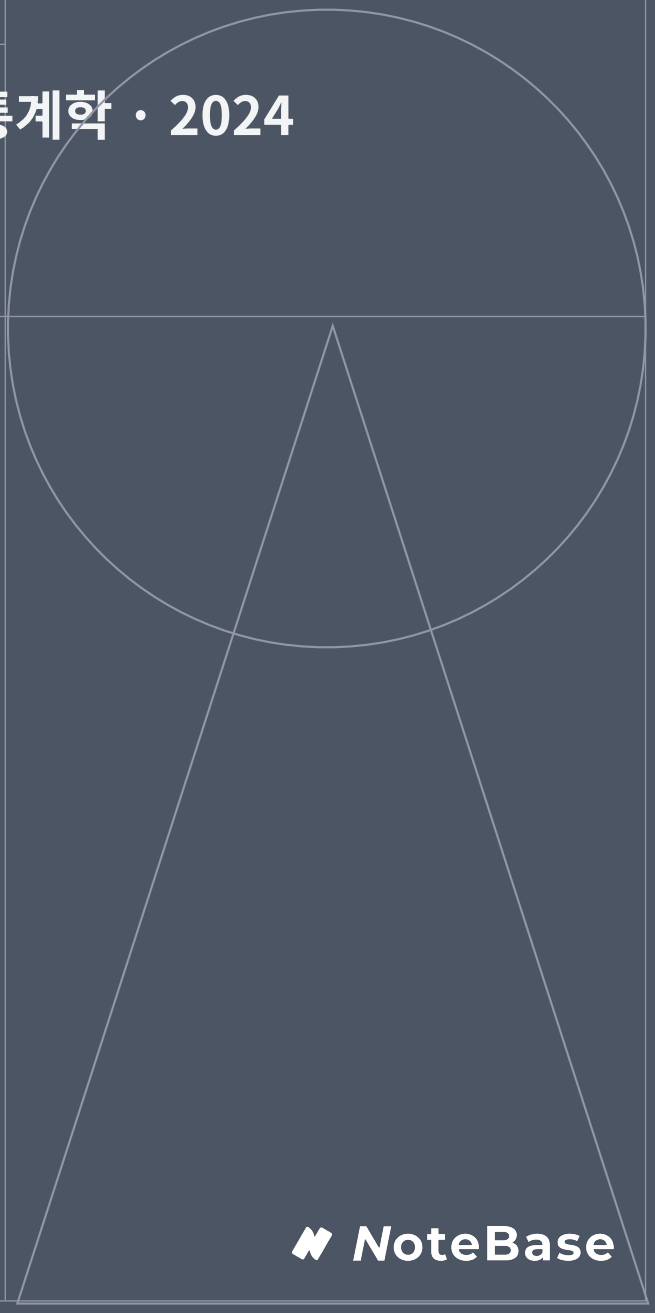
기출OX 해설집

공무원/경찰/공채(경위) · 통계학 · 2024



NoteBase

기출OX 해설집

공무원/경찰/공채(경위) · 통계학 · 2024

공무원/경찰/공채(경위) · 통계학 · 2024

문 1. 다음 중 평균이 가장 작은 분포는?

① 성공확률이 0.4이고 시행횟수가 10인 이항분포

② 자유도가 5인 t 분포

③ 분산이 5인 포아송분포

④ 분산이 10인 카이제곱분포

정답 2번

간단한 해설

자유도가 5인 t 분포는 **평균이 존재하며 0**이다. 다른 분포의 평균은 각각 4, 5, 5로 모두 양수이므로 이 분포의 평균이 가장 작다.

선지별 해설

① 이항분포: 시행횟수 n, 성공확률 p인 이항분포의 평균은 $E(X) = np$이다. 따라서 평균은 $10 \times 0.4 = 4$이다.

② t 분포: 표준 t 분포는 0을 중심으로 대칭이며, **자유도가 1보다 크면 평균이 존재하고 그 값은 0**이다. 자유도 5는 이 조건을 만족하므로 평균은 0이다.

③ 포아송분포: 모수가 λ인 포아송분포에서는 **평균과 분산이 모두 λ**이다. 분산이 5이므로 평균도 5이다.

④ 카이제곱분포: 자유도가 k이면 평균은 k, 분산은 2k이다. 주어진 분산이 10이므로 $2k = 10$, $k = \frac{10}{2} = 5$이다. 따라서 평균은 5이다.

네 분포의 평균은 차례로 4, 0, 5, 5이다. 가장 작은 값은 0이므로 ②의 자유도 5인 t 분포가 해당한다.

NoteBase

문 2. 그릇 B_1에는 빨간 공 3개와 흰 공 3개, 그릇 B_2에는 빨간 공 1개와 흰 공 2개, 그릇 B_3에는 빨간 공 5개와 흰 공 3개가 들어 있다. 각 그릇이 선택될 확률은 $P(B_1)=\frac{1}{3}$, $P(B_2)=\frac{1}{2}$, $P(B_3)=\frac{1}{6}$이다. 실험은 주어진 확률로 그릇을 선택하고, 그 그릇에서 임의로 하나의 공을 추출한다. 추출된 공이 빨간색일 때, 이 공이 그릇 B_2에서 나왔을 확률은?

① $\frac{1}{9}$
② $\frac{7}{16}$
③ $\frac{8}{17}$
④ $\frac{8}{21}$

정답 4번

간단한 해설

정답은 ④ **821**이다. 빨간 공이 나올 전체확률은 716, 그릇 B_2에서 빨간 공이 나올 결합확률은 16이다. 베이즈 정리에 따라 (1/6)/(7/16)=821을 얻는다.

계산 과정

빨간 공이 나오는 사건을 R이라고 하자. 구하려는 값은 빨간 공을 관찰한 뒤의 그릇 선택 확률인 $P(B_2 \mid R)$이며, 베이즈 정리를 적용한다.

각 그릇에서 빨간 공을 뽑을 조건부확률은 $P(R \mid B_1) = 3/6 = 1/2$, $P(R \mid B_2) = 1/3$, $P(R \mid B_3) = 5/8$이다.

전체확률법칙으로 빨간 공이 나올 확률을 구하면 P(R)=13·12+12·13+16·58=16+16+548=716이다.
그릇마다 선택 확률이 다르므로 단순히 세 비율의 평균을 내면 안 된다.

그릇 B_2를 선택하고 빨간 공을 뽑을 결합확률은 P(B_2 R)=P(B_2)P(R B_2)=12·13=16이다.

따라서 P(B_2 R)=(P(B_2 R))/(P(R))=(1/6)/(7/16)=821이므로 정답은 ④이다. ②의 7/16은 빨간 공이 나오는 전체확률이지, 빨간 공이라는 조건 아래 그릇을 판단하는 확률이 아니다. ①의 1/9, ③의 8/17도 계산 결과와 다르다.

NoteBase

문 3. 앞면이 나올 확률이 동일한 동전 3개를 독립적으로 던져 관측된 앞면의 수를 X라 하자. 이 실험을 80번 반복해 앞면이 나온 동전의 수가 0, 1, 2, 3인 빈도가 각각 15, 25, 25, 15번이다. 유의수준 5%에서 귀무가설 $H_0 : X \sim B(3, \frac{1}{2})$을 검정하고자 할 때, 다음 중 옳은 것은? 단, $\chi^2_\alpha(k)$는 자유도 k인 카이제곱분포의 제 100(1-α) 백분위수이다.

① 귀무가설 하에서 카이제곱 검정통계량은 자유도가 3인 카이제곱분포를 근사적으로 따르고, 계산한 값이 임곗값 $\chi^2_{0.05}(3) = 7.815$보다 작으므로 귀무가설을 채택한다.

② 귀무가설 하에서 카이제곱 검정통계량은 자유도가 3인 카이제곱분포를 근사적으로 따르고, 계산한 값이 임곗값 $\chi^2_{0.05}(3) = 7.815$보다 크기 때문에 귀무가설을 기각한다.

③ 귀무가설 하에서 카이제곱 검정통계량은 자유도가 4인 카이제곱분포를 근사적으로 따르고, 계산한 값이 임곗값 $\chi^2_{0.05}(4) = 9.488$보다 작으므로 귀무가설을 채택한다.

④ 귀무가설 하에서 카이제곱 검정통계량은 자유도가 4인 카이제곱분포를 근사적으로 따르고, 계산한 값이 임곗값 $\chi^2_{0.05}(4) = 9.488$보다 크기 때문에 귀무가설을 기각한다.

정답 1번

간단한 해설

기대도수는 10, 30, 30, 10이고 카이제곱 검정통계량은 $\frac{20}{3} \approx 6.667$이다. **자유도 3의 임곗값 7.815보다 작으므로** 귀무가설을 기각하지 않는다.

NoteBase

계산 과정

귀무가설이 범주별 확률을 완전히 지정하므로 카이제곱 적합도 검정을 사용한다. 관측도수는 앞면 수 0, 1, 2, 3에 대해 각각 15, 25, 25, 15이고 총 시행 수는 80이다.

귀무가설 $X \sim B(3, \frac{1}{2})$에서 네 범주의 확률은 $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$이다. 각 확률에 80을 곱하면 기대도수는 10, 30, 30, 10이며, 모두 충분히 커서 카이제곱 근사를 적용할 수 있다.

관측도수와 기대도수를 각각 O_i, E_i라 하면 검정통계량은 $T = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$이다. 따라서 aligned&T=((15-10)^2)/(10)+((25-30)^2)/(30)+((25-30)^2)/(30)&+((15-10)^2)/(10)=(20)/(3)≈6.667aligned
이다.

범주는 4개이고 도수 합계가 80으로 고정되어 한 개의 제약이 있다. 또한 성공확률은 귀무가설에서 이미 정해져 있어 자료로 추정한 모수가 없다. 따라서 자유도는 $4 - 1 = 3$이며, 귀무가설 아래에서 T는 근사적으로 자유도 3인 카이제곱분포를 따른다.

적합도 검정은 관측도수와 기대도수의 차이가 클수록 귀무가설을 기각하는 우측 검정이다.
$6.667 < \chi^2_{0.05}(3) = 7.815$이므로 유의수준 5%에서 귀무가설을 기각하지 않는다. 선지의 ‘채택’은 이 비기각 결정을 뜻하며, 귀무가설이 참임을 입증했다는 의미는 아니다.

자유도가 3이고 통계량이 임곗값보다 작다는 판단에 해당하는 것은 ①이다. ②는 임곗값과의 대소관계를 뒤집었고, ③·④는 자유도를 4로 잘못 정했다.

NoteBase

문 4. 어느 고등학교 학생 전체의 국어와 영어 성적이 이변량정규분포를 따른다고 할 때, 이들 성적 사이의 모상관계수(ρ)에 대한 가설 $H_0 : \rho = 0$ 대 $H_1 : \rho \neq 0$을 유의수준 5%에서 검정하고자 한다. 이를 위해 학생 $n = 20$명을 임의로 추출하여 표본상관계수 $r = 0.6$, 검정통계량 $t = 3.182$를 얻었을 때 옳은 설명은? 단, $t_\alpha(k)$는 자유도 k인 t 분포의 제$100(1-\alpha)$ 백분위수이다.

① 검정통계량 $t = 3.182$가 $t_{0.025}(18) = 2.101$보다 크므로 귀무가설을 기각한다.

② 검정통계량 $t = 3.182$가 $t_{0.025}(18) = 2.101$보다 크므로 귀무가설을 채택한다.

③ 검정통계량 $t = 3.182$가 $t_{0.025}(19) = 2.093$보다 크므로 귀무가설을 기각한다.

④ 검정통계량 $t = 3.182$가 $t_{0.025}(19) = 2.093$보다 크므로 귀무가설을 채택한다.

