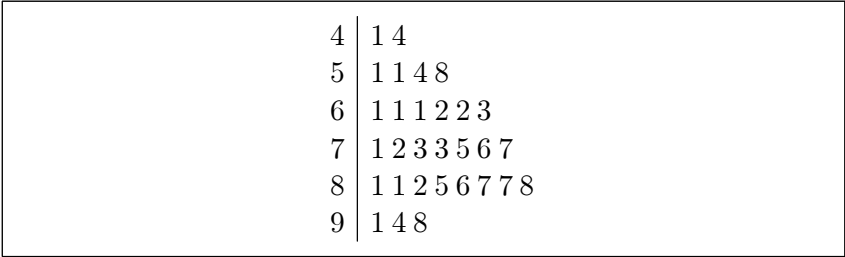


통 계 학

1. 다음은 통계학을 수강하는 학생 30명 성적에 대한 줄기－잎 그림이다. 제1사분위수와 최빈값을 더한 값은?



- ① 122
- ② 124
- ③ 126
- ④ 128

2. 세 사건 A , B , C 에 대해 A 와 B 는 서로 독립, A 와 C 도 서로 독립, B 와 C 는 서로 배반이다. $P(A) = 0.2$, $P(B) = 0.3$, $P(C) = 0.5$ 일 때, $P(A \cup B \cup C)$ 의 값은?

- ① 0.78
- ② 0.84
- ③ 0.90
- ④ 0.96

3. 정규분포 $N(\mu, 1)$ 을 따르는 확률표본 X_1 , X_2 , X_3 을 이용하여 모평균 μ 를 추정하려고 한다. 다음에 제시된 μ 에 대한 추정량 $\hat{\mu}_A$, $\hat{\mu}_B$, $\hat{\mu}_C$ 중 평균제곱오차(MSE)가 작은 것부터 순서대로 바르게 나열한 것은?

$$\hat{\mu}_A = \frac{1}{2}X_1 + \frac{1}{2}X_2$$
$$\hat{\mu}_B = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3$$
$$\hat{\mu}_C = \frac{3}{2}X_1 - \frac{1}{4}X_2 - \frac{1}{4}X_3$$

- ① $\hat{\mu}_A$, $\hat{\mu}_B$, $\hat{\mu}_C$
- ② $\hat{\mu}_A$, $\hat{\mu}_C$, $\hat{\mu}_B$
- ③ $\hat{\mu}_B$, $\hat{\mu}_A$, $\hat{\mu}_C$
- ④ $\hat{\mu}_C$, $\hat{\mu}_A$, $\hat{\mu}_B$

4. 자료 $(x_i, y_i)(i = 1, 2, \dots, 10)$ 에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소제곱법으로 적합한 회귀계수의 추정값은 $\hat{\beta}_0 = 5$, $\hat{\beta}_1 = 2$ 이다.

$\sum_{i=1}^{10} x_i = 120$ 일 때 $\sum_{i=1}^{10} y_i$ 의 값은? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

- ① 190
- ② 240
- ③ 290
- ④ 340

5. 세 가지 백신에 따른 항체 생성률의 차이가 있는지를 확인하기 위해 백신 $k(k = 1, 2, 3)$ 를 접종한 후, 항체가 미생성된 개체 수(O_{k0})와 항체가 생성된 개체 수(O_{k1})를 조사한 결과는 다음과 같다.

항체 생성 여부 백신	미생성	생성	계
백신 1	O_{10}	O_{11}	n_1
백신 2	O_{20}	O_{21}	n_2
백신 3	O_{30}	O_{31}	n_3
계	O_0	O_1	n

백신 k 를 접종한 후 항체가 생성된 개체의 비율을 p_k 라 할 때, 가설 H_0 : ‘ $p_1 = p_2 = p_3$ 이다’ 대 H_1 : ‘ p_1 , p_2 , p_3 중 적어도 하나는 다르다’에 대해 카이제곱검정을 하고자 한다. 귀무가설이 참일 때 백신 k 를 접종한 후 항체가 미생성된 개체의 추정 기대도수를 $\hat{E}_{k0}$ 이라 하고 항체가 생성된 개체의 추정 기대도수를 $\hat{E}_{k1}$ 이라 하자. 이에 대한 설명으로 옳은 것만을 모두 고르면?

