

통계학개론

문 1. 다음은 반복수가 같은 일원배치 분산분석을 실시하여 얻은 결과의 일부이다. 이 결과에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	F -값	p -값
처리	2000	4		0.0027
잔차	3500	35		

- ① 처리의 수는 5이다.
- ② 각 처리에서의 반복수는 10이다.
- ③ 유의수준 0.01에서 처리효과 간에 유의한 차이가 있다.
- ④ F -값은 5이다.

문 2. 분산분석법에 대한 설명으로 옳지 않은 것은?

- ① 일원배치 분산분석에서 인자(factor)의 각 수준에서의 반복수는 다를 수 있다.
- ② 반복이 있는 이원배치 분산분석에서 두 인자의 교호작용(interaction)제곱합의 자유도는 두 인자의 제곱합의 자유도를 곱한 것과 같다.
- ③ 인자가 2개이면 주로 t -검정을 실시하지만 인자가 3개 이상이면 일원배치 분산분석을 이용하여 인자들을 비교한다.
- ④ 반복수 $r(r \geq 2)$ 이 동일한 이원배치 분산분석에서 잔차제곱합의 자유도는 교호작용제곱합의 자유도보다 크다.

문 3. 수술 후 완치율(p)이 30%로 알려져 있는 어느 암에 걸린 환자 60명을 대상으로 수술과 방사선 치료를 병행하였더니, 30명이 완치되었다고 한다. 이 결과로부터 수술과 방사선 치료를 병행하는 것이 수술 후 완치율을 높이는지 검정하고자 할 때, 가장 적절한 귀무가설(H_0)과 대립가설(H_1)은?

- ① $H_0 : p = 0.3, H_1 : p > 0.3$
- ② $H_0 : p = 0.3, H_1 : p \neq 0.3$
- ③ $H_0 : p = 0.5, H_1 : p > 0.5$
- ④ $H_0 : p = 0.5, H_1 : p \neq 0.5$

문 4. 다음은 두 변수 X 와 Y 에 대하여 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, 2, \dots, n$ 의 기본가정을 전제로 하여 분석했을 때 출력된 분산분석표이다. 이 결과로부터 얻은 결론으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	F -값	p -값
회귀	100	1	100	33.33	0.0004
잔차	24	8	3		
계	124	9			

- ① 오차항(ϵ_i)의 분산의 불편추정값은 3이다.
- ② 추정된 회귀식은 유의수준 0.05에서 유의하다.
- ③ 위 분석에 사용된 자료의 크기(n)는 9이다.
- ④ 결정계수는 $\frac{100}{124}$ 이다.

문 5. 두 정규모집단 A와 B에서 각각 표본을 취해 구한 표본분산값이 각각 $S_A^2 = 47.3, S_B^2 = 36.4$ 이다. A집단의 모분산 σ_A^2 이 B집단의 모분산 σ_B^2 보다 크다고 할 수 있는지를 유의수준 5%로 검정하려고 할 때 가장 알맞은 검정방법은?

- ① F -검정
- ② t -검정
- ③ Z -검정
- ④ χ^2 -검정

문 6. 상자 A에 빨간 공 2개와 파란 공 4개가 들어 있다. 상자 A에서 3개의 공을 랜덤하게 비복원추출하여 비어 있는 상자 B에 넣었다. 상자 B에서 공 1개를 랜덤하게 꺼냈을 때 그 공이 빨간 공이었다면, 빨간 공 2개와 파란 공 1개가 상자 B로 옮겨졌을 확률은?

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

문 7. 어떤 음료수 공장에서 생산되는 음료수의 용량은 평균 μ , 분산 16인 확률분포를 따른다고 한다. 이 공장에서 n 개의 음료수를 독립적으로 랜덤하게 추출하여 표본평균 $\bar{X}$ 를 구할 때, $\bar{X}$ 의 기댓값은 200, 표준편차는 0.2였다. $\mu + n$ 의 값은?

