

통계학개론

1. 다음은 대통령 선거 출구 조사 방법에 대한 설명이다. 이에 해당하는 표본 추출 기법은?

모든 투표자를 대상으로 조사하는 것이 아니라, 투표자들이 나오는 것을 조사원이 출구 근처에서 보다가 일정 간격으로 투표자를 추출해 조사합니다. 간격이 너무 넓은 경우에는 조사원의 집중력 저하로 조사해야 할 투표자를 놓칠 가능성이 큼니다. 반대로 간격이 좁으면 조사원이 너무 바빠 응답자를 정확하게 선정하지 못할 수 있습니다. 16대 대선의 경우는 간격이 6이었고, 19대 대선은 5로 정해 조사했습니다.

- ① 단순임의추출법(simple random sampling)
- ② 계통추출법(systematic sampling)
- ③ 층화임의추출법(stratified random sampling)
- ④ 집락추출법(cluster sampling)

2. 세 사건 A, B, C 가 서로 독립이고 $P(A) = \frac{1}{4}, P(B) = 3P(C)$,

$P(A \cap B^c \cap C) = \frac{1}{48}$ 일 때, $P(B \cap C)$ 는?

- ① $\frac{1}{12}$
- ② $\frac{5}{12}$
- ③ $\frac{7}{12}$
- ④ $\frac{11}{12}$

3. 확률변수 X, Y 에 대하여 $E[X(Y+1)] = 9, E[Y(X+1)] = 10, E(X+Y) = 7$ 일 때, 공분산 $Cov(X, Y)$ 의 값은?

- ① -6
- ② -5
- ③ 5
- ④ 6

4. ‘두 확률변수 X, Y 의 상관계수가 0’, 즉 ‘ $Corr(X, Y) = 0$ ’과 동치가 아닌 것은?

- ① $Var(X+Y) = Var(X-Y)$
- ② $E(XY) = E(X)E(Y)$
- ③ $Cov(-2X+3, 3Y-2) = 0$
- ④ $Cov(X+Y, X-Y) = 0$

5. 두 연속확률변수 X, Y 의 상관계수에 대한 설명으로 옳지 않은 것은?

- ① 상관계수가 0이면 X, Y 는 독립이다.
- ② 상관계수는 측정단위의 영향을 받지 않는다.
- ③ 상관계수가 -0.7인 경우가 +0.5인 경우보다 선형관계가 강하다.
- ④ X, Y 의 표본상관계수가 0.5일 때, 동일 자료에서 $3X$ 와 $5Y$ 의 표본상관계수도 0.5이다.

6. 정규분포 $N(\mu, \sigma^2)$ 에서 표준편차(σ)를 이용하여 사분위수범위를 나타낸 것으로 옳은 것은? (단, z_α 는 표준정규분포의 제 $100 \times (1-\alpha)$ 백분위수이다)

- ① $z_{0.025} \times \sigma$
- ② $z_{0.25} \times \sigma$
- ③ $2 \times z_{0.025} \times \sigma$
- ④ $2 \times z_{0.25} \times \sigma$

7. 미지의 모평균이 μ 이고 모분산이 σ^2 인 모집단으로부터 확률표본 X_1, X_2, X_3, X_4 를 추출하여 μ 에 대한 점추정량을 다음과 같이 정의하였다. 이 점추정량 중에서 μ 에 대한 불편추정량이며 그중 최소분산을 갖는 것은? (단, $\sigma^2 > 0$ 이다)

- $\hat{\mu}_1 = \frac{1}{5}(X_1 + X_2 + 2X_3 + X_4)$
- $\hat{\mu}_2 = \frac{1}{4}(X_1 + X_2 + 2X_3 + X_4)$
- $\hat{\mu}_3 = \frac{1}{5}(2X_1 + X_2 + X_3 + 2X_4)$
- $\hat{\mu}_4 = \frac{1}{4}(2X_1 + X_2 + X_3)$

- ① $\hat{\mu}_1$
- ② $\hat{\mu}_2$
- ③ $\hat{\mu}_3$
- ④ $\hat{\mu}_4$

8. 확률변수 X 의 확률분포가 다음과 같다.

x	-1	0	1	계
$P(X=x)$	(가)	(나)	0.25	1

다음 항목 중 옳은 것만을 모두 고르면?