- ㉠. $\hat{E}_{10} + \hat{E}_{11} = n_1$ 이다.
- ㉡. $\sum_{k=1}^3 \hat{E}_{k1} = O_1$ 이다.
- ㉢. 카이제곱 검정통계량의 자유도는 2이다.

- ① ㉠
- ② ㉠, ㉢
- ③ ㉡, ㉢
- ④ ㉠, ㉡, ㉢

6. 다음 자료의 중앙값과 산술평균이 각각 50, 55일 때, $x - y$ 의 값은?
(단, $x \geq y$ 이다)

60, 30, 40, 50, 70, 40, x , y

- ① 20
- ② 30
- ③ 40
- ④ 50

7. 어느 대학교는 메일을 스팸 메일과 정상 메일로 분류하는 프로그램을 사용하고 있다. 이 프로그램은 스팸 메일의 99 %를 스팸 메일로 분류하고, 정상 메일의 1 %는 스팸 메일로 분류한다. 이 대학교 전체 메일의 1 %는 스팸 메일이다. 이 프로그램이 어떤 메일을 스팸 메일로 분류했을 때, 해당 메일이 정상 메일일 확률은?

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

8. 볼펜을 제조하는 어느 회사에서는 자사 제품의 판매 비율 p 를 추정하고자 한다. 판매되는 볼펜을 대상으로 임의추출한 n 개 표본 중 자사 제품의 비율을 이용하여 p 를 추정할 때, 95.44 % 신뢰수준에서 p 에 대한 추정량의 오차한계를 0.05 이하로 하는 최소 표본의 크기는?
(단, 확률변수 Z 가 표준정규분포를 따를 때 $P(Z \geq 1) = 0.1587$, $P(Z \geq 2) = 0.0228$ 이다)

- ① 100
- ② 400
- ③ 900
- ④ 1,600

9. 모평균이 μ 인 정규분포를 따르는 모집단에서 임의추출한 10개 자료의 표본평균이 11.5이다. 가설 $H_0: \mu = 10$ 대 $H_1: \mu \neq 10$ 에 대한 검정통계량의 p -값이 0.052일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면?

- ㄱ. 가설 $H_0: \mu = 10$ 대 $H_1: \mu \neq 10$ 에 대하여 유의수준 5 %에서 귀무가설을 기각한다.
- ㄴ. 가설 $H_0: \mu = 10$ 대 $H_1: \mu > 10$ 에 대한 p -값은 0.026이다.
- ㄷ. 모평균 μ 에 대한 95 % 신뢰구간은 10을 포함한다.

- ① ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

10. 자료 $(x_i, y_i)(i = 1, 2, \cdots, 11)$ 에서 표본평균 $\bar{x}$, $\bar{y}$ 는 각각 2, 3이다. 이 자료에 단순선형회귀모형 $y_i = \beta x_i + \epsilon_i$ 를 최소제곱법으로 적합하여 얻은 β 의 추정량을 $\hat{\beta}$ 이라 할 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

- ㄱ. 잔차제곱합의 자유도는 10이다.
- ㄴ. 추정 회귀직선은 (2,3)을 지난다.
- ㄷ. $\hat{\beta}$ 은 β 의 비편향추정량(unbiased estimator)이다.

