

## 통계학개론

1. 두 자료 A, B의 줄기-잎 그림에 대한 설명으로 옳지 않은 것은?

A	B
0   9	5   9
1   5 8	6   5 8
2   4 4 4 8 8 8	7   4 4 4 8 8 8
3   2 5	8   2 5
4   1	9   1

- ① 변동계수는 서로 같다.
  - ② 범위는 서로 같다.
  - ③ 사분위수범위는 서로 같다.
  - ④ 표준편차는 서로 같다.
2. 두 확률변수  $X$ 와  $Y$ 의 결합확률분포가 다음과 같을 때, 이에 대한 설명으로 옳은 것은?

	X	0	2
Y	0	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{3}{8}$	$\frac{3}{8}$	

- ①  $E(X) = \frac{3}{4}$
  - ②  $Var(X) = \frac{1}{2}$
  - ③  $E(X+Y) = \frac{3}{2}$
  - ④  $P(X+Y=2) = \frac{1}{2}$
3. 정규분포를 따르는 확률변수  $X$ 에 대하여  $P(X \leq 2) = 0.2$ 이고  $P(X \leq 12) = 0.8$ 일 때,  $P(2 \leq X \leq 7)$ 의 값은?
- ① 0.35
  - ② 0.3
  - ③ 0.25
  - ④ 0.2
4. 두 확률변수  $X$ 와  $Y$ 가 각각 성공의 확률이 0.3과 0.7인 베르누이 분포를 따르고 서로 독립일 때,  $P(X \neq Y)$ 의 값은?
- ① 0.09
  - ② 0.21
  - ③ 0.49
  - ④ 0.58

5. 어느 회사에서 경력에 따라 연봉에 차이가 있는지 알아보기 위해 1년 미만, 1년 이상 3년 미만, 3년 이상의 경력을 가진 직원을 각각 100명씩 임의추출하여 연봉(단위: 만 원)을 조사한 결과 다음과 같은 분할표를 얻었다.

	연봉	3,000 미만	3,000 이상 4,000 미만	4,000 이상 5,000 미만	5,000 이상
경력	1년 미만	30	50	20	0
1년 이상 3년 미만	20	30	40	10	
3년 이상	10	10	30	50	

- “경력에 따른 연봉에 차이는 없다.”라는 가정하에서 구한 기대도수에 대한 설명으로 옳지 않은 것은?
- ① 기대도수와 관측도수 차의 절댓값의 최솟값은 0이다.
  - ② 기대도수와 관측도수 차의 절댓값의 최댓값은 30이다.
  - ③ 기대도수의 최솟값은 10이다.
  - ④ 기대도수의 최댓값은 30이다.

6. 정규분포를 따르는 두 모집단  $N(\mu_X, \sigma_X^2)$ 과  $N(\mu_Y, \sigma_Y^2)$ 에서 각각 독립표본  $X_1, X_2, \dots, X_5$ 와  $Y_1, Y_2, \dots, Y_6$ 을 임의추출하여 구한 표본 분산을 순서대로 24와 9라고 하자. 확률변수  $W$ 는 분자와 분모의 자유도가 각각 4와 5인  $F$ 분포를 따를 때, 가설  $H_0: \frac{\sigma_X^2}{\sigma_Y^2} = 1$  대

$H_1: \frac{\sigma_X^2}{\sigma_Y^2} > 1$ 에 대한 검정에서 유의확률과 같은 것은?

- ①  $P\left(W \geq \frac{3}{8}\right)$
- ②  $P\left(W \geq \frac{8}{3}\right)$
- ③  $P\left(W \geq \frac{15}{32}\right)$
- ④  $P\left(W \geq \frac{32}{15}\right)$

7. 고혈압 환자 20명 중에서 임의추출한 10명으로 이루어진 집단 A에는 신약을 처방하고 나머지 10명으로 이루어진 집단 B에는 기존 약을 처방한 뒤 4주 후에 혈압을 측정하였다. 다음은 두 집단의 혈압이 각각 동일한 분산을 갖는 정규분포를 따른다고 가정하고 분석한 결과이다.

공통분산의 추정량의 자유도는  $k$ 이고  $t_{0.025}(k) = 2.1$ 이라고 할 때, 두 집단의 평균 혈압 차에 대한 95% 신뢰구간은  $(-12.2, -3.8)$ 이다.