정답 1번

간단한 해설

정답은 ①이다. 자유도는 **20-2=18**이고, 양측 5% 검정의 임계값은 2.101이다. **3.182가 2.101보다 크므로 귀무가설을 기각**한다.

계산 과정

이변량정규분포에서 모상관계수가 0이라는 귀무가설을 검정할 때 검정통계량은 $t = r\sqrt{(n-2)/(1-r^2)}$이며, 귀무가설 아래에서 **자유도 $n-2$인 t분포**를 따른다. 따라서 이 문제의 자유도는 $20 - 2 = 18$이다.

값을 대입하면 $t = 0.6\sqrt{18/(1-0.6^2)} = 0.6\sqrt{18/0.64} \approx 3.182$로 주어진 통계량과 일치한다.

대립가설이 ρ"22600인 **양측검정**이므로 유의수준 0.05를 양쪽 꼬리에 각각 0.025씩 배분한다. 문제의 분위수 정의에 따라 임계값은 $t_{0.025}(18) = 2.101$이고, 기각역은 $|t| > 2.101$이다.

관측값은 $|3.182| > 2.101$이므로 **귀무가설을 기각**한다. 따라서 자유도 18의 임계값을 사용하여 기각한다고 한 ①이 옳다. ②는 기각 판단을 반대로 했고, ③·④는 자유도를 19로 잘못 사용했다.

결론은 유의수준 5%에서 모상관계수가 0이 아니라는 통계적 증거가 있다는 것이다. 표본상관계수는 양수이지만, 이 검정만으로 국어 성적이 영어 성적의 원인이라고 결론 내릴 수는 없다.

NoteBase

문 5. 두 확률변수 X, Y가 서로 독립이고 각각 정규분포 $N(0,2)$, $N(1,2)$를 따를 때 확률 $P(X>Y)$는? 단, 확률변수 Z는 표준정규분포 $N(0,1)$을 따른다.

① $P\left(Z>-\frac{1}{2}\right)$
② $P\left(Z>-\frac{1}{4}\right)$
③ $P\left(Z>\frac{1}{2}\right)$
④ $P\left(Z>\frac{1}{4}\right)$

정답 3번

간단한 해설

독립인 두 정규확률변수의 차 X − Y는 평균 −1, 분산 4인 정규분포를 따른다. 표준편차 2로 표준화하면 **0보다 큰 차이**가 표준정규변수의 1/2 초과에 대응한다.

계산 과정

차이를 W = X − Y로 두면 구할 확률은 P(W0)이다. 서로 독립인 정규확률변수의 선형결합도 정규분포를 따르므로 W의 평균과 분산을 구해 표준화한다.

정규분포 $N(\mu,\sigma^2)$의 두 번째 모수는 분산이다. 따라서 $E(W)=E(X)-E(Y)=0-1=-1$이다. 독립이므로 공분산은 0이며, 차의 분산은 빼는 것이 아니라 **두 분산을 더한다**.
$\mathrm{Var}(W)=\mathrm{Var}(X)+\mathrm{Var}(Y)=2+2=4$
그러므로 $W\sim N(-1,4)$이고 표준편차는 2이다.

평균을 빼고 표준편차로 나누면 $Z=\frac{W+1}{2}\sim N(0,1)$이다. 따라서

aligned&P(X Y)=P(W0)=P≤ft((W+1)/(2)(1)/(2))&=P≤ft(Z(1)/(2))aligned

이는 ③과 일치한다.

NoteBase

문 6. 확률분포에 대한 다음 설명 중 옳지 않은 것은?

① 확률변수 X가 $F(r_1, r_2)$ 분포를 따를 때, $Y = \frac{1}{X}$은 $F(r_2, r_1)$ 분포를 따른다.

② 서로 독립인 두 확률변수 X_1, X_2가 각각 이항분포 $B(n_1, p_1), B(n_2, p_2)$를 따를 때, $Y = X_1 + X_2$는 $B(n_1 + n_2, p_1 + p_2)$ 분포를 따른다.

③ 서로 독립인 두 확률변수 X_1, X_2가 각각 $\chi^2(r_1), \chi^2(r_2)$ 분포를 따를 때, $Y = X_1 + X_2$는 $\chi^2(r_1 + r_2)$ 분포를 따른다.

④ 서로 독립인 두 확률변수 X_1, X_2가 각각 정규분포 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$를 따를 때, $Y = X_1 - X_2$는 $N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ 분포를 따른다.

정답 2번

간단한 해설

독립인 이항확률변수의 합에서는 **성공확률이 같을 때 시행횟수만 합한다**. 공통 성공확률이 p이면 합의 분포는 $B(n_1 + n_2, p)$이며, 성공확률을 $p_1 + p_2$로 더하지 않는다.

선지별 해설

① O: 서로 독립인 $U \sim \chi^2(r_1), V \sim \chi^2(r_2)$에 대해 $X = \frac{U/r_1}{V/r_2}$는 $F(r_1, r_2)$ 분포를 따른다. 역수를 취하면 $\frac{1}{X} = \frac{V/r_2}{U/r_1}$가 되어 **분자와 분모의 자유도 순서가 바뀌므로** $F(r_2, r_1)$ 분포를 따른다.

② X: 독립인 이항확률변수의 합이 통상적인 이항분포가 되려면 **성공확률이 같아야** 한다. $p_1 = p_2 = p$이면 $X_1 + X_2 \sim B(n_1 + n_2, p)$이다. 시행횟수는 합하지만 성공확률은 합하지 않는다. 성공확률이 다르면 합은 일반적으로 하나의 이항분포로 표현되지 않는다.

③ O: 카이제곱분포는 독립인 표준정규확률변수들의 제곱합으로 설명할 수 있다. 서로 독립인 두 카이제곱확률변수를 더하면 이 제곱합의 항 수가 합쳐지므로 **자유도도 합해져** $X_1 + X_2 \sim \chi^2(r_1 + r_2)$가 된다.

④ O: 독립인 정규확률변수의 선형결합은 정규분포를 따른다. 차의 평균은 $E(X_1 - X_2) = \mu_1 - \mu_2$이고, 독립이므로 공분산이 0이 되어 $\mathrm{Var}(X_1 - X_2) = \sigma_1^2 + \sigma_2^2$이다. **차를 구하더라도 분산은 더한다**는 점이 중요하며, 따라서 제시된 정규분포를 따른다.

NoteBase

문 7. 어느 대학 학부생들 중 임의로 추출된 400명 학생들을 그들이 속한 단과대학과 성별에 따라 분류한 표가 다음과 같다. 학생이 소속된 단과대학이 성별과 독립이라는 귀무가설 검정에 대한 설명 중 옳지 않은 것은?

	인문대	공대	경영대	간호대	자연대	합계
남	23	60	30	17	70	200
여	17	20	50	43	70	200
합계	40	80	80	60	140	400

① 귀무가설 하에서 카이제곱 검정통계량은 근사적으로 자유도가 4인 카이제곱분포를 따른다.

② 계산된 카이제곱 검정통계량 37.17이 유의수준 1%에서 임곗값 13.28보다 크므로 귀무가설을 채택한다.

③ 귀무가설 하에서 자연대 여학생 셀에 해당하는 기대도수의 추정값은 70명이다.

④ 이 독립성검정의 검정 결과는 각 단과대학에 속하는 확률이 남녀 간에 모두 같은지에 대한 동질성검정의 결과와 동일하다.

정답 2번

간단한 해설

②가 옳지 않다. 자유도는 4이며, 검정통계량 37.17이 1% 임곗값 13.28보다 크므로 **독립이라는 귀무가설을 기각**한다. 자연대 여학생의 기대도수는 $200 \times 140/400 = 70$명이다.

NoteBase

선지별 해설

① (O) 성별 2개 범주와 단과대학 5개 범주의 분할표이므로 카이제곱 독립성검정의 자유도는 $(2-1)(5-1)=4$이다. 각 성별의 기대도수는 인문대 20, 공대 40, 경영대 40, 간호대 30, 자연대 70으로 충분히 크므로 카이제곱 근사를 적용하기에 적절하다.

② (X) 카이제곱 독립성검정에서는 관측도수와 기대도수의 차이가 클수록 검정통계량이 커지며, 큰 값이 기각역에 해당한다. $\chi^2=\sum(O-E)^2/E\approx 37.17$은 1% 유의수준 임곗값 13.28보다 크므로 **귀무가설을 기각**한다. 따라서 단과대학과 성별이 독립이라고 보기 어렵다는 통계적 근거가 있으며, 귀무가설을 채택한다는 설명은 틀리다.

③ (O) 독립성 가정 아래 기대도수는 행 합계와 열 합계의 곱을 전체 표본 수로 나누어 구한다. 자연대 여학생의 기대도수는 $E=200\times 140/400=70$명이다. 이 셀은 관측도수도 70명이어서 검정통계량에 대한 기여분이 0이다.

④ (O) 동일한 분할표에서 성별에 따라 단과대학의 분포가 같은지를 검정하는 동질성검정은 독립성검정과 같은 기대도수·검정통계량·자유도를 사용한다. 따라서 이 표에 대한 통계적 기각 여부도 같다. 다만 독립성검정과 동질성검정은 표본추출 설계와 귀무가설을 서술하는 관점에서 차이가 있다.

NoteBase

문 8. 공정한 육면체 주사위를 던져 나온 눈의 수를 X라 하면, 확률변수 X의 확률질량함수는 $f(x)=\frac{1}{6},\ x=1,2,\cdots,6$ 이다. 다음 중 가장 작은 값은?

① $E(X)$

② $E(2X-4)$

③ $Var(X)$

④ $Var(2X-4)$

정답 3번

간단한 해설

각 값은 차례로 $\frac{7}{2}$, 3, $\frac{35}{12}$, $\frac{35}{3}$이다. **분산은 배율의 제곱만큼 변하고 상수 이동에는 변하지 않으므로** $Var(X)=\frac{35}{12}$가 가장 작다.

선지별 해설

네 선지의 값을 같은 기준으로 계산하여 크기를 비교한다. 기대값은 선형성이 성립하지만, 분산은 상수를 더하거나 빼도 변하지 않고 배율의 제곱만큼 변한다.

① $E(X)=\frac{1+2+3+4+5+6}{6}=\frac{7}{2}=3.5$이다. 각 눈의 확률이 같으므로 여섯 값의 산술평균이다.

② 기대값의 선형성에 따라 $E(2X-4)=2E(X)-4=2\times\frac{7}{2}-4=3$이다.

③ $E(X^2)=\frac{1^2+2^2+3^2+4^2+5^2+6^2}{6}=\frac{91}{6}$이다. 따라서
aligned&Var(X)=E(X^2)-\E(X)\^2=(91)/(6)&-(49)/(4)=(35)/(12)≈2.917aligned
이다.

④ 분산의 변환 성질에 따라
aligned&Var(2X-4)=2^2Var(X)=4×(35)/(12)&=(35)/(3)≈11.667aligned
이다. 상수 −4는 중심 위치만 옮기므로 분산에 영향을 주지 않는다.

비교하면 $\frac{35}{12}<3<\frac{7}{2}<\frac{35}{3}$이므로 가장 작은 값은 ③ $Var(X)$이다.

NoteBase

문 9. 이산확률변수 X의 가능한 값이 1, 2, 3이고 $P(X=1)=2P(X=2)=4P(X=3)$일 때, X의 분산은?

① 3

② 2

③ $\frac{11}{7}$

④ $\frac{26}{49}$

정답 4번

간단한 해설

정답은 ④이다. 확률은 차례로 4/7, 2/7, 1/7이고, $E[X]=11/7$, $E[X^2]=3$이다. 따라서 **분산은** $3-(11/7)^2=26/49$이다.