- ㄱ. (가)+(나) = $\frac{3}{4}$
- ㄴ. $E(X) = \frac{1}{8}$ 일 때, (나) = $\frac{3}{8}$
- ㄷ. $E(X) = \frac{1}{12}$ 일 때, 위 모집단에서 크기가 2인 확률표본을 추출한다면 표본평균 \bar{X} 의 분산은 $Var(\bar{X}) = \frac{59}{144}$ 이다.

- ① ㄱ
- ② ㄴ
- ③ ㄱ, ㄷ
- ④ ㄴ, ㄷ

9. 독립변수 X 와 종속변수 Y 를 갖는, 절편이 있는 단순선형회귀모형에서 최소제곱법을 적용했을 때, 결정계수에 대한 설명으로 옳은 것은?
- ① 결정계수가 0이면 모든 잔차가 0이다.
 - ② 결정계수가 큰 모형일수록 설명력이 낮은 모형이다.
 - ③ 결정계수는 X, Y 의 표본상관계수를 이용하여 구할 수 있다.
 - ④ 만약 새로운 독립변수를 모형에 추가한다면 결정계수는 감소한다.

10. 어느 도시의 고등학생 중 남학생 100명과 여학생 100명을 무작위 추출한 뒤 수학과 영어 과목에 대한 선호도를 조사하여 다음과 같은 분할표를 얻었다. “성별에 따라 과목 선호도에 차이가 없다.”라는 귀무가설을 검정하기 위한 카이제곱검정통계량의 값이 8일 때, (가)의 값과 귀무가설 하에서의 이 검정통계량의 근사적인 분포를 바르게 연결한 것은? (단, (가) ≥ 50 이다)

성별 \ 과목	수학	영어	계
남학생	(가)		100
여학생			100
계	100	100	200

(가) 근사적인 분포

- ① 55 자유도가 1인 카이제곱분포
- ② 55 자유도가 2인 카이제곱분포
- ③ 60 자유도가 1인 카이제곱분포
- ④ 60 자유도가 2인 카이제곱분포

11. 어느 도시에서 하루에 발생하는 교통사고 건수를 지난 200일 동안 조사하여 다음 표와 같이 정리하였다.

교통사고 건수	0	1	2	3	4	계
해당일 수	90	50	30	20	10	200

이 자료에서 구한 하루 발생 교통사고 건수의 통계량의 값으로 옳지 않은 것은?

- ① 평균은 1.05이다.
- ② 중앙값은 1이다.
- ③ 최빈값은 0이다.
- ④ 제3사분위수는 3이다.

12. 분산이 같은 두 정규모집단 A와 B로부터 크기가 각각 8인 확률 표본을 독립적으로 추출하였더니 표본분산이 각각 19와 13이었다. 두 모평균의 차 $\mu_A - \mu_B$ 에 대한 95% 신뢰구간의 길이는? (단, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $t_{0.025}(7) = 2.365, t_{0.025}(14) = 2.145$ 이다)

- ① 4.29
- ② 4.73
- ③ 8.58
- ④ 9.46

13. 앞면이 나올 확률이 p 인 동전에 대해 가설 $H_0: p = \frac{1}{2}, H_1: p > \frac{1}{2}$ 을 검정하려고 한다. 이 동전을 다섯 번 던져 보니 앞면이 네 번 나왔다면, 유의확률(p -value)은?

- ① $\frac{1}{16}$
- ② $\frac{2}{16}$
- ③ $\frac{3}{16}$
- ④ $\frac{4}{16}$

14. 다음은 연령대별로 스마트폰 과의존 위험군 비율을 조사한 것이다.

연령대	10대 이하	20 ~ 30대	40 ~ 50대	60대 이상
위험군 비율	28%	37%	22%	16%
표본 크기	50명	100명	100명	50명

이 자료를 바탕으로 다음의 가설을 검정하고자 한다.

- H_0 : 연령대에 따라 과의존 위험군 비율에 차이가 없다.
- H_1 : 연령대에 따라 과의존 위험군 비율에 차이가 있다.

귀무가설 H_0 가 참일 때 구한, 위험군에 속하는 사람 수의 추정된 기대도수와 관측도수 간에 차이가 존재한다. 이 차이의 절댓값이 가장 큰 연령대는?

