

통계학개론

문 1. 상자그림(box plot)으로부터 알 수 없는 통계량은?

- ① 중앙값(median)
- ② 25 % 절사평균(trimmed mean)
- ③ 사분위수 범위(interquartile range)
- ④ 범위(range)

문 2. 휴대전화를 제조하는 세 회사 A, B, C의 시장 점유율은 각각 50 %, 30 %, 20 %로 알려져 있다. 신제품의 출시가 시장 점유율에 영향을 미치는지 알아보기 위하여 C회사가 신제품을 출시하고 6개월 후, 200명의 휴대전화 사용자를 임의로 추출하여 다음과 같이 조사하였다. 시장 점유율의 변화가 있는지 알아보기 위하여 사용하는 검정법으로 옳은 것은?

회사명	A	B	C
관측도수	98	48	54

- ① 자유도가 3인 카이제곱분포를 이용한 독립성 검정
- ② 자유도가 3인 카이제곱분포를 이용한 적합도 검정
- ③ 자유도가 2인 카이제곱분포를 이용한 독립성 검정
- ④ 자유도가 2인 카이제곱분포를 이용한 적합도 검정

문 3. 선형회귀모형에 대한 가정으로 옳지 않은 것은?

- ① 오차의 등분산성
- ② 오차 간의 독립성
- ③ 독립변수와 종속변수의 선형성
- ④ 독립변수 간 상관관계 존재성

문 4. 두 변수의 상관계수(correlation coefficient)에 대한 설명으로 옳은 것은?

- ① 항상 0과 1 사이의 값이다.
- ② 0에 가까울수록 두 변수는 서로 독립이다.
- ③ 공분산과 부호가 같다.
- ④ 두 변수의 인과관계를 설명하는 값이다.

문 5. 서로 독립인 두 확률변수 X 와 Y 의 결합확률분포표가 다음과 같을 때, $P(X=2, Y=1)$ 의 값은?

$X \backslash Y$	1	3
2	a	b
4	0.2	0.4

- ① $\frac{4}{15}$
- ② $\frac{1}{5}$
- ③ $\frac{2}{15}$
- ④ $\frac{1}{15}$

문 6. 가설 $H_0: \mu = 4$ 대 $H_1: \mu = 6$ 에 대한 검정통계량 T 의 분포는 $N(\mu, 0.5^2)$ 을 따르고 검정의 기각역이 $\{T > 5\}$ 이다. 이 검정법의 제1종의 오류와 제2종의 오류를 범할 확률을 옳게 짝지은 것은? (단, 표준정규분포를 따르는 확률변수 Z 에 대하여, $P(Z > 1) = 0.159$, $P(Z > 2) = 0.023$ 이다)

	제1종의 오류를 범할 확률	제2종의 오류를 범할 확률
①	0.023	0.023
②	0.159	0.023
③	0.023	0.159
④	0.159	0.159

문 7. 다음은 여행 프로그램 A, B, C에 대한 선호도가 연령대와 상관성이 있는지 알아보기 위하여 설문조사를 한 결과이다.

	A	B	C
20대	6	27	19
30대	8	36	17
40대	21	45	33
50대	14	18	6

프로그램에 대한 선호도가 연령대와 상관성이 없다는 귀무가설을 검정하기 위한 카이제곱 검정통계량이 14.15일 때, 유의수준 5%에서 검정 결과로 옳은 것은? (단, $\chi^2_\alpha(n)$ 은 자유도가 n 인 카이제곱분포의 제 $100 \times (1 - \alpha)$ 백분위수이고 $\chi^2_{0.05}(6) = 12.59$, $\chi^2_{0.05}(8) = 15.51$ 이다)

- ① 검정통계량의 값이 12.59보다 크므로 프로그램의 선호도가 연령대와 상관성이 없다고 할 수 있다.
- ② 검정통계량의 값이 12.59보다 크므로 프로그램의 선호도가 연령대와 상관성이 없다고 할 수 없다.
- ③ 검정통계량의 값이 15.51보다 작으므로 프로그램의 선호도가 연령대와 상관성이 없다고 할 수 있다.
- ④ 검정통계량의 값이 15.51보다 작으므로 프로그램의 선호도가 연령대와 상관성이 없다고 할 수 없다.

문 8. 어느 지역에 사는 주민의 월 소득은 정규분포 $N(200, 20^2)$ 을 따른다. 이 지역에서 두 명의 주민을 임의로 추출하여 월 소득을 조사할 때 옳지 않은 것은? (단, 표준정규분포를 따르는 확률변수 Z 에 대하여 $P(Z > 1.645) = 0.05$ 이다)

- ① 두 명의 월 소득 평균이 구간 (167.1, 232.9)에 속할 확률은 0.9보다 크다.
- ② 한 명의 월 소득이 167.1보다 작고 다른 한 명의 월 소득이 232.9보다 클 확률은 0.05이다.
- ③ 두 명 중 최소한 한 명의 월 소득이 구간 (167.1, 232.9)에 속할 확률은 0.99이다.
- ④ 두 명 모두 월 소득이 200보다 작을 확률은 0.25이다.