계산 과정

$P(X=3)=a$라고 놓으면 조건에 의해 $P(X=2)=2a$, $P(X=1)=4a$이다. 확률의 합이 1이므로 $4a+2a+a=7a=1$이다. 따라서 **확률은 1·2·3에 대하여 각각** 4/7, 2/7, 1/7이다. 식에 적힌 계수 1·2·4를 그대로 확률비로 읽으면 안 된다.

기댓값은 E[X]=1×47+2×27+3×17=(11)/(7)이다. 이는 ③의 값이지만 문제에서 구하는 것은 기댓값이 아니라 분산이다.

제곱의 기댓값은 E[X^2]=1^2×47+2^2×27+3^2×17=(21)/(7)=3이다. 따라서 ①은 제곱의 기댓값에 해당한다.

분산 공식 $\mathrm{Var}(X)=E[X^2]-(E[X])^2$에 대입하면 $3-(\frac{11}{7})^2=\frac{147-121}{49}=\frac{26}{49}$이다. 따라서 정답은 ④ 26/49이고, ②의 2도 계산된 분산과 일치하지 않는다.

NoteBase

문 10. $Z \sim N(0,1)$**일 때, 다음 식이 성립한다.**
$P(-0.675 < Z < 0.675) = 0.5$
이 사실을 이용하여 $N(0,4)$ **분포의 사분위수범위(IQR)를 구하면?**

① $0.675 \times \frac{1}{4}$
② $0.675 \times \frac{1}{2}$
③ 0.675×2
④ 0.675×4

정답 4번

간단한 해설

중앙 50% 구간의 양 끝은 제1·제3사분위수이다. N(0,4)의 **표준편차는 2**이므로 사분위수는 ±1.35이고, 두 값의 차인 IQR은 2.7이다.

계산 과정

사분위수범위는 제3사분위수에서 제1사분위수를 뺀 값, 즉 $IQR = Q_3 - Q_1$이다. 표준정규분포는 0을 중심으로 대칭이므로 주어진 중앙 50% 구간 바깥의 확률은 양쪽에 각각 25%씩 놓인다. 따라서 Z의 제1사분위수는 −0.675, 제3사분위수는 0.675이다.

구하려는 확률변수를 X라 하면 $X \sim N(0,4)$에서 분산은 4, **표준편차는 2**이다. $Z = X/2$이므로 X의 사분위수는 Z의 사분위수에 2를 곱한 값이다.
$Q_1 = 2 \times (-0.675) = -1.35$, $Q_3 = 2 \times 0.675 = 1.35$

두 사분위수 사이의 폭을 구하면
$IQR = 1.35 - (-1.35) = 0.675 \times 4 = 2.7$
따라서 ④와 일치한다.

NoteBase

문 11. 어떤 직물의 가공 시 처리액의 농도가 직물의 인장강도에 영향을 미치는지 알아보기 위해 네 가지 농도에서 반복 4회씩, 총 16회를 랜덤하게 실험하여 얻은 분산분석표의 일부가 다음과 같다. 처리액의 농도가 직물의 인장강도에 영향을 미치지 않는다는 귀무가설을 유의수준 5%에서 검정한 결과로 옳은 것은?

요인	제곱합	자유도	평균제곱	F
처리	1800			
잔차	1600			
합계	3400			

① 검정통계량(F 값)이 $F_{0.05}(3, 12) = 3.49$보다 작으므로 귀무가설을 채택한다.

② 검정통계량(F 값)이 $F_{0.05}(3, 12) = 3.49$보다 크므로 귀무가설을 기각한다.

③ 검정통계량(F 값)이 $F_{0.05}(4, 12) = 3.26$보다 작으므로 귀무가설을 채택한다.

④ 검정통계량(F 값)이 $F_{0.05}(4, 12) = 3.26$보다 크므로 귀무가설을 기각한다.

정답 2번

간단한 해설

처리·잔차 자유도는 각각 3과 12이고, 검정통계량은 $F = \frac{1800/3}{1600/12} = 4.5$이다. **임곗값 3.49보다 크므로** 평균 인장강도가 모두 같다는 귀무가설을 기각한다.

NoteBase

계산 과정

네 농도 집단의 평균 인장강도가 모두 같다는 귀무가설을 일원배치 분산분석으로 검정한다. 통상적인 F 검정은 오차의 독립성·정규성과 집단 간 등분산성을 전제로 한다. 집단 수는 4, 전체 관측 수는 16이며, 표에서 처리제곱합 1800, 잔차제곱합 1600, 총제곱합 3400을 읽을 수 있다.

처리 자유도는 $4-1=3$, 잔차 자유도는 $16-4=12$이다. 총자유도는 $16-1=15$이며 두 자유도의 합과 일치한다. 따라서 검정에 쓰는 F 분포의 분자·분모 자유도는 각각 3과 12이다.

평균제곱은 각 제곱합을 해당 자유도로 나눈 값이다. 처리평균제곱은 $MS_T=\frac{1800}{3}=600$, 잔차평균제곱은 $MS_E=\frac{1600}{12}=\frac{400}{3}$이다.

검정통계량은 처리평균제곱을 잔차평균제곱으로 나누어 $F=\frac{MS_T}{MS_E}=\frac{600}{400/3}=4.5$가 된다. 집단 내 변동에 비해 집단 간 변동이 클수록 F 값이 커지므로 우측 기각역을 사용한다.

$4.5>F_{0.05}(3,12)=3.49$이므로 유의수준 5%에서 귀무가설을 기각한다. 이는 네 집단의 평균이 모두 같지는 않다는 근거이지, 모든 집단 쌍의 평균이 서로 다르다는 결론은 아니다.

따라서 분자·분모 자유도를 3과 12로 적용하고 귀무가설을 기각하는 ②가 해당한다. ①은 대소관계와 결정을 뒤집었으며, ③·④는 분자 자유도를 집단 수 4와 혼동했다.

NoteBase

문 12. 자료의 상자그림 요약에 대한 설명으로 옳은 것만을 모두 고른 것은?

가. 수염(whisker)은 안울타리 경계를 벗어나지 않는다.
나. 상자그림은 범주형보다 연속형 자료에 적합한 요약 방법이다.
다. 이상점이 한 개만 존재한다면, 이 값은 자료의 최솟값 또는 최댓값이다.

① 가, 나
② 가, 다
③ 나, 다
④ 가, 나, 다

정답 4번

간단한 해설

정답은 ④ **가·나·다**이다. 수염은 안울타리 안의 가장 바깥 관측값까지 이어지고, 상자그림은 수치형 자료의 요약에 적합하다. **이상점이 하나뿐이면 최솟값 또는 최댓값**이어야 한다.

보기별 해설

가. (O) 사분위범위를 $IQR = Q_3 - Q_1$이라고 하면 안울타리는 $Q_1 - 1.5IQR$과 $Q_3 + 1.5IQR$이다. 일반적인 수정 상자그림의 수염은 이 범위 안에 있는 관측값 중 가장 바깥쪽 값까지 이어지므로 안울타리를 벗어나지 않는다. 수염 끝이 항상 울타리의 수치 자체와 같지는 않다.

나. (O) 상자그림은 중앙값·사분위수·산포·이상점 등을 이용하여 수치형 자료의 분포를 요약한다. 따라서 명목적 범주형 자료보다는 연속형 자료를 요약하는 데 적합하다. 값의 순서와 간격을 해석할 수 있는 이산형 수치 자료에도 사용할 수 있다.

다. (O) 이상점이 아래쪽 안울타리보다 작다면 그보다 더 작은 관측값 역시 울타리 밖에 있어 이상점이 된다. 따라서 이상점이 하나뿐이라면 아래쪽 이상점은 전체 최솟값이어야 한다. 위쪽도 같은 논리로 유일한 이상점은 전체 최댓값이어야 한다.

가·나·다가 모두 옳으므로 정답은 ④이다. 특히 수염 끝은 울타리가 아니라 울타리 안의 실제 관측값이며, 유일한 이상점은 분포의 한쪽 끝에 위치한다는 점을 기억해야 한다.

NoteBase

문 13. X가 평균이 λ인 포아송분포를 따르는 확률변수일 때, 다음의 가설을 검정하려고 한다.
$H_0 : \lambda \leq 8$ **대** $H_1 : \lambda > 8$
기각역을 $X \geq 13$**으로 정할 때,** $\lambda = 15$**에서의 검정력은?**

① 1-Σ_x=0^12(e^-88^x)/(x!)

② Σ_x=0^12(e^-88^x)/(x!)

③ 1-Σ_x=0^12(e^-1515^x)/(x!)

④ Σ_x=0^12(e^-1515^x)/(x!)

정답 3번

간단한 해설

정답은 ③이다. 검정력은 $\lambda = 15$**일 때 X≥13이 될 확률**이므로, $1 - \sum_{x=0}^{12} e^{-15}15^x/x!$이다.

계산 과정

검정력은 대립가설에 속하는 특정 모수값이 실제 값일 때 **귀무가설을 기각할 확률**이다. 따라서 $\lambda = 15$에서의 검정력은 평균 15인 포아송분포를 사용하여 기각역에 들어갈 확률을 계산해야 한다.

기각역이 X≥13이므로 검정력은 P_15(X≥13)이다. X가 정수값을 갖는 이산확률변수이므로 여사건은 **X≤12**이며, X≤13이 아니다.

평균이 15인 포아송분포의 확률질량함수는 $P_{15}(X = x) = e^{-15}15^x/x!$이다. 따라서 검정력은 **P_15(X≥13)=1-Σ_x=0^12(e^-1515^x)/(x!)**로 계산되며, 이에 해당하는 선택지는 ③이다.

④는 $\lambda = 15$에서 기각하지 않을 확률, 즉 제2종 오류확률 $\beta(15)$이다. ①은 귀무가설의 경계 $\lambda = 8$에서 기각할 확률이고, ②는 그 경계에서 기각하지 않을 확률이다. 검정력을 구할 때에는 **모수값 15와 기각사건**을 함께 맞추어야 한다.

NoteBase

문 14. 모의고사 점수 (X)와 본고사 점수 (Y) 간의 관계를 알아보기 위해 다음과 같은 단순회귀모형을 고려한다.

$Y_i = \alpha + \beta(x_i - \overline{x}) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$

최소제곱법에 의한 회귀계수 추정을 위해 어느 반 학생 10명의 점수를 이용하여 다음과 같은 결과를 얻었다.

$\sum_{i=1}^{10} y_i = 810, \quad \sum_{i=1}^{10} y_i^2 = 68710$

또한 잔차제곱합이 $SSE = 310$일 때, 결정계수는?

① 0.10

② 0.85

③ 0.90

④ 0.95

정답 3번

간단한 해설

평균 주위 총제곱합은 $SST = 68{,}710 - \frac{810^2}{10} = 3{,}100$이다. 따라서 결정계수는 $R^2 = 1 - \frac{310}{3{,}100} = 0.90$이다.

계산 과정

결정계수는 종속변수 Y의 평균 주위 총변동 중 회귀모형으로 설명되는 비율이다. 주어진 모형에는 절편 α가 포함되어 있으므로 최소제곱 회귀에서 $R^2 = 1 - \frac{SSE}{SST}$를 사용한다. **설명변수를 평균 중심화해도 절편이 없어지는 것은 아니다.**

학생 수가 10명이고 점수 합이 810이므로 y=(810)/(10)=81이다. 총제곱합 SST는 점수의 제곱합 자체가 아니라 평균으로부터의 편차제곱합이다.

$SST = \sum_{i=1}^{10} y_i^2 - \frac{(\sum_{i=1}^{10} y_i)^2}{10}$

$SST = 68{,}710 - \frac{810^2}{10} = 3{,}100$

잔차제곱합 SSE = 310은 모형으로 설명되지 않은 변동이다. 따라서

$R^2 = 1 - \frac{310}{3{,}100} = 0.90$

회귀모형이 본고사 점수의 표본 총변동 중 90%를 설명한다는 뜻이며, 개별 점수를 90%의 확률로 맞힌다는 뜻은 아니다.