- ① 220
- ② 280
- ③ 450
- ④ 600

문 8. 어느 경제학자가 조사한 100가구의 연간소득(X , 단위: 100만원)과 연간식료품지출액(Y , 단위: 10만원)에 대해 회귀분석을 수행한 결과 추정된 회귀식이 $\hat{Y} = 12.9 + 0.8X$ 이다. 옳지 않은 것은? (단, 모든 항목은 100가구의 연간소득 범위 안에서 고려한다)

- ① 종속변수는 연간식료품지출액이고 독립변수는 연간소득이다.
- ② 연간소득과 연간식료품지출액의 표본상관계수는 0.8이다.
- ③ 연간소득 3,000만원인 가구의 연간식료품지출액의 기댓값은 369만원으로 추정된다.
- ④ 연간소득이 100만원 증가하면 연간식료품지출액은 평균적으로 8만원 증가하는 것으로 추정된다.

문 9. 어느 지역 주민의 수축기 혈압이 평균 120이고 표준편차 20인 정규분포를 따른다고 한다. 이 지역 주민 중 랜덤하게 선택된 25명의 수축기 혈압 평균을 $\bar{X}$ 라 할 때, $P(\bar{X} \geq x) = 0.05$ 인 x 의 값은? (단, 아래의 표는 Z 가 표준정규분포를 따르는 확률변수일 때 $P(Z \leq z)$ 의 값에 상응하는 z 의 값을 나타낸 것이다)

$P(Z \leq z)$	z
0.010	-2.33
0.025	-1.96
0.050	-1.65
0.075	-1.44
0.100	-1.28

- ① 112.16
- ② 113.40
- ③ 126.60
- ④ 127.84

문 10. 멘델의 이론에 따르면, 노란색 순종 완두와 녹색 순종 완두를 교배하면 노란색 완두와 녹색 완두가 3:1의 비율로 생산된다고 한다. “멘델의 이론이 맞다.”라는 귀무가설을 검정하기 위해 100번의 교배실험을 하여 노란색 완두 80개와 녹색 완두 20개를 얻었다. 카이제곱검정통계량 값(㉠)과 유의수준 0.05에서 귀무가설의 기각 여부(㉡)를 바르게 연결한 것은? (단, $\chi^2_{\alpha}(k)$ 는 자유도가 k 인 카이제곱분포의 $(1-\alpha) \times 100$ 번째 백분위수를 나타내고, $\chi^2_{0.05}(1)=3.84$, $\chi^2_{0.01}(1)=6.63$ 이다)

㉠ ㉡

- ① $\frac{4}{3}$ 귀무가설 기각
 ② $\frac{4}{3}$ 귀무가설 채택
 ③ $\frac{25}{16}$ 귀무가설 기각
 ④ $\frac{25}{16}$ 귀무가설 채택

문 11. 중심 위치의 측도에 대한 설명으로 옳지 않은 것은?

- ① 산술평균보다 중앙값이 이상값의 영향을 더 많이 받는다.
 ② 산술평균은 자료값들의 합을 자료값의 수로 나눈 것을 의미한다.
 ③ 자료값들 중에 그 자료의 중앙값과 동일한 값이 없을 수도 있다.
 ④ 최빈값은 자료값들 중에서 가장 많이 나타나는 값을 의미한다.

문 12. 절편이 있는 단순선형회귀모형에 최소제곱법을 적용하여 회귀모수를 추정할 때, 옳지 않은 것은?

- ① 잔차의 합은 0이다.
 ② 결정계수는 두 변수의 표본상관계수의 제곱과 같다.
 ③ 오차항의 분산에 대한 불편추정량은 종속변수의 표본분산과 같다.
 ④ 추정회귀직선은 두 변수의 표본평균에 해당하는 점을 지난다.

문 13. 사건 A, B, C 에 대하여 다음 중 옳지 않은 것은?

- ① $P(A \cup B) + P(A \cap B) = P(A) + P(B)$
 ② $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C)$
 ③ 두 사건 A 와 B 가 독립이면 A 와 $\bar{B}$ 도 독립이다. ($\bar{B}$ 는 B 의 여사건)
 ④ 둘씩 독립인 세 사건 A, B, C 가 서로 독립이 아닐 수 있다.

