의 고정된 시행 횟수가 아니다. $p = 1/2$이면 원분포와 같아지지만 일반식은 성공확률을 $1 - p$로 바꾼 것이다.

따라서 $Y \sim Bin(n, 1 - p)$로 선택지 ③이다.

NoteBase