구한 결정계수는 0.90이므로 ③에 해당한다.

NoteBase

문 15. 두 확률변수 간의 공분산과 상관계수에 대한 설명으로 옳은 것은? 단, a, b, c, d는 상수이다.

① ac<0이면 $Cov(aX+b, cY+d) = -acCov(X, Y)$이다.

② $Corr(X, Y) = 0$이면 X와 Y는 서로 독립이다.

③ 표준화된 두 확률변수 간의 공분산 $Cov\left(\frac{X-\mu_1}{\sigma_1}, \frac{Y-\mu_2}{\sigma_2}\right)$는 $Corr(X, Y)$와 같다.

④ $Cov(X, Y)$와 $Corr(X, Y)$는 모두 X와 Y의 단위에 무관한 측도이다.

정답 3번

간단한 해설

각 변수에서 평균을 빼고 표준편차로 나누면, 공분산도 두 표준편차의 곱으로 나누어진다. 그 값이 **피어슨 상관계수**이다. 이 관계는 두 변수의 분산이 모두 **유한하고 0보다 클 때** 성립한다.

NoteBase

선지별 해설

공분산은 두 변수가 각 평균에서 벗어난 정도의 곱을 평균한 값이다. 이하 $C = \mathrm{Cov}(X, Y)$라 하자. 피어슨 상관계수는 C를 두 표준편차의 곱으로 나눈 값이다. 피어슨 상관계수와 표준화를 사용하려면 두 변수의 분산이 모두 **유한하고 0보다 커야 한다**.

① X: 상수를 더해도 평균에서의 편차는 바뀌지 않고, 상수를 곱하면 편차도 그만큼 변한다. $U = aX + b$, $V = cY + d$라 두면 $\mathrm{Cov}(U, V) = acC$이다. $ac < 0$이어도 계수는 **−ac가 아니라 ac**이다. 원래 공분산이 0이 아니라면 변환 후 공분산의 부호가 원래와 반대가 된다.

② X: 상관계수가 0이라는 것은 **선형 상관이 없다는 뜻이며 독립을 보장하지 않는다**. 독립이면 공분산과 상관계수가 0이지만, 그 역은 일반적으로 성립하지 않는다. 예를 들어 X가 −1, 0, 1을 같은 확률로 갖고 $Y = X^2$이면 $E[X] = E[X^3] = 0$이다. 또한 $E[XY] = E[X^3] = 0$이므로 공분산과 상관계수는 0이다. 그러나 X가 0이면 Y도 반드시 0이 되어 Y의 분포가 달라지므로 두 변수는 독립이 아니다.

③ O: X와 Y의 평균을 각각 μ_1, μ_2, 표준편차를 각각 σ_1, σ_2라 하자. 표준화는 **평균을 빼고 표준편차로 나누는** 변환이다.
$Z_1 = \frac{X - \mu_1}{\sigma_1}, Z_2 = \frac{Y - \mu_2}{\sigma_2}$로 두면, 평균을 빼는 것은 공분산에 영향을 주지 않고 각 표준편차로 나누는 효과만 남는다.
따라서 $\mathrm{Cov}(Z_1, Z_2) = \frac{C}{\sigma_1 \sigma_2}$이다.
한편 $\mathrm{Corr}(X, Y) = \frac{C}{\sigma_1 \sigma_2}$이므로 표준화된 두 변수의 공분산은 원래 두 변수의 상관계수와 같다.

④ X: 공분산은 **두 변수의 단위의 곱**을 단위로 갖고, 상관계수는 **단위가 없는 측도**이다. 예를 들어 X의 단위를 m에서 cm로 바꾸어 값이 100배가 되면 공분산도 100배가 된다. 상관계수는 분자의 공분산과 분모의 X의 표준편차가 모두 100배가 되어 그 비율이 상쇄되므로 변하지 않는다.

NoteBase

문 16. 다중회귀모형 $Y=\beta_0+\beta_1x_1+\beta_2x_2+\varepsilon,\ \varepsilon\sim N(0,\sigma^2)$을 적합하여 아래의 분산분석표를 얻었을 때, 다음 중 옳은 것은?

요인	제곱합	자유도	평균제곱	F	p-값
회귀	140	2	(ㄱ)	(ㄷ)	<0.001
잔차	10	20	(ㄴ)		
합계	150	22			

① F 검정의 귀무가설은 $H_0:\beta_0=\beta_1=\beta_2=0$이다.

② (ㄱ)+(ㄴ)+(ㄷ)=200.5이다.

③ $X\sim F(2,20)$일 때, p-값은 $P(X>140)$이다.

④ 오차분산에 대한 불편추정값은 $\hat{\sigma}^2=70$이다.

정답 3번

간단한 해설

회귀평균제곱은 70, 잔차평균제곱은 0.5이므로 $F=140$이다. 귀무가설 아래에서 자유도는 2와 20이며, p-값은 **관측된 F 값 이상이 나올 우측 꼬리확률**이다.

NoteBase

선지별 해설

표의 회귀제곱합은 140, 잔차제곱합은 10이고 자유도는 각각 2와 20이다. 총제곱합 150과 총자유도 22도 두 부분의 합과 일치한다. 평균제곱은 제곱합을 해당 자유도로 나눈 값이며, 전체 회귀의 유의성을 보는 F 통계량은 회귀평균제곱을 잔차평균제곱으로 나눈다.

① X: 전체 회귀의 F 검정은 두 설명변수의 기울기가 모두 0인지 검정한다. 귀무가설은 $H_0 : \beta_1 = \beta_2 = 0$이며, 절편 β_0까지 0으로 제한하지 않는다. 귀무가설이 성립하면 반응변수의 평균은 설명변수에 따라 변하지 않는 상수이다.

② X: (char"3131) $= \frac{140}{2} = 70$, (char"3134) $= \frac{10}{20} = 0.5$, (char"3137) $= \frac{70}{0.5} = 140$이다. 따라서 합은 $70 + 0.5 + 140 = 210.5$이며 200.5가 아니다.

③ O: 귀무가설 아래에서 검정통계량은 분자 자유도 2, 분모 자유도 20의 F 분포를 따른다. 관측값이 140이고 큰 값일수록 귀무가설에 반하는 증거이므로, $X \sim F(2, 20)$일 때 p-값은 $P(X > 140)$이다. 연속분포이므로 '이상'과 '초과'의 확률은 같다. 이는 표의 p-값이 0.001보다 작다는 표시와도 부합한다.

④ X: 오차분산의 불편추정값은 회귀평균제곱이 아니라 잔차평균제곱이다. 따라서 σ^2=(SSE)/(20)=(10)/(20)=0.5이며, 70은 회귀평균제곱이다.

NoteBase

문 17. $X \sim B\left(150, \frac{4}{10}\right)$**일 때, 확률** $P(60 \le X < 80)$**에 대한 연속성 수정을 고려한 근사식은? 단,** $\Phi(\cdot)$**는 표준정규분포의 누적분포함수이다.**

① $\Phi\left(\frac{19.5}{6}\right) - \Phi\left(\frac{1}{12}\right)$
② $\Phi\left(\frac{19.5}{6}\right) - \Phi\left(-\frac{1}{12}\right)$
③ $\Phi\left(\frac{20.5}{6}\right) - \Phi\left(\frac{1}{12}\right)$
④ $\Phi\left(\frac{20.5}{6}\right) - \Phi\left(-\frac{1}{12}\right)$

정답 2번

간단한 해설

정답은 ②이다. 60≤ X80은 정수 60부터 79까지이므로 연속성 수정 구간은 **59.5부터 79.5**이다. 평균 60, 표준편차 6으로 표준화하면 ≤ft((19.5)/(6))-≤ft(-112)을 얻는다.

계산 과정

이항분포의 평균과 분산은 $\mu = np = 150 \times 0.4 = 60$, $\sigma^2 = np(1-p) = 150 \times 0.4 \times 0.6 = 36$이다. 따라서 표준편차는 $\sigma = 6$이며, 정규근사 변수 Y는 평균 60, 분산 36을 갖는다.

X는 정수값을 가지므로 60≤ X80은 60≤ X≤79와 같다. 포함되는 양 끝 정수 60과 79의 막대 전체를 덮도록 연속성 수정을 하면 59.5≤ Y≤79.5가 된다. 80은 포함되지 않으므로 위쪽 경계를 80.5로 두면 안 된다.

표준화하면 아래 경계는 $(59.5 - 60)/6 = -1/12$, 위 경계는 $(79.5 - 60)/6 = 19.5/6$이다. 따라서 구하는 확률은 P≤ft(-112≤ Z≤(19.5)/(6))로 근사한다.

누적분포함수로 표현하면 ≤ft((19.5)/(6))-≤ft(-112)이므로 정답은 ②이다. ①은 아래 경계의 부호가 틀리고, ③은 양쪽 경계가 모두 틀리며, ④는 위쪽 경계를 80.5로 잘못 수정한 식이다.

NoteBase

문 18. 다음은 반복이 있는 이원배치 분산분석을 수행한 결과의 일부이다. 다음 중 옳지 않은 것은? 단, 인자 A, B는 모두 모수인자이다.

요인	제곱합	자유도	평균제곱	F
A	120	3	(ㄱ)	(ㅁ)
B	160	2	(ㄴ)	(ㅂ)
$A \times B$	72	6	(ㄷ)	(ㅅ)
잔차	60	12	(ㄹ)	
합계	412	23		

① (ㄹ)<(ㄷ)<(ㄱ)<(ㄴ)

② 인자 A와 B의 수준 수는 다르다.

③ 반복 수는 3으로 동일하다.

④ (ㅁ)+(ㅂ)+(ㅅ) > 25

정답 3번

간단한 해설

정답은 ③이다. A와 B의 수준 수는 각각 4와 3이고, 잔차 자유도 $4 \times 3 \times (r - 1) = 12$에서 **반복 수는 2**이다.

NoteBase

선지별 해설

① (O) 평균제곱은 제곱합을 자유도로 나누어 구한다. 표에서 (ㄱ)은 120/3 = 40, (ㄴ)은 160/2 = 80, (ㄷ)은 72/6 = 12, (ㄹ)은 60/12 = 5이다. 따라서 **5124080**이므로 (ㄹ), (ㄷ), (ㄱ), (ㄴ)의 오름차순이 성립한다.

② (O) 인자 A의 수준 수를 a, B의 수준 수를 b라 하면 각각의 자유도는 $a-1$, $b-1$이다. 따라서 $a-1=3$, $b-1=2$이므로 **A는 4수준, B는 3수준**으로 서로 다르다. 교호작용의 자유도도 $(4-1)(3-1)=6$으로 표와 일치한다.

③ (X) 각 처리조합의 반복 수를 r라 하면 잔차 자유도는 $ab(r-1)$이다. $4\times3\times(r-1)=12$이므로 **반복 수는** $r=2$이다. 전체 관측 수가 $4\times3\times2=24$이고 총 자유도가 $24-1=23$인 점도 표와 일치한다. 따라서 반복 수가 3이라는 설명은 틀리다.

④ (O) A와 B가 모두 모수인자이므로 각 효과의 F통계량은 해당 평균제곱을 잔차 평균제곱 5로 나눈다. (ㅁ)은 40/5 = 8, (ㅂ)은 80/5 = 16, (ㅅ)은 12/5 = 2.4이다. 따라서 $8+16+2.4=26.4>25$이다.

NoteBase

문 19. 단순회귀모형 $Y_i = \alpha + \beta x_i + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2)$을 최소제곱법으로 적합한 결과를 $\hat{y}_i - \bar{y} = A(x_i - \bar{x})$로 나타낼 때, $(E(A),\ Var(A))$로 옳은 것은? 단, S_xx=Σ_i=1^n(x_i-x)^2이다.