문 14. 0 또는 1의 값만 가질 수 있는 확률변수 X 가 $E(X) = 4Var(X)$ 를 만족할 때 X 의 기댓값은? (단, $Var(X) \neq 0$ 이다)

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

문 15. 어느 의료 기관에서 한 달 동안 내원한 환자들 중 10명을 랜덤추출하여 연령을 조사한 결과가 다음과 같다. 이 결과에 대한 설명으로 옳지 않은 것은? (단, 10명 연령의 합은 400이고, 연령 제곱의 합은 18,286이다)

14, 21, 29, 33, 40, 45, 49, 50, 52, 67

- ① 표본분산은 220보다 크다.
 ② 표본범위는 53이다.
 ③ 표본사분위(수)범위는 20보다 크다.
 ④ 표본변동계수는 0.5보다 크다.

문 16. 확률변수 X 는 시행횟수가 10이고 성공확률이 p 인 이항분포를 따른다. 가설 $H_0: p=0.5$ 대 $H_1: p>0.5$ 의 검정에서 기각역이 $X \geq 9$ 라면, 제1종 오류를 범할 확률은?

- ① $(0.5)^{10}$
 ② $10(0.5)^9$
 ③ $10(0.5)^{10}$
 ④ $11(0.5)^{10}$

문 17. 확률변수 X 와 Y 의 상관계수가 0.3일 때 옳은 것은?

- ① $2X$ 와 Y 의 상관계수는 0.6이다.
 ② $X+3$ 과 Y 의 상관계수는 0.3이다.
 ③ $-X$ 와 Y 의 상관계수는 0.3이다.
 ④ X 와 Y 의 상관계수의 단위는 $(X \text{의 단위})/(Y \text{의 단위})$ 이다.

문 18. 평균 μ , 표준편차 5인 정규모집단에서 크기 25인 랜덤포본을 추출했을 때, 모평균 μ 에 대한 95% 신뢰구간의 길이는? (단, Z 가 표준정규분포를 따르는 확률변수일 때, $P(|Z| < 2) = 0.95$ 로 가정한다)

- ① 1 ② 2
 ③ 3 ④ 4

문 19. 어느 지역에서 60대 이상의 성인 중 흡연 경험이 있는 사람 40명과 흡연 경험이 없는 사람 60명을 랜덤하게 선택하여 심근경색 증세가 있는지를 조사한 결과가 다음과 같다.

구분	심근경색 증세 있음	심근경색 증세 없음	합계
흡연 경험 있음	12	28	40
흡연 경험 없음	8	52	60
합계	20	80	100

심근경색 증세 비율이 흡연 경험 여부에 따라 다른지를 검정하기 위한 카이제곱검정통계량의 값이 4.17로 얻어졌다. 이에 대한 설명으로 옳은 것만을 <보기>에서 모두 고르면? (단, $\chi^2_\alpha(k)$ 는 자유도가 k 인 카이제곱분포의 $(1-\alpha)\times 100$ 번째 백분위수를 나타내고, $\chi^2_{0.05}(1)=3.84$, $\chi^2_{0.01}(1)=6.63$ 이다)

· <보 기>

- ㉑. p -값(p -value 또는 유의확률)은 0.01과 0.05 사이이다.
- ㉒. 유의수준 1%에서 흡연 경험 여부에 따라 심근경색 증세 비율이 다르다고 할 수 있다.
- ㉓. 흡연 경험이 있는 사람 중에서 심근경색 증세가 없다고 조사되었던 사람 중 한 명이 심근경색 증세가 있는 것으로 수정되어 12가 13으로 변경되고 28이 27로 변경되면 카이제곱검정통계량의 값은 커진다.

- ① $\neg$
② $\neg, \perp$
③ $\neg, \sqsubset$
④ $\perp, \sqsubset$

문 20. 어느 지역의 남자 고등학생들의 키는 평균 172 cm, 표준편차 5 cm인 정규분포를 따른다고 한다. 한 명의 남학생을 랜덤하게 선택했을 때, 이 남학생의 키가 170 cm 이하일 확률과 같은 것은? (단, Z 는 표준정규분포를 따르는 확률변수이다)

- ① $P(Z \geq \frac{172-170}{5})$
- ② $P(Z \leq \frac{172-170}{5})$
- ③ $P(Z \geq \frac{170-172}{\sqrt{5}})$
- ④ $P(Z \leq \frac{170-172}{\sqrt{5}})$