문 9. 다음은 범씨의 네 가지 종류에 따라 비의 평균 수확량에 차이가 있는지 알아보기 위한 분산분석표의 일부이다. 유의수준 5%에서 이 가설을 검정할 때, ㉠, ㉡의 값과 검정 결과를 옳게 짝지은 것은?

요인	제공합	자유도	F-값	p-값
범씨의 종류	6	㉠	㉡	0.0519
오차	4	8		
계	10	11		

- | | | | |
|---|---|---|---------|
| | ㉠ | ㉡ | 검정 결과 |
| ① | 3 | 4 | 유의하지 않다 |
| ② | 3 | 4 | 유의하다 |
| ③ | 4 | 5 | 유의하지 않다 |
| ④ | 4 | 5 | 유의하다 |

문 10. 공기업 기관장의 재임기간은 적어도 10년이라고 주장하는 정부는 이 사실을 입증하기 위하여 25개의 공기업을 조사하였다. 25개 공기업 기관장의 평균 재임기간이 10.5년이라고 할 때, 옳은 것만을 모두 고른 것은? (단, 재임기간의 모표준편차는 3년으로 알려져 있으며, 표준정규분포를 따르는 확률변수 Z 에 대하여, $P(Z > 1.645) = 0.05$, $P(Z > 2.576) = 0.005$ 이다)

- ㄱ. 대립가설은 '공기업 기관장의 재임기간은 10년 이하이다.'이다.
- ㄴ. Z 검정통계량의 값은 $\frac{5}{6}$ 이다.
- ㄷ. 유의수준이 5%일 때 Z 검정통계량과 1.645를 비교한다.
- ㄹ. 기각역의 형태는 $\{\bar{X} < c\}$ 이다.

- ① ㄱ, ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄷ, ㄹ

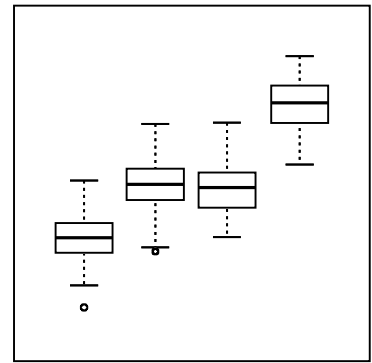
문 11. $P(A) = 0.4$, $P(B) = 0.3$, $P(B|A) = 0.5$ 일 때, $P(A \cup B)$ 의 값은?

- ① 0.45
- ② 0.5
- ③ 0.55
- ④ 0.7

문 12. 확률변수 X 가 정규분포 $N(100, 10^2)$ 을 따른다고 할 때, 옳지 않은 것은?

- ① $(X - 100)/10$ 은 표준정규분포를 따른다.
- ② $(X - 100)$ 의 분산은 100이다.
- ③ $X/100$ 의 표준편차는 1이다.
- ④ $X/10$ 의 분산은 1이다.

문 13. 다음은 4개 시의 단위면적당 주택가격의 자료를 이용하여 그린 상자그림이다. 단위면적당 평균 주택가격이 4개 시에 따라 차이가 있는지 검정하기 위한 분석법은?



- ① 분산분석
- ② 상관분석
- ③ 교차분석
- ④ 시계열분석

문 14. 서로 다른 값을 가지는 세 개 이상의 연속형 자료를 요약할 때 사용하는 기술통계량에 대한 설명으로 옳은 것은?

- ① 자료의 개수가 적을수록 평균과 분산에서 극단값의 영향은 작아진다.
- ② 최댓값이 현재 값보다 큰 값으로 바뀌면 중앙값은 커진다.
- ③ 최댓값이 α 만큼 증가하고, 최솟값이 α 만큼 감소하여도 변동계수(coefficient of variation) 값은 바뀌지 않는다.
- ④ 평균보다 작은 자료의 개수가 평균보다 큰 자료의 개수보다 많으면 중앙값은 평균보다 작다.

문 15. 다음 기댓값 또는 분산 중에서 값이 가장 큰 것은?