① $\left(\beta,\ \frac{\sigma^2}{S_{xx}}\right)$
② $\left(\beta,\ \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2\right)$
③ $\left(\alpha,\ \frac{\sigma^2}{S_{xx}}\right)$
④ $\left(\alpha,\ \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2\right)$

정답 1번

간단한 해설

평균점을 기준으로 쓴 회귀직선에서 A는 **기울기 β의 최소제곱 추정량**이다. 통상적인 단순회귀모형에서 그 평균은 β, 분산은 σ^2/S_{xx}이다.

핵심 풀이

절편을 포함한 최소제곱 회귀직선은 표본평균점 $(\bar{x}, \bar{y})$을 지난다. 따라서 "0177_i-y=β(x_i-x)이며, 주어진 A는 절편이 아니라 **기울기의 최소제곱 추정량**이다.

기울기 추정량은 $A = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{S_{xx}}$이다. 회귀모형을 대입하고 $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$을 이용하면 $A = \beta + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i}{S_{xx}}$
오차의 평균이 0이므로 $E(A) = \beta$이다. 즉 A는 기울기 β의 불편추정량이다.

통상적인 단순회귀모형에서 설명변수 값은 고정하고, 오차들은 서로 비상관이며 공통 분산을 갖는 것으로 둔다. 이때 S_xx0이면
$\mathrm{Var}(A) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma^2}{S_{xx}^2} = \frac{\sigma^2}{S_{xx}}$
따라서 설명변수의 산포가 클수록 기울기 추정의 분산이 작아진다.

$\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2$는 기울기가 아닌 **절편 추정량의 분산**이다. 절편 추정량의 평균은 α이고, 여기서 묻는 기울기 추정량의 평균은 β이다. 그러므로 $(E(A), \mathrm{Var}(A)) = \left(\beta, \frac{\sigma^2}{S_{xx}}\right)$이며 ①과 일치한다.

NoteBase

문 20. 두 모평균의 비교에서, 다음의 조건을 모두 만족하는 경우에 해당하는 검정법은?

가. 두 모집단은 정규분포를 따른다.
나. 모집단의 분산은 미지이며 같지 않다($\sigma_1^2 \neq \sigma_2^2$).
다. 표본의 크기(n_1, n_2)는 소표본으로 20보다 작다.

① t 검정
② Z 검정
③ χ^2 검정
④ 이항검정

정답 1번

간단한 해설

정규모집단의 두 평균을 비교하면서 모분산을 모르고 표본도 작으므로 t 검정을 사용한다. 특히 분산이 같지 않으므로 독립표본 비교에서는 **Welch t 검정**이 적합하다.

핵심 풀이

두 모평균을 비교할 때 모분산이 알려져 있으면 정규분포를 이용한 Z 검정을 생각할 수 있다. 그러나 이 문항은 **모분산이 미지**이므로 표본분산으로 표준오차를 추정해야 한다. 표본이 작아 분산 추정의 불확실성을 무시하기 어려우므로 t 검정을 선택한다. 두 모집단의 정규성은 소표본에서 이 접근의 중요한 조건이다.

두 분산이 같지 않다는 조건 때문에 공통분산을 가정하여 표본분산을 합치는 합동분산 t 검정은 적합하지 않다. 서로 독립인 두 표본을 비교하는 통상적인 경우에는 **등분산을 가정하지 않는 Welch t 검정**을 사용한다. 표준오차는 두 표본분산을 각각 반영하며, 자유도는 Welch–Satterthwaite 방식으로 근사한다.

카이제곱 검정은 정규모집단의 분산 검정이나 범주형 자료의 적합도·독립성 검정 등에 쓰이며, 이항검정은 이항자료의 성공확률을 검정한다. 이들은 주어진 소표본의 두 모평균 비교에 해당하지 않는다.

따라서 선택지에서는 ① t 검정이 해당한다.

NoteBase

문 21. X_1, X_2를 아래 분포로부터의 크기 2인 확률표본이라 할 때, 확률 $P\left(\frac{X_1+X_2}{2}=2\right)$는?

x	0	1	2	3
$P(X=x)$	0.1	0.2	0.4	0.3

① 0.12

② 0.22

③ 0.28

④ 0.44

정답 3번

간단한 해설

평균이 2이려면 두 관측값의 합이 4여야 한다. **가능한 순서쌍 세 개의 확률을 모두 더하면** $0.2 \times 0.3 + 0.4^2 + 0.3 \times 0.2 = 0.28$이다.

계산 과정

표에서 확률변수의 가능한 값은 0, 1, 2, 3이고 각 확률은 0.1, 0.2, 0.4, 0.3이다. 확률표본이므로 X_1, X_2는 이 분포를 공통으로 따르며 서로 독립이다.

표본평균이 2라는 사건은 $\frac{X_1+X_2}{2}=2$, 즉 $X_1+X_2=4$이다. 가능한 순서쌍은 $(1,3), (2,2), (3,1)$뿐이다. 값 4는 분포에 없으므로 $(0,4), (4,0)$은 포함하지 않는다.

독립성에 따라 각 순서쌍의 확률은 곱으로 계산한다. $P(X_1=1, X_2=3) = 0.2 \times 0.3 = 0.06$, $P(X_1=2, X_2=2) = 0.4 \times 0.4 = 0.16$, $P(X_1=3, X_2=1) = 0.3 \times 0.2 = 0.06$이다.

세 사건은 서로 배반하므로 확률을 더한다. 따라서 $P\left(\frac{X_1+X_2}{2}=2\right) = 0.06 + 0.16 + 0.06 = 0.28$이다. 서로 다른 순서인 $(1,3)$과 $(3,1)$을 모두 세되 $(2,2)$는 한 번만 세어야 한다.

계산 결과 0.28에 해당하는 선택지는 ③이다.

NoteBase

문 22. 추정량과 관련된 다음 성질 중 옳지 않은 것은?

① 일치성은 표본이 클 때의 추정량의 성질이다.

② 불편추정량의 평균제곱오차는 분산과 같다.

③ 두 불편추정량 중 분산이 작은 쪽이 보다 유효한 추정량이다.

④ 불편성은 추정량이 반드시 가져야 하는 성질이다.

정답 4번

간단한 해설

정답은 ④이다. **추정량이 반드시 불편추정량이어야 하는 것은 아니다.** 편향과 분산을 함께 고려한 평균제곱오차나 일치성 등도 중요한 판단 기준이다.

선지별 해설

① (O) 일치성은 **표본 크기가 증가할 때 추정량이 참모수에 가까워지는 성질**이다. 통상적으로 추정량이 참모수에 확률수렴한다는 뜻이며, 특정한 유한 표본에서 기댓값이 모수와 같은지를 따지는 불편성과는 구별된다.

② (O) 추정량 T의 평균제곱오차는 $E[(T-\theta)^2] = \mathrm{Var}(T) + (E[T]-\theta)^2$이다. 불편추정량은 $E[T]=\theta$이므로 편향이 0이고, **평균제곱오차는 분산과 같다.**

③ (O) 같은 모수를 추정하는 두 불편추정량을 비교하면 분산이 작은 쪽의 평균제곱오차도 작다. 따라서 동일한 조건에서 **분산이 작은 불편추정량이 더 유효**하다고 평가한다. 다만 모수값에 따라 분산의 대소가 바뀌면 어느 한 추정량이 모든 모수값에서 더 유효하다고 단정할 수는 없다.

④ (X) **불편성은 바람직한 기준 중 하나이지 모든 추정량의 필수 조건은 아니다.** 편향이 조금 있더라도 분산이 충분히 작아 평균제곱오차가 더 작은 추정량을 선택할 수 있다. 또한 유한 표본에서는 편향되어 있지만 표본 크기가 커지면 참모수로 수렴하는 일치추정량도 존재한다.

NoteBase

문 23. 두 학교의 학생들이 동일한 시험을 치렀다. 학생 수는 A학교가 B학교의 2배이다. B학교 학생들의 평균 점수가 75점이고 두 학교를 합한 전체 학생의 평균 점수가 65점이었다면, A학교 학생들의 평균 점수는?

① 50점

② 55점

③ 60점

④ 62점

정답 3번

간단한 해설

정답은 ③ **60점**이다. 학생 수의 비가 A:B=2:1이므로 전체 평균은 $(2x + 75)/3 = 65$이다. 이를 풀면 $x = 60$이 된다.

계산 과정

B학교 학생 수를 n명, A학교의 평균 점수를 x점이라고 하자. A학교 학생 수는 2n명이므로 두 학교 전체 학생 수는 3n명이다.

평균은 총점수를 총인원으로 나눈 값이므로 $65 = \frac{2n \cdot x + n \cdot 75}{3n} = \frac{2x + 75}{3}$이다. A학교와 B학교의 인원이 다르므로 두 학교 평균을 같은 비중으로 평균 내면 안 된다.

양변에 3을 곱하면 $195 = 2x + 75$이고, 따라서 $2x = 120$, $x = 60$이다. A학교 평균은 **60점**이므로 정답은 ③이다.

선택지의 값을 같은 식에 대입하면 ① 50점은 전체 평균 약 58.33점, ② 55점은 약 61.67점, ③ 60점은 65점, ④ 62점은 약 66.33점을 만든다. 주어진 전체 평균 65점을 만족하는 것은 ③뿐이다.

NoteBase

문 24. 가설검정에서 유의수준과 유의확률에 대한 설명으로 옳지 않은 것은?

① 유의수준은 검정통계량의 관측값과 무관하며, 유의확률은 유관하다.

② 유의수준은 제1종 오류를 범할 확률의 최대 허용값으로 정한다.

③ 유의확률이 작을수록 관측된 자료가 귀무가설로부터 벗어난 정도가 큼을 의미한다.

④ 유의확률은 검정통계량의 대립가설 하에서의 분포로부터 구해진다.

정답 4번

간단한 해설

유의확률은 **귀무가설이 참이라고 가정**했을 때 관측값만큼 또는 더 극단적인 검정통계량이 나올 확률이다. 대립가설하 분포에서 계산하는 값이 아니다.

선지별 해설

① O: 유의수준은 자료를 분석하기 전에 정하는 기각 판단의 기준이므로 실제 관측된 검정통계량 값에 따라 달라지지 않는다. 반면 유의확률은 **관측된 검정통계량을 기준으로 계산**하므로 자료에 따라 달라진다.

② O: 제1종 오류는 참인 귀무가설을 기각하는 오류이다. 유의수준은 이 오류 확률이 넘지 않도록 허용하는 상한으로 정한다. 특히 이산분포를 이용하는 검정 등에서는 실제 제1종 오류 확률이 설정한 유의수준보다 작을 수도 있다.

③ O: 유의확률이 작다는 것은 귀무가설이 참일 때 관측값만큼 또는 더 극단적인 결과가 나타날 확률이 작다는 뜻이다. 따라서 검정통계량을 기준으로 자료가 귀무가설과 더 양립하기 어렵다는 증거가 된다. 다만 작은 유의확률 자체가 효과의 크기가 크다거나 귀무가설이 참일 확률이 작다는 뜻은 아니다.

④ X: 유의확률은 대립가설이 아니라 **귀무가설하의 검정통계량 분포**로 구한다. 이 분포에서 대립가설의 방향에 따라 관측값 이상으로 극단적인 영역의 확률을 계산한다. 단측·양측 여부는 극단적인 영역을 정하는 데 관여하지만 기준 분포를 대립가설하 분포로 바꾸지는 않는다. 따라서 옳지 않은 설명은 ④이다.