- ① ㄷ
- ② ㄱ, ㄴ
- ③ ㄱ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

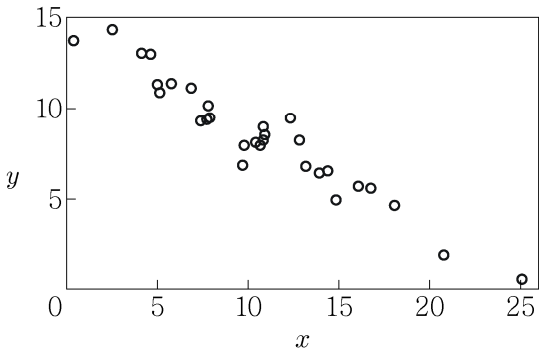
11. 어느 놀이공원은 검은 공과 흰 공이 들어 있는 상자에서 흰 공을 뽑으면 선물을 증정하고 있다. 상자 속의 흰 공의 비율 p 는 $\frac{1}{2}$ 로 알려져 있다. 이 비율이 $\frac{1}{4}$ 로 낮아졌다는 주장을 확인하기 위하여 다음 가설에 대한 검정을 실시하고자 한다.

$$H_0: p = \frac{1}{2} \text{ 대 } H_1: p = \frac{1}{4}$$

상자에서 임의로 복원추출한 4개의 공이 모두 검은색이면 귀무가설을 기각한다고 할 때, 이 검정법의 제1종 오류를 범할 확률과 검정력을 바르게 연결한 것은?

	제1종 오류를 범할 확률	검정력
①	$\frac{1}{32}$	$\frac{1}{256}$
②	$\frac{1}{32}$	$\frac{81}{256}$
③	$\frac{1}{16}$	$\frac{1}{256}$
④	$\frac{1}{16}$	$\frac{81}{256}$

12. 다음과 같은 산점도를 갖는 자료 (x_i, y_i) 에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소제곱법으로 적합하여 얻은 추정 회귀직선이 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ 일 때, 다음 값 중 가장 큰 것은? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)



- ① 추정 회귀직선의 기울기($\hat{\beta}_1$)
- ② x 와 y 의 표본상관계수
- ③ x 와 $\hat{y}$ 의 표본상관계수
- ④ y 와 $\hat{y}$ 의 표본상관계수

13. 어떤 작물의 생산량이 비료(A)의 4가지 수준과 토양(B)의 3가지 수준에 따라 차이가 있는지를 알아보려고 한다. 모든 수준 조합마다 동일한 반복수로 측정하여 얻은 분산분석표의 일부는 다음과 같다. 이에 대한 설명으로 옳은 것만을 모두 고르면?

요인	제곱합	자유도	평균제곱	F -값	p -값
비료(A)	145		48.3	9.67	0.0002
토양(B)	(가)				< 0.0001
교호작용($A \times B$)	35	(나)			0.3564
오차	120	24	5.0		
계	600				

ㄱ. (가)의 값은 300이다.
ㄴ. (나)의 값은 5이다.
ㄷ. 모든 수준 조합에서 반복수는 4이다.

- ① ㄱ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

14. 다음은 자료 $(x_{1i}, x_{2i}, y_i)(i = 1, 2, \dots, 20)$ 에 다중선형회귀모형 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$ 를 최소제곱법으로 적합한 결과이다.

계수	추정 회귀계수	표준오차	t -값	p -값
β_0	-6.053	15.663	-0.386	0.704
β_1	0.955	0.354	2.699	0.015
β_2	0.477	0.294	1.624	0.123

이 회귀모형에 대한 총제곱합(SST)은 980, 회귀제곱합(SSR)은 538일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

ㄱ. 잔차평균제곱(MSE)은 26이다.
ㄴ. 모든 설명변수는 유의수준 5%에서 유의하다.
ㄷ. 동일한 신뢰수준에서 β_1 에 대한 신뢰구간의 폭이 β_2 에 대한 신뢰구간의 폭보다 넓다.

- ① ㄱ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

15. 자격증 시험의 응시자 중 100명을 임의추출하여 온라인 강의 수강과 자격증 시험 합격 여부를 조사한 결과는 다음과 같다.