- ① 10대 이하
- ② 20 ~ 30대
- ③ 40 ~ 50대
- ④ 60대 이상

15. 어느 마트에서 특정 상품에 대한 진열대의 높이와 폭이 매출액에 영향을 주는지 알아보기 위해, 진열대 높이(A)를 상, 중, 하 3수준으로, 폭(B)을 대, 소 2수준으로 하여 실험하였다. 다른 조건들이 동일한 12일 동안 6개의 실험조건을 무작위 순서로 2일씩 적용하여 매출액을 조사한 후 얻은 분산분석표가 다음과 같다. 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	F-값
인자 A	1,544			
인자 B	12			
교호작용 A×B	24			
오차	62			
계				

- ① 총제곱합은 1,642이다.
- ② 인자 B의 평균제곱은 교호작용 A×B의 평균제곱과 동일하다.
- ③ 오차 자유도는 6이다.
- ④ 유의성검정에서의 유의확률은 인자 A가 교호작용 A×B보다 크다.

16. 다음은 자료 (x_{1i}, x_{2i}, y_i) ($i = 1, 2, \dots, 23$)에 다중선형회귀모형 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$ 를 최소제곱법으로 적합하여 얻은 분산분석표이다. 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

요인	제곱합	자유도	평균제곱	F-값	p-값
회귀	420		(가)	(다)	< 0.001
잔차			(나)		
계	480				

ㄱ. 결정계수는 0.875이다.
 ㄴ. 유의수준 1%에서 이 모형이 유의하다.
 ㄷ. (가) + (나) + (다) = 283이다.

- ① ㄱ, ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

17. 다음 설명 중 옳지 않은 것은?

- ① 유의수준 $\alpha = 0.01$ 에서 귀무가설을 기각하는 검정 결과가 나왔다면 유의수준 $\alpha = 0.05$ 에서도 귀무가설을 기각하게 된다.
- ② 모든 자연수 n 에 대해 $t_{0.05}(n) < t_{0.05}(n+1)$ 이다(단, $t_{0.05}(k)$ 는 자유도가 k 인 t분포의 제95백분위수).
- ③ 모든 자연수 n 에 대해 $\chi_{0.05}^2(n) < \chi_{0.05}^2(n+1)$ 이다(단, $\chi_{0.05}^2(k)$ 는 자유도가 k 인 카이제곱분포의 제95백분위수).
- ④ 두 사건 A_1, A_2 가 서로 배반이고 $P(B) > 0$ 일 때, $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$ 이다.

18. 다음은 자료 (x_i, y_i) ($i = 1, 2, \dots, n$)에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 적합하여 구할 수 있는 값들이다.

○ $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	○ $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
○ $e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$	○ $SSE = \sum_{i=1}^n e_i^2$
○ $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$	○ $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
○ $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$	

다음 식 중 옳은 것은? (단, $\hat{\beta}_0, \hat{\beta}_1$ 은 각각 β_0, β_1 의 최소제곱 추정값이다)

- ① $\hat{\beta}_1 = \frac{S_{xy}}{S_{yy}}$
- ② $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$
- ③ $r^2 = \frac{S_{xy}}{S_{xx} S_{yy}}$ (단, r^2 은 결정계수)
- ④ $\sum_{i=1}^n y_i e_i = 0$

19. 두 이산확률변수 X, Y 의 결합확률질량함수가 다음과 같을 때, $P(Y = 2X)$ 의 값은? (단, c 는 상수이다)

$$f(x, y) = \frac{x+y}{c}, \quad x = 1, 2, 3, \quad y = 1, 2, 3, 4$$

- ① $\frac{1}{18}$
- ② $\frac{1}{9}$
- ③ $\frac{1}{6}$
- ④ $\frac{2}{9}$

20. 수준이 3개이고 반복수가 각각 $n_1 = 5, n_2 = 10, n_3 = 15$ 인 일원 배치법으로부터 얻은 자료를 y_{ij} ($i = 1, 2, 3, j = 1, 2, \dots, n_i$)라 하고, i 번째 수준에서의 평균을 $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ 라 하자. $\bar{y}_1 = 6, \bar{y}_2 = 3, \bar{y}_3 = 2$ 일 때, 분산분석표에서 처리제곱합은?

- ① 60
- ② 90
- ③ 270
- ④ 330