NoteBase

문 25. 특정 자료에 대해 단순회귀모형 $Y_i = \alpha + \beta x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, \cdots, n$을 적합한 결과가 다음과 같을 때, 상관계수는?

$$\hat{y} = -10x + 5, \quad \sum_{i=1}^{n} (y_i - \bar{y})^2 = 100, \quad \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 36$$

① 0.64

② 0.80

③ −0.64

④ −0.80

정답 4번

간단한 해설

결정계수는 $R^2 = 1 - \frac{36}{100} = 0.64$이므로 상관계수의 절댓값은 0.80이다. **회귀직선의 기울기가 음수**이므로 상관계수는 −0.80이다.

계산 과정

절편이 포함된 단순선형회귀를 최소제곱법으로 적합하면 결정계수는 표본상관계수의 제곱과 같다. 즉 $R^2 = r^2$이며, 상관계수의 부호는 추정 회귀계수의 부호와 일치한다.

제시된 총제곱합은 SST=Σ_i=1^n(y_i- y)^2=100, 잔차제곱합은 SSE=Σ_i=1^n(y_i-"0177_i)^2=36이다. 따라서 회귀가 설명하는 제곱합은 $SSR = 100 - 36 = 64$이다.

결정계수는 $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{36}{100} = 0.64$이다. 이는 설명된 변동의 비율이며, 상관계수 자체가 아니다. 상관계수의 절댓값은 $|r| = \sqrt{0.64} = 0.80$이다.

추정 회귀직선 "0177=-10x+5의 기울기는 −10이다. β=r(s_y)/(s_x)에서 표준편차의 비는 양수이므로 기울기가 음수이면 상관계수도 음수이다. 절편 5는 상관계수의 부호를 결정하지 않는다.

따라서 상관계수는 $r = -0.80$으로 ④에 해당한다. 0.64는 결정계수이고, 0.80은 부호를 반영하기 전의 절댓값이다.

NoteBase

문 26. 사건 A와 B에 대한 확률이 다음과 같이 주어졌다.
$P(A) = 0.7,\ P(B) = 0.6,\ P(A \cap B) = 0.4$
이때 조건부확률 $P(B^c \mid A^c)$는?

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

정답 2번

간단한 해설

A와 B가 모두 일어나지 않을 확률은 0.1이고, **조건 사건인 A의 여사건 확률**은 0.3이다. 따라서
$P(B^c \mid A^c) = \frac{0.1}{0.3} = \frac{1}{3}$이다.

계산 과정

조건부확률 $P(B^c \mid A^c)$는 A가 일어나지 않았다는 조건 아래에서 B도 일어나지 않을 확률이다. 따라서 **분모는 조건 사건 A의 여사건 확률**이다.
$P(A^c) = 1 - P(A) = 1 - 0.7 = 0.3$

분자는 A와 B가 모두 일어나지 않을 확률이다. 드모르간 법칙으로 $A^c \cap B^c = (A \cup B)^c$이고, 포함배제 공식으로
$P(A \cup B) = 0.7 + 0.6 - 0.4 = 0.9$
따라서 $P(A^c \cap B^c) = 1 - 0.9 = 0.1$이다.

조건부확률의 정의에 대입하면
$P(B^c \mid A^c) = \frac{P(A^c \cap B^c)}{P(A^c)} = \frac{0.1}{0.3} = \frac{1}{3}$
주어진 교집합 확률을 이용하는 계산이므로 A와 B의 독립성을 별도로 가정할 필요가 없다.

구한 값은 $\frac{1}{3}$이므로 ②에 해당한다.

NoteBase

문 27. 크기 5인 다음의 자료로부터 최소제곱법을 사용하여 단순회귀모형 $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$, ε_i는 서로 독립을 적합한 결과 추정된 회귀식이 $\hat{y} = 11.8 - 1.6x$로 얻어졌다. σ^2의 불편추정값은?

x	1	2	3	4	5
y	10	9	7	5	4

① $\frac{2}{5}$

② $\frac{2}{15}$

③ $\frac{2}{25}$

④ $\frac{2}{35}$

정답 2번

간단한 해설

정답은 ②이다. 잔차제곱합은 0.4이고, 절편과 기울기를 추정했으므로 잔차 자유도는 3이다. 따라서 **오차분산의 불편추정값은** 0.4/3 = 2/15이다.

계산 과정

절편과 기울기를 추정하는 단순선형회귀에서 오차분산의 불편추정량은 $s^2 = SSE/(n-2)$이다. 여기서 SSE=Σ_i=1^n(y_i-"0177_i)^2이며, 두 회귀계수를 추정하므로 잔차 자유도는 $5 - 2 = 3$이다.

표의 $x = 1, 2, 3, 4, 5$를 회귀식 "0177=11.8-1.6x에 대입하면 예측값은 차례로 10.2, 8.6, 7.0, 5.4, 3.8이다. 관측값 10, 9, 7, 5, 4에서 예측값을 빼면 잔차는 **-0.2, 0.4, 0, -0.4, 0.2**이다.

잔차제곱합은
aligned&SSE=(-0.2)^2+0.4^2+0^2+(-0.4)^2+0.2^2&=0.04+0.16+0+0.16+0.04=0.4=2/5aligned
이다.

따라서 $s^2 = (2/5)/3 = 2/15$이며 정답은 ②이다. ①의 2/5는 잔차제곱합 자체이고, ③의 2/25는 이를 표본크기 5로 나눈 값으로 불편추정값이 아니다. ④의 2/35 역시 잔차 자유도 3을 사용한 결과와 다르다.

NoteBase

문 28. $X_1, X_2, \ldots, X_{11}$**은 정규분포** $N(\mu_1, \sigma^2)$**으로부터의 확률표본이고** $Y_1, Y_2, \ldots, Y_9$**은 정규분포** $N(\mu_2, \sigma^2)$**으로부터의 확률표본이며,** $X_1, X_2, \ldots, X_{11}$**과** $Y_1, Y_2, \ldots, Y_9$**은 서로 독립이다. 두 확률표본의 분산을 각각** S_1^2, S_2^2**이라 할 때,** $\mu_1 - \mu_2$**에 대한 95% 신뢰구간의 길이는? 단,** $t_\alpha(k)$**는 자유도** k**인** t **분포의 제100**$(1-\alpha)$**백분위수이다.**

① $2 \times t_{0.025}(18)\sqrt{\left(\frac{10S_1^2 + 8S_2^2}{18}\right)\left(\frac{1}{11} + \frac{1}{9}\right)}$
② $t_{0.025}(18)\sqrt{\left(\frac{10S_1^2 + 8S_2^2}{18}\right)\left(\frac{1}{11} + \frac{1}{9}\right)}$
③ $2 \times t_{0.025}(18)\sqrt{\left(\frac{10S_1^2 + 8S_2^2}{18}\right)\left(\frac{1}{10} + \frac{1}{8}\right)}$
④ $t_{0.025}(18)\sqrt{\left(\frac{10S_1^2 + 8S_2^2}{18}\right)\left(\frac{1}{10} + \frac{1}{8}\right)}$

정답 1번

간단한 해설

정답은 ①이다. 합동표본분산은 $(10S_1^2 + 8S_2^2)/18$, 자유도는 18이다. 표준오차에는 1/11 + 1/9를 사용하고, 신뢰구간의 전체 길이는 **임곗값×표준오차의 두 배**이다.

계산 과정

두 모집단은 정규분포이고 분산이 같은 σ^2이며, 두 표본은 독립이다. 따라서 공통분산을 합동표본분산으로 추정하는 두 표본 t 신뢰구간을 사용한다. 표본 크기는 $n_1 = 11, n_2 = 9$이다.

합동표본분산은 $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} = \frac{10S_1^2 + 8S_2^2}{18}$이며 자유도는 $11 + 9 - 2 = 18$이다.

두 표본평균 차이의 추정 표준오차는 SE=S_p^2≤ft(1n_1+1n_2)=≤ft((10S_1^2+8S_2^2)/(18))≤ft(111+19)이다. 평균의 분산에는 표본 크기 11과 9를 사용하며 자유도 10과 8을 넣지 않는다.

95% 신뢰구간은 (X- Y)± t_0.025(18)SE이다. 전체 길이는 상한에서 하한을 뺀 값이므로 **오차한계의 두 배**인 $2t_{0.025}(18)SE$이다.

따라서 길이는 2t_0.025(18)≤ft((10S_1^2+8S_2^2)/(18))≤ft(111+19)로 정답은 ①이다. ②는 반길이만 제시했고, ③은 표준오차에서 표본 크기 대신 자유도를 사용했으며, ④는 두 오류를 모두 포함한다.

NoteBase

문 29. 주머니 속에 흰 공 5개와 검은 공 2개가 들어 있다. A, B 두 사람이 게임을 하는데, A부터 시작해서 교대로 공을 하나씩 무작위로 비복원추출하여 처음으로 검은 공을 꺼내는 사람이 승리하는 것으로 정하였다. A가 이길 확률은?

① $\frac{3}{7}$
② $\frac{1}{2}$
③ $\frac{15}{28}$
④ $\frac{4}{7}$

정답 4번

간단한 해설

A가 이기려면 **첫 검은 공이 1·3·5번째에 나와야 한다**. 각 경우의 확률을 더하면 $\frac{2}{7}+\frac{4}{21}+\frac{2}{21}=\frac{4}{7}$이다.

계산 과정

A는 1, 3, 5, 7번째 차례에 공을 뽑는다. 그러나 흰 공이 5개뿐이므로 첫 검은 공은 늦어도 6번째에 나온다. 따라서 A가 이기는 경우는 첫 검은 공이 1, 3, 5번째에 나오는 세 경우이다.

첫 번째에 검은 공을 뽑을 확률은 $P_1=\frac{2}{7}$이다.

세 번째에 처음 검은 공이 나오려면 앞의 두 공이 모두 흰 공이어야 한다. 비복원추출이므로 남은 공의 수를 순서대로 반영하면 $P_3=\frac{5}{7}\times\frac{4}{6}\times\frac{2}{5}=\frac{4}{21}$이다.

다섯 번째에 처음 검은 공이 나오려면 앞의 네 공이 모두 흰 공이어야 한다. 따라서 $P_5=\frac{5}{7}\times\frac{4}{6}\times\frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}=\frac{2}{21}$이다. 네 번의 흰 공 추출 후에는 흰 공 1개와 검은 공 2개가 남는다.

세 경우는 첫 검은 공의 위치가 서로 달라 배반사건이다. 따라서 P(A 승리)=P_1+P_3+P_5=(2)/(7)+(4)/(21)+(2)/(21)=(4)/(7)이다.

계산 결과 $\frac{4}{7}$에 해당하는 선택지는 ④이다.

NoteBase

문 30. 불량품이 만들어질 확률이 0.002인 공정에서 100개의 제품을 생산할 때, 만들어지는 불량품의 수를 X라 하자. 확률 $P(X \geq 2)$를 포아송분포 근사를 이용하여 구하면?

① $1 - 1.2 \times e^{-0.2}$

② $1 - e^{-0.2}$

③ $1.2 \times e^{-0.2}$

④ $1.22 \times e^{-0.2}$

정답 1번

간단한 해설

포아송 근사의 평균은 $\lambda = np = 0.2$이다. 2개 이상일 확률은 전체 확률에서 **0개와 1개일 확률을 빼서** 구하므로 $1 - 1.2e^{-0.2}$로 근사된다.

계산 과정

각 제품의 불량 여부를 서로 독립이고 불량 확률이 일정한 시행으로 보면, 불량품 수 X는 시행 횟수 n = 100, 성공 확률 p = 0.002인 이항분포를 따른다. 시행 횟수가 많고 불량 확률이 작으므로 문제의 지시대로 평균이 같은 포아송분포로 근사한다. 이때 모수는 **기대 불량품 수**인 $\lambda = np = 100 \times 0.002 = 0.2$이다.