- ① 이항분포 $B(20, 0.2)$ 를 따르는 확률변수 X 에 대하여 $Var(X)$ 의 값
- ② 서로 독립이고 분산이 각각 3과 1인 확률변수 X 와 Y 에 대하여 $Var(X - Y)$ 의 값
- ③ 평균 $\frac{1}{2}$ 인 확률변수 X 에 대하여 $E(6X)$ 의 값
- ④ 구간 $[1, 3]$ 에서 정의된 균일(uniform)분포를 따르는 확률변수 X 에 대하여 $E(X)$ 의 값

문 16. 다음 자료에 단순선형회귀모형 $Y_i = \alpha + \beta X_i + \epsilon_i$, ($i = 1, 2, \dots, 7$)을 적용하려고 한다.

i	1	2	3	4	5	6	7
X	1	2	3	5	7	9	8
Y	6	9	10	12	15	13	17

회귀분석 결과가 다음과 같을 때, 옳은 것만을 모두 고른 것은?

	회귀계수	표준오차	t -값	p -값
상수항	6.3695	1.4031	4.539	0.006
x	1.0690	0.2432	4.395	0.007
$R^2 = 0.794$, $R_{adj}^2 = 0.753$, $F = 19.32$ (p -값 = 0.007)				

- ㄱ. 유의수준 5%에서 귀무가설 $H_0 : \beta = 0$ 을 기각한다.
- ㄴ. 결정계수는 75% 이상이다.
- ㄷ. X 와 Y 의 표본상관계수는 0.9보다 크다.
- ㄹ. 유의수준 1%에서 단순선형회귀직선은 통계적으로 유의하다.

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

문 17. X_1, X_2, \dots, X_n 이 정규분포 $N(\mu, \sigma^2)$ 을 따르는 모집단에서의 임의표본(random sample)이고 \bar{X} 를 표본평균, S^2 을 표본분산이라고 할 때, 가설 $H_0 : \mu = \mu_0$ 대 $H_1 : \mu \neq \mu_0$ 의 검정에 대한 설명으로 옳지 않은 것은? (단, 모분산 σ^2 은 알려져 있지 않으며, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 제 $100 \times (1 - \alpha)$ 백분위수이다)

- ① 검정통계량으로 $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ 을 사용한다.
- ② 검정통계량 t 에 대하여 $|t| < t_{\alpha/2}(n-1)$ 일 때 귀무가설을 기각한다.
- ③ 유의수준 5%에서 귀무가설을 기각하지 못하면 μ 의 95% 신뢰구간에 μ_0 가 포함된다.
- ④ 유의확률(p -값)이 유의수준보다 작으면 귀무가설을 기각한다.

문 18. 다음은 일원분산분석(one-way ANOVA)을 수행하여 얻은 분산분석표의 일부이다. 이에 대한 설명으로 옳지 않은 것은?

요인	자유도	F -값	p -값
처리	2	2.78	0.08
오차	27		

- ① 유의수준 1%에서 기각역이 $\{F > c\}$ 일 때 $c < 2.78$ 이다.
- ② 유의수준 5%에서 각 처리의 평균이 모두 동일하다는 귀무가설을 채택한다.
- ③ 자료의 수는 30이다.
- ④ 처리의 수는 3이다.

문 19. 정규분포를 따르는 모집단에서 n 개의 임의표본을 추출하여 모평균 μ 에 대한 추론을 하려고 한다. 옳은 것만을 모두 고른 것은? (단, 모표준편차는 알려져 있는 값이다)

- ㄱ. n 이 일정할 때 μ 에 대한 신뢰구간의 길이는 신뢰수준이 증가할수록 길어진다.
- ㄴ. 오차의 한계가 d 로 주어질 때 μ 를 추정하기 위한 표본의 크기는 신뢰수준이 증가할수록 커진다.
- ㄷ. 가설 $H_0 : \mu = 5$ 대 $H_1 : \mu > 5$ 에서 Z 검정의 유의확률(p -값)은 표본평균의 관측값이 증가할수록 작아진다.
- ㄹ. 가설 $H_0 : \mu = 5$ 대 $H_1 : \mu > 5$ 에서 Z 검정의 검정력은 μ 가 5보다 클수록 증가한다.

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

문 20. 다음은 지난 1년간 두 손해보험 상품 A, B에 대하여 계약자 및 해지자의 수를 조사한 결과이다. 두 손해보험 상품의 계약 해지 비율이 서로 같다고 할 수 있는지 유의수준 5%에서 검정할 때 옳지 않은 것은?

	계약자 수	계약 해지자 수
A	300	60
B	200	40

- ① 두 손해보험 상품에 대한 계약 해지 비율의 추정값은 같다.
- ② 귀무가설이 참일 때 조사 대상 전체 인원에 대해 손해보험 상품의 계약을 해지한 사람의 수는 이항분포를 따른다.
- ③ 두 손해보험 상품의 계약 해지 비율의 추정값에 대한 표준오차는 같다.
- ④ 검정통계량의 값은 0이다.