합격 여부 온라인 강의 수강 여부	합격 여부		
	합격	불합격	계
수강함	40	20	60
수강하지 않음	30	10	40
계	70	30	100

온라인 강의 수강 여부와 자격증 시험 합격 여부는 연관성이 없다는 귀무가설을 유의수준 10%에서 검정하고자 할 때, 카이제곱 검정 통계량의 값과 검정 결과를 바르게 연결한 것은? (단, $\chi^2_\alpha(k)$ 는 자유도가 k 인 카이제곱분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $\chi^2_{0.1}(1) = 2.71$, $\chi^2_{0.1}(3) = 6.25$ 이다)

카이제곱 검정통계량의 값		검정 결과
①	$\frac{50}{63}$	귀무가설을 기각함
②	$\frac{50}{63}$	귀무가설을 기각하지 않음
③	$\frac{200}{63}$	귀무가설을 기각함
④	$\frac{200}{63}$	귀무가설을 기각하지 않음

16. 두 철강 공장 A, B 에서 생산하는 소재의 인장강도는 모분산이 각각 σ_A^2, σ_B^2 인 정규분포를 따른다고 한다. 두 철강 공장 A, B 의 소재를 각각 13개, 10개 임의추출하여 구한 인장강도의 표본분산은 각각 $s_A^2 = 27, s_B^2 = 9$ 이다. 두 공장에서 생산하는 소재의 인장강도가 서로 독립일 때, 두 모분산의 비 $\frac{\sigma_A^2}{\sigma_B^2}$ 에 대한 95 % 신뢰구간은? (단, $F_\alpha(k_1, k_2)$ 는 분자와 분모의 자유도가 각각 k_1, k_2 인 F 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $F_{0.025}(12, 9) = 3.87, F_{0.025}(9, 12) = 3.44$ 이다)

① $\left(\frac{3}{3.87}, 3 \times 3.44\right)$
② $\left(\frac{3}{3.87}, 3 \times 3.87\right)$
③ $\left(\frac{3}{3.44}, 3 \times 3.44\right)$
④ $\left(\frac{3}{3.44}, 3 \times 3.87\right)$

17. 세 그룹 A, B, C 에서 각각 4개, 5개, 6개의 관측값을 조사한 결과 그룹별 표본평균과 표본분산은 다음과 같다. 그룹에 따라 평균의 차이가 있는지를 알아보기 위한 분산분석에서 오차제곱합과 그 자유도를 바르게 연결한 것은?

그룹	자료의 수	표본평균	표본분산
A	4	19	7.0
B	5	10	9.5
C	6	14	8.0

	오차제곱합	자유도
①	99	12
②	99	14
③	180	12
④	180	14

18. 5명의 학생을 대상으로 특별 교육 프로그램 이수 전후의 학업성취도를 측정한 결과는 다음과 같다. 이수 전후 학업성취도의 평균이 같은지를 검정하기 위한 t -검정통계량의 절댓값은? (단, 이수 전후 학업성취도 차이는 서로 독립이고, 정규분포를 따른다)

학생	A	B	C	D	E
이수 전	43	35	32	18	22
이수 후	41	41	42	32	44

- ① 2.0
② 2.5
③ 3.0
④ 3.5

19. 1부터 8까지의 숫자가 각각 하나씩 적혀 있는 8장의 카드가 있다. 이 중에서 3장의 카드를 임의로 비복원추출할 때, 선택된 3장의 카드에 적혀 있는 수 중에서 가장 작은 값을 X 라 하자. X 의 기댓값 $E(X)$ 는?

- ① $\frac{3}{2}$
② $\frac{7}{4}$
③ 2
④ $\frac{9}{4}$

20. 자유도가 10인 t 분포를 따르는 확률변수 T 가 평균으로부터 사분위수 범위의 m 배 이상 떨어져 있는 값을 극단값으로 정의하자. 확률변수 T 가 극단값을 가질 확률이 0.02 이하가 되도록 하기 위한 m 의 최솟값을 m_0 이라 할 때, $14m_0$ 의 값은? (단, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $t_{0.25}(10) = 0.70, t_{0.02}(10) = 2.36, t_{0.01}(10) = 2.76$ 이다)

- ① 23.6
② 27.6
③ 47.2
④ 55.2

21. 두 확률변수 X, Y 의 결합확률밀도함수가 다음과 같을 때, $P\left(X > \frac{1}{2}\right)$ 의 값은?

$$f(x, y) = \begin{cases} 6(1 - x - y), & x > 0, y > 0, x + y < 1 \\ 0, & \text{그 외} \end{cases}$$

- ① $\frac{1}{8}$
② $\frac{1}{6}$
③ $\frac{1}{4}$
④ $\frac{1}{2}$

22. 네 가지 약품 A, B, C, D 에 대해 각각 3회씩 반복하여 측정한 약효는 다음과 같다. 약품에 따라 약효의 차이가 있는지를 검정하기 위한 F -검정통계량의 값은?