근사에 쓰는 확률변수를 Y라 하면 $Y \sim \text{Poisson}(0.2)$이고, 확률질량함수는 $P(Y = k) = e^{-0.2}\frac{0.2^k}{k!}$이다. 따라서 불량품이 0개일 확률은 $P(Y = 0) = e^{-0.2}$, 1개일 확률은 $P(Y = 1) = 0.2e^{-0.2}$이다.

'2개 이상'의 여사건은 **0개 또는 1개**이다. 그러므로

aligned&P(X≥2)≈ P(Y≥2)=1&-P(Y=0)-P(Y=1)=1-1.2e^-0.2aligned

이는 ①과 일치한다. 원래 이항확률과의 정확한 등식이 아니라 포아송 근삿값이다.

NoteBase

문 31. 연속형 자료의 n개 관측값을 아래 표와 같이 적절히 나누어진 k개의 각 구간에 몇 개의 관측값이 속하는지를 빈도로 요약하였다. 각 구간의 빈도는 충분히 크다고 가정한다. 위의 자료를 이용하여 다음 가설에 대한 유의수준 α인 검정을 수행하고자 한다. 여기서 μ와 σ^2은 미지의 모수이다.

H_0: 자료가 $N(\mu, \sigma^2)$을 따른다.
H_1: H_0가 아니다.

이때 (검정법, 임곗값)으로 올바르게 짝지어진 것은?

단, $\chi^2_\alpha(r)$과 $F_\alpha(r_1, r_2)$는 각각 자유도가 r인 카이제곱분포와 자유도가 (r_1, r_2)인 F분포의 제100(1-α) 백분위수이다.

구간	I_1	I_2	$\cdots$	I_k	합계
관측 빈도	n_1	n_2	$\cdots$	n_k	n

① (카이제곱검정, $\chi^2_\alpha(k-1)$)
② (카이제곱검정, $\chi^2_\alpha(k-3)$)
③ (F 검정, $F_\alpha(1, k-1)$)
④ (F 검정, $F_\alpha(1, k-3)$)

정답 2번

간단한 해설

정답은 ②이다. 구간별 관측빈도와 정규분포의 기대빈도를 비교하므로 카이제곱 적합도검정을 쓴다. 두 미지모수 μ, σ^2를 추정하므로 자유도는 $k-1-2=k-3$이다.

NoteBase

핵심 풀이

관측값을 k개의 구간으로 나눈 빈도가 주어졌으므로, 정규분포에서 예상되는 구간별 빈도와 실제 빈도를 비교하는 카이제곱 적합도검정을 사용한다. 두 분산의 비를 비교하는 F 검정이 아니다.

귀무가설 아래 구간 I_i에 들어갈 확률을 추정한 값을 p_i라 하면 기대빈도는 E_i=np_i이고, 검정통계량은 $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E_i)^2}{E_i}$이다. 적절한 카이제곱 근사를 위해서는 기대빈도도 충분히 커야 한다.

총빈도의 합이 n으로 고정되어 기본 자유도는 $k-1$이다. 여기에 같은 자료로 정규분포의 미지 모수인 평균 μ와 분산 σ^2 두 개를 추정하므로 자유도를 2 더 차감한다. 따라서 **자유도는** $k-3$이다.

검정통계량이 클수록 정규분포와의 차이가 크므로 오른쪽 꼬리에서 기각한다. 유의수준 α의 임곗값은 $\chi^2_{\alpha}(k-3)$이므로 정답은 ②이다. ①은 두 모수 추정을 반영하지 않았고, ③·④는 검정법을 F 검정으로 잘못 선택하였다.

NoteBase

문 32. 남자 8명과 여자 7명이 있다. 전체 15명 중 9명을 비복원추출할 때, 뽑히는 남자의 수로 가능한 값은 몇 가지인가?

① 7

② 8

③ 9

④ 10

정답 1번

간단한 해설

정답은 ①이다. 남자는 최소 2명, 최대 8명까지 뽑힐 수 있으므로 가능한 값은 $8 - 2 + 1 = 7$**개**이다.

계산 과정

뽑힌 남자의 수를 X라 하자. 여자는 전체 7명뿐이므로 9명을 뽑으면, 여자를 모두 뽑더라도 남자가 **최소** $9 - 7 = 2$**명**은 포함되어야 한다.

남자는 전체 8명이고 비복원추출이므로 같은 사람을 다시 뽑을 수 없다. 따라서 뽑히는 남자의 수는 **최대 8명**이다. 남자 8명을 모두 뽑고 여자 1명을 뽑는 경우에 이 최댓값을 얻는다.

가능한 값은 **2, 3, 4, 5, 6, 7, 8**이다. 각 값에서 필요한 여자 수 $9 - X$는 1명부터 7명 사이이므로 모두 실제로 가능한 구성이다. 정수값의 개수는 $8 - 2 + 1 = 7$개이므로 정답은 ①이다.

문제는 9명을 뽑는 서로 다른 조합의 수가 아니라 **남자의 수가 가질 수 있는 값의 개수**를 묻는다. 최댓값 8을 그대로 답하거나, 0명·1명처럼 여자 수 제한을 만족하지 못하는 경우를 포함하면 안 된다. 따라서 ②·③·④의 8·9·10은 해당하지 않는다.

NoteBase

문 33. 확률분포와 관련된 다음 설명 중 옳지 않은 것은?

① $X \sim N(\mu, \sigma^2)$일 때, $((X-\mu)/\sigma)^2$의 분포는 $\chi^2(1)$이다.

② $T \sim t(r)$일 때, T^2의 분포는 $F(1, r)$이다.

③ $X_1, X_2, \cdots, X_n$이 $N(0, \sigma^2)$으로부터의 확률표본일 때, (Σ _i=1^nX_i^2)/(σ^2)의 분포는 $\chi^2(n)$이다.

④ $X_1, X_2, \cdots, X_n$이 임의의 연속인 모집단으로부터의 확률표본일 때, 표본평균($\bar{X}$)과 표본분산(S^2)은 서로 독립이다.

정답 4번

간단한 해설

표본평균과 표본분산의 독립성은 **정규모집단에서의 확률표본**에 대해 성립한다. 이를 임의의 연속분포로 일반화할 수는 없다.

선지별 해설

① O: 정규확률변수를 표준화한 $Z = \frac{X-\mu}{\sigma}$는 표준정규분포를 따른다. **표준정규확률변수 하나의 제곱**은 자유도 1인 카이제곱분포이므로 $Z^2 \sim \chi^2(1)$이다.

② O: 서로 독립인 $Z \sim N(0,1)$, $U \sim \chi^2(r)$를 사용하면 t 확률변수는 $T = \frac{Z}{\sqrt{U/r}}$로 표현된다. 따라서 $T^2 = \frac{Z^2/1}{U/r}$이다. 분자와 분모가 각각 독립인 카이제곱확률변수를 자유도로 나눈 것이므로 $F(1, r)$ 분포를 따른다.

③ O: 확률표본의 각 값을 σ로 나눈 X_i/σ는 서로 독립인 표준정규확률변수이다. 그 제곱 n개를 더한 $\frac{\sum_{i=1}^{n} X_i^2}{\sigma^2}$는 자유도 n인 카이제곱분포를 따른다. **알려진 모평균 0을 기준으로 한 제곱합**이므로 표본평균을 빼서 얻는 자유도 n−1과 구별한다.

④ X: 표본평균과 표본분산의 독립성은 **정규모집단에서 얻은 확률표본**에 대해 성립하는 중요한 성질이다. 모집단이 연속분포라는 조건만으로는 보장되지 않는다. 또한 두 통계량의 공분산이 0인 경우가 있더라도 그것만으로 독립이라고 결론낼 수 없다.

NoteBase

문 34. 어떤 농구선수의 자유투 성공률 p에 대해 다음의 가설을 검정하려고 한다.
$H_0 : p \le 0.6$ **대** $H_1 : p > 0.6$
이 선수가 자유투를 15회 시도하여 12회 성공했을 때 유의확률은?

① $\binom{15}{12}(0.6)^{12}(0.4)^3$
② $\sum_{x=12}^{15}\binom{15}{x}(0.6)^x(0.4)^{15-x}$
③ $\sum_{x=3}^{15}\binom{15}{x}(0.4)^x(0.6)^{15-x}$
④ $\sum_{x=0}^{3}\binom{15}{x}(0.4)^x(0.6)^{15-x}$

정답 4번

간단한 해설

대립가설이 성공률 증가를 주장하므로 **12회 이상 성공할 확률**을 구한다. 이는 실패확률 0.4인 이항분포에서 실패가 3회 이하일 확률과 같다.

계산 과정

각 자유투가 동일한 성공확률로 독립 시행된다는 이항모형에서 성공 횟수를 S라 하면 $S \sim B(15, p)$이다. 관측 성공 횟수는 12이고 대립가설이 $p > 0.6$이므로 성공 횟수가 클수록 대립가설을 지지한다.

유의확률은 귀무가설 아래에서 관측값만큼 또는 더 극단적인 성공 횟수가 나올 확률이다. p≤0.6 범위에서 $P_p(S \ge 12)$는 p가 클수록 커지므로 경계값 $p = 0.6$을 사용한다. 따라서 유의확률은 $\sum_{s=12}^{15}\binom{15}{s}(0.6)^s(0.4)^{15-s}$이다.

선택지의 형태와 연결하기 위해 실패 횟수를 $F = 15 - S$로 놓는다. 경계값에서 실패확률은 0.4이므로 $F \sim B(15, 0.4)$이고, S≥12는 F≤3과 같은 사건이다. 따라서 유의확률은 $\sum_{x=0}^{3}\binom{15}{x}(0.4)^x(0.6)^{15-x}$로 쓸 수 있다.

①은 정확히 12회 성공할 확률만 포함하여 13~15회 성공을 빠뜨린다.

②는 성공 12회 이하의 확률로 꼬리 방향이 다르다.

③은 실패확률 0.4를 사용하면서 12회 이상을 더하므로 실패가 12회 이상인 사건을 나타낸다.

관측된 성공 12회 이상을 실패 3회 이하로 표현한 식은 ④이다.

NoteBase

문 35. 어떤 회사원의 출근 소요 시간이 평균 40분, 표준편차 5분인 정규분포를 따른다고 한다. 이 사람이 회사에 늦지 않게 도착할 확률이 1-α 이상 되게 하려면 늦어도 업무 시작 시각보다 몇 분 이전에 출발하여야 하는가? 단, 0 < α < 0.5이며, z_α는 표준정규분포의 제100(1-α) 백분위수이다.

① 40+5$z_{\alpha/2}$

② 40+5z_α

③ 40-5z_α

④ 40-5$z_{\alpha/2}$

정답 2번

간단한 해설

정시 도착 확률을 $1-\alpha$ 이상으로 하려면 출근 시간의 **상위 α 꼬리를 제외한 분위수**만큼 시간을 확보해야 한다. 이를 원래 단위로 바꾸면 $40 + 5z_\alpha$분이다.

계산 과정

출근 소요 시간을 T, 업무 시작 전에 확보한 시간을 t분이라 하자. 늦지 않게 도착하는 사건은 $T \le t$이므로 필요한 조건은 P(T≤ t)≥1-α이다. '늦어도' 출발해야 하는 시각은 이 조건을 만족하는 최소 t를 구해 업무 시작 시각에서 빼면 된다.

평균은 40분, 표준편차는 5분이므로 $Z = \frac{T-40}{5} \sim N(0, 1)$이다. 따라서
$P(T \le t) = P\left(Z \le \frac{t-40}{5}\right)$
문제에서 z_α는 $P(Z \le z_\alpha) = 1-\alpha$를 만족하는 분위수이다. 표준정규분포의 누적확률은 증가하므로 $\frac{t-40}{5} \ge z_\alpha$, 즉 t≥40+5z_α이어야 한다.

지각이라는 **한쪽 꼬리의 확률만 α 이하**로 제한하므로 양측 구간에 쓰는 $z_{\alpha/2}$를 사용하지 않는다. 또한 0α0.5이므로 z_α0이며, 평균 소요 시간 40분보다 더 여유를 두어야 한다. 따라서 늦어도 업무 시작 $40 + 5z_\alpha$분 전에 출발해야 하며, ②와 일치한다.

NoteBase

문 36. 평균이 5로 알려져 있고 분산은 알려져 있지 않은 정규모집단으로부터의 크기 15인 확률표본을 $X_1, X_2, \cdots, X_{15}$라 할 때, 모분산 σ^2에 대한 90% 신뢰구간은? 단, $\chi^2_\alpha(k)$는 자유도 k인 카이제곱분포의 제100$(1-\alpha)$ 백분위수이며, $\overline{X}$는 표본평균이다.

① ≤ft((Σ_i=1^15(X_i-5)^2)/(^2_0.05(15)),(Σ_i=1^15(X_i-5)^2)/(^2_0.95(15)))

② ≤ft((Σ_i=1^15(X_i-5)^2)/(^2_0.05(14)),(Σ_i=1^15(X_i-5)^2)/(^2_0.95(14)))

③ ≤ft(Σ_i=1^15(X_i-X)^2^2_0.05(15),Σ_i=1^15(X_i-X)^2^2_0.95(15))

④ ≤ft(Σ_i=1^15(X_i-X)^2^2_0.05(14),Σ_i=1^15(X_i-X)^2^2_0.95(14))

정답 1번

간단한 해설

평균이 5로 알려져 있으므로, 5를 중심으로 한 편차제곱합을 모분산으로 나누면 **자유도 15인 카이제곱분포**를 따른다. 중앙확률 90%를 모분산에 대해 풀면, 하한에는 큰 분위수인 $\chi^2_{0.05}(15)$, 상한에는 작은 분위수인 $\chi^2_{0.95}(15)$가 분모에 들어간다.

핵심 풀이

알려진 평균을 이용한 제곱합을 $Q=\sum_{i=1}^{15}(X_i-5)^2$라 하자. 정규모집단에서 독립적으로 뽑은 표본이므로 $Z_i=(X_i-5)/\sigma$는 서로 독립인 표준정규확률변수이다. 따라서 $W=Q/\sigma^2=\sum_{i=1}^{15}Z_i^2$는 자유도 15인 카이제곱분포를 따른다. **평균을 표본에서 추정하지 않았으므로** 자유도를 하나 줄이지 않는다.

문제의 $\chi^2_\alpha(k)$는 누적확률이 $1-\alpha$인 분위수이다. $a=\chi^2_{0.95}(15)$는 하위 5% 분위수이고, $b=\chi^2_{0.05}(15)$는 하위 95% 분위수이므로 $a<b$이다. 양쪽 꼬리에 각각 5%를 남기면 $P(a<W<b)=0.90$이 된다.

여기에 $W=Q/\sigma^2$를 넣으면 $a<Q/\sigma^2<b$이다. 양수인 모분산에 대해 정리하면 $Q/b<\sigma^2<Q/a$를 얻는다. 따라서 하한은 $Q/\chi^2_{0.05}(15)$, 상한은 $Q/\chi^2_{0.95}(15)$이며, 이는 ①의 구간과 같다. **큰 분위수로 나눈 값이 하한**이 되는 것은 역수를 취할 때 크기 관계가 바뀌기 때문이다.

표본평균을 중심으로 한 제곱합은 분포가 다르다. $R=\sum_{i=1}^{15}(X_i-\overline{X})^2$라 하면 R/σ^2는 자유도 14인 카이제곱분포를 따른다. 따라서 **제곱합의 중심과 자유도를 맞추어야 한다.** 5를 중심으로 한 제곱합에 자유도 14를 쓰거나, 표본평균을 중심으로 한 제곱합에 자유도 15를 쓰면 해당 분포와 맞지 않는다. ④처럼 R과 자유도 14를 사용해도 별도의 유효한 90% 신뢰구간을 만들 수 있으며, ①은 알려진 평균 5를 직접 활용하여 얻은 구간이다.

NoteBase

문 37. 과거 자료를 분석한 결과, 두 종류 주가(A, B)의 일별 변화(움직임)에 대해 다음과 같은 확률분포표를 얻었다. 오늘 주가 A가 상승했다는 것을 알고 있을 때, 주가 B가 하락했을 확률은?

		주가 B	
		하락	상승
주가 A	하락	0.10	0.30
	상승	0.45	0.15

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

정답 2번

간단한 해설

정답은 ② **34**이다. A가 상승할 확률은 0.45 + 0.15 = 0.60이고, A 상승과 B 하락의 결합확률은 0.45이다. 따라서 조건부확률은 0.45/0.60 = 3/4이다.

계산 과정

A가 상승한 사건을 U, B가 하락한 사건을 D라고 하자. 주어진 조건이 A의 상승이므로 구할 값은 $P(D \mid U)$이다.

표에서 A 상승 행의 두 확률은 B 하락에 대해 0.45, B 상승에 대해 0.15이다. 따라서
$P(U) = 0.45 + 0.15 = 0.60$이고, A 상승과 B 하락이 함께 일어날 확률은 $P(D \cap U) = 0.45$이다.

조건부확률 공식에 따라 P(D U)=(P(D U))/(P(U))=(0.45)/(0.60)=34이다. 따라서 정답은 ②이며, ①의 2/3, ③의 3/5, ④의 4/5는 구하는 조건부확률과 다르다.

③의 3/5 = 0.60은 A가 상승할 확률 자체이다. 문제는 이미 A가 상승했다는 조건 아래에서 B가 하락할 비율을 묻기 때문에 0.45를 0.60으로 나누어야 한다. B 하락 열의 합 0.55를 분모로 쓰면 조건의 방향이 바뀐다.

NoteBase

문 38. 크기 n인 자료로부터 최소제곱법을 사용하여 단순회귀모형 $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$를 적합한 결과 추정된 회귀식이 $\hat{y} = 0.2 + 1.6x$로 얻어졌다. 변환 $x_i^* = 2x_i$ $(i = 1, 2, \cdots, n)$를 통해 새로운 변수 x^*를 생성한 뒤 x^*를 독립변수, Y를 종속변수로 하는 단순회귀모형을 최소제곱법으로 적합하였을 때 얻어지는 회귀식은?

① $\hat{y} = 0.4 + 3.2x^*$

② $\hat{y} = 0.2 + 3.2x^*$

③ $\hat{y} = 0.4 + 0.8x^*$

④ $\hat{y} = 0.2 + 0.8x^*$

정답 4번

간단한 해설

$x^* = 2x$이므로 $x = \frac{x^*}{2}$를 기존 회귀식에 대입한다. **절편은 그대로이고 기울기는 절반**이 되어 "0177=0.2+0.8x^*이다.

계산 과정

변환은 설명변수의 척도만 두 배로 바꾸고 종속변수 Y는 그대로 두는 것이다. $x^* = 2x$이므로 기존 변수를 새 변수로 나타내면 $x = \frac{x^*}{2}$이다.

기존 회귀식에 대입하면 "0177=0.2+1.6≤ft((x^*)/(2))=0.2+0.8x^*이다. 새 설명변수의 한 단위 변화는 기존 설명변수의 절반 단위 변화이므로 기울기는 절반이 된다.

이 대입식이 다시 최소제곱법을 적용한 결과와 같은 이유는, 0이 아닌 상수배 변환으로 표현 가능한 직선의 집합이 바뀌지 않기 때문이다. 같은 관측치에서 동일한 예측값과 잔차를 만들 수 있으므로 최소 잔차제곱합도 그대로이다. 또한 $x^* = 0$과 $x = 0$은 같은 위치이므로 절편은 0.2로 유지된다.

따라서 새 회귀식은 "0177=0.2+0.8x^*로 ④에 해당한다. 기울기를 3.2로 바꾸는 것은 척도 변환의 방향을 뒤집은 것이고, 절편을 0.4로 바꾸는 것은 변환하지 않은 Y까지 두 배로 한 것과 혼동한 것이다.

NoteBase

문 39. 크기 n인 자료로부터 최소제곱법을 사용하여 아래의 단순회귀모형을 적합하였다.
Y_i=_0+_1x_i+_i, _i N(0,σ^2), _i는 서로 독립
다음 중 옳은 것만을 모두 고른 것은?

가. $\hat{\beta}_1$은 β_1의 불편추정량이다.
나. $\hat{\beta}_0$은 β_0의 불편추정량이다.
다. 독립변수의 값이 x로 주어질 때 Y의 기댓값에 대한 신뢰구간의 길이는 x에 무관하다.

① 가, 나
② 가, 다
③ 나, 다
④ 가, 나, 다

정답 1번

간단한 해설

오차의 평균이 0인 주어진 회귀모형에서 기울기와 절편의 최소제곱추정량은 각각 불편추정량이다. 평균반응의 신뢰구간은 **x가 표본평균에서 멀어질수록 넓어지므로** x와 무관하지 않다.

보기별 해설

가. O: S_xx=Σ_i=1^n(x_i- x)^2>0이라 두 회귀계수를 추정할 수 있다고 하자. 기울기의 최소제곱추정량은 β_1=_1+(Σ_i=1^n(x_i- x)_i)/(S_xx)로 표현된다. **오차의 기댓값이 0**이므로 추가된 항의 기댓값도 0이며, E(β_1)=_1이다.

나. O: 절편의 최소제곱추정량은 β_0= Y-β_1 x이다. E(Y)=_0+_1 x이고 기울기 추정량이 불편하므로, E(β_0)=_0이다. 따라서 절편 추정량도 불편추정량이다.

다. X: 주어진 x에서 평균반응의 추정값은 β_0+β_1x이며 그 표준오차는 s(1)/(n)+((x- x)^2)/(S_xx)이다. 여기서 $s^2 = \frac{SSE}{n-2}$는 오차분산 추정값이다. 같은 신뢰수준에서 신뢰구간의 길이는 이 표준오차에 비례하므로 **x가 표본평균에 가까울수록 짧고 멀수록 길다**. 개별 새 관측값의 예측구간이 아니라 평균반응의 신뢰구간이라는 점도 구별해야 한다.

따라서 옳은 보기는 가, 나이며 ①에 해당한다.

NoteBase

문 40. 공정한 3개의 육면체 주사위를 던질 때 나오는 수의 합이 6일 확률은?

① $\frac{1}{6}$
② $\frac{5}{54}$
③ $\frac{5}{108}$
④ $\frac{1}{36}$

정답 3번

간단한 해설

합이 6인 눈의 종류는 (1, 1, 4), (1, 2, 3), (2, 2, 2)이다. **주사위를 구별한 배열**은 각각 3·6·1가지이므로 확률은 $\frac{10}{216} = \frac{5}{108}$이다.

계산 과정

세 주사위를 첫째·둘째·셋째로 구별하면 각 결과는 순서 있는 세 수로 나타난다. 각 주사위가 독립적으로 공정하게 나오므로 전체 결과는 $6^3 = 216$가지이며 모두 같은 확률을 갖는다.

합이 6이 되는 수를 작은 순서로 정리하면 (1, 1, 4), (1, 2, 3), (2, 2, 2)의 세 종류뿐이다. 다만 주사위는 구별되므로 **각 종류의 순열 수**를 세어야 한다.

(1, 1, 4)는 4가 나오는 주사위를 고르는 3가지이다.

(1, 2, 3)은 서로 다른 세 수의 배열이므로 3! = 6가지이다.

(2, 2, 2)는 모든 눈이 같으므로 1가지이다.

따라서 합이 6인 결과는 3 + 6 + 1 = 10가지이다.

등확률인 결과 중 조건을 만족하는 경우의 비율을 구하면
P(합이 6)=(10)/(216)=(5)/(108)
이는 ③과 일치한다.

NoteBase