약품 A	약품 B	약품 C	약품 D
3	2	0	-1
2	1	-1	-2
1	0	-2	-3

- ① 1
- ② 8
- ③ 10
- ④ 30

23. 자료 $(x_i, y_i)(i = 1, 2, \dots, n)$ 에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소제곱법으로 적합하여 얻은 β_0 의 추정값, 회귀제곱합, 잔차평균 제곱을 각각 $\hat{\beta}_0, SSR, MSE$ 라 하자. 설명변수를 $x_i^* = 2x_i$ 로 변환한 자료 $(x_i^*, y_i^*)(i = 1, 2, \dots, n)$ 에 단순선형회귀모형 $y_i^* = \beta_0^* + \beta_1^* x_i^* + \epsilon_i^*$ 를 최소제곱법으로 적합하여 얻은 β_0^* 의 추정값, 회귀제곱합, 잔차평균 제곱을 각각 $\hat{\beta}_0^*, SSR^*, MSE^*$ 라 하자. 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i, ϵ_i^* 는 각각 $N(0, \sigma^2), N(0, (\sigma^*)^2)$ 을 따르고 서로 독립이다)

㉠. $\hat{\beta}_0 = \hat{\beta}_0^*$

㉡. $SSR = SSR^*$

㉢. $MSE = MSE^*$

- ① ㉠
- ② ㉠, ㉡
- ③ ㉡, ㉢
- ④ ㉠, ㉡, ㉢

24. 어느 공장에서 생산하는 반도체 웨이퍼 1장당 결함 수는 평균이 0.5인 포아송 분포를 따르고, 평균이 $m(m > 0)$ 인 포아송 분포의 확률질량함수는 다음과 같다.

$$f(x) = \begin{cases} \frac{e^{-m} m^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{그 외} \end{cases}$$

웨이퍼당 결함 수가 서로 독립일 때, 갑과 을의 대화에서 (가), (나)에 들어갈 값을 바르게 연결한 것은? (단, 확률변수 Z 가 표준정규분포를 따를 때 $P(Z \geq 1) = 0.1587, P(Z \geq 2) = 0.0228$ 이다)

갑: 포아송 분포의 성질을 이용하면 웨이퍼 5장에서 총 결함이 2개 이상일 확률은 (가)임을 알 수 있어.

을: 나는 50장의 웨이퍼를 고려해 보겠어. 중심극한정리를 이용하면 50장의 웨이퍼에서 총 결함이 30개 이상일 확률은 (나)로 근사할 수 있어.

- | (가) | (나) |
|---------------------|--------|
| ① $1 - e^{-0.5}$ | 0.0228 |
| ② $1 - e^{-0.5}$ | 0.1587 |
| ③ $1 - 3.5e^{-2.5}$ | 0.0228 |
| ④ $1 - 3.5e^{-2.5}$ | 0.1587 |

25. 금융상품 A, B 의 월간 수익률을 각각 확률변수 R_A, R_B 라 하자. 금융상품 A, B 에 분산 투자하였을 때의 수익률 R_p 는 다음과 같다.

$$R_p = aR_A + (1-a)R_B \text{ (단, } 0 \leq a \leq 1)$$

R_A, R_B 의 분산과 공분산이 다음과 같을 때, R_p 의 분산 $Var(R_p)$ 를 최소로 하는 상수 a 의 값은?

$$Var(R_A) = 5, \quad Var(R_B) = 20, \quad Cov(R_A, R_B) = -5$$

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