공통분산의 추정량을  $S_p^2$ 이라고 할 때,  $k$ 와  $S_p^2$ 의 값을 바르게 연결한 것은? (단,  $t_\alpha(k)$ 는 자유도가  $k$ 인  $t$ 분포의 제  $100 \times (1 - \alpha)$  백분위수이다)

- | $k$  | $S_p^2$ |
|------|---------|
| ① 9  | 10      |
| ② 9  | 20      |
| ③ 18 | 10      |
| ④ 18 | 20      |

8. 자료  $(x_i, y_i)$  ( $i = 1, 2, \dots, 6$ )에 단순선형회귀모형  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소제곱법으로 적합하여 다음 값을 얻었다.

$$\sum_{i=1}^6 (y_i - \hat{y}_i)^2 = 150, \quad \sum_{i=1}^6 (y_i - \bar{y})^2 = 270, \quad \sum_{i=1}^6 (x_i - \bar{x})^2 = 30$$

$x$ 와  $y$ 의 표본상관계수가 음수일 때,  $\beta_1$ 의 추정값은? (단,  $\hat{y}_i$ 는  $y_i$ 의 적합값이다)

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

9. 다음 표는 세 학생 집단 A, B, C의 시험성적을 비교하기 위해 각 집단에서 4명씩 임의추출하여 작성한 분산분석표의 일부이다.

요인	제곱합	자유도	평균제곱	F-값
처리	(가)			
오차		(나)		
계	(다)	(라)		

집단 A와 B의 시험성적의 모평균이 동일하다고 가정하고 하나의 집단으로 묶어서 집단 C와 비교하기 위한 분산분석표를 다시 작성한다고 할 때, (가) ~ (라) 중에서 변하지 않는 것만을 모두 고르면?

- ① (가), (나)
- ② (가), (다)
- ③ (나), (라)
- ④ (다), (라)

10. 어느 영업사원에게 걸려오는 전화통화 수는 한 시간에 평균 6인 포아송분포를 따른다. 두 확률변수  $U$ 와  $V$ 는 각각 평균이 2와 6인 포아송분포를 따른다고 할 때, 오전 9시 이후 이 영업사원에게 처음 걸려오는 전화가 같은 날 오전 9시 20분 이후일 확률과 같은 것은?

- ①  $P(U=0)$
- ②  $P(U=1)$
- ③  $P(V=0)$
- ④  $P(V=1)$

11. 자료의 퍼진 정도를 나타내는 측도 중에서 원자료와 같은 측정 단위를 갖는 것만을 모두 고르면?

- ㄱ. 표준편차
- ㄴ. 사분위수범위
- ㄷ. 분산
- ㄹ. 변동계수

- ① ㄱ, ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄹ
- ④ ㄷ, ㄹ

12. 동전 한 개를 연속해서 10번 던질 때, 처음 6번의 시행에서는 뒷면이 2번 나오고 나머지 4번의 시행에서는 앞면이 2번 나올 확률은?

- ①  $\frac{15}{2^9}$
- ②  $\frac{15}{2^{10}}$
- ③  $\frac{45}{2^9}$
- ④  $\frac{45}{2^{10}}$

13. 다음 표는 단순선형회귀모형  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  ( $i = 1, 2, \dots, 26$ )에 대한 분산분석표의 일부이다. 이 모형의 결정계수의 값은? (단,  $\epsilon_i$ 는  $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

요인	제곱합	자유도	평균제곱	F-값
회귀				5
잔차	120			
계		25		

- ①  $\frac{4}{29}$
- ②  $\frac{5}{29}$
- ③  $\frac{6}{29}$
- ④  $\frac{7}{29}$

14. 두 확률변수  $X$ 와  $Y$ 가 이변량 정규분포를 따르고 상관계수가  $\rho$ 일 때, 가설  $H_0: \rho=0$  대  $H_1: \rho>0$ 을 검정하고자 한다. 표본의 크기가  $n$ 이고 표본상관계수가  $r$ 일 때, 위 가설을 검정하기 위한 검정통계량

$$T = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}}$$

는  $H_0$ 하에서 자유도가  $(n-2)$ 인  $t$ 분포를 따른다고 한다.  $n=11$ 이고  $r=0.6$ 일 때, 유의수준 5%에서

검정 방법에 대한 설명으로 옳은 것은? (단,  $t_\alpha(k)$ 는 자유도가  $k$ 인  $t$ 분포의 제  $100 \times (1-\alpha)$  백분위수를 나타낼 때,  $t_{0.025}(9) = 2.26$ ,  $t_{0.05}(9) = 1.83$ 이다)

- ①  $T$ 의 값이  $t_{0.025}(9)$ 보다 커서 귀무가설을 기각한다.
- ②  $T$ 의 값이  $t_{0.025}(9)$ 보다 작아서 귀무가설을 기각하지 못한다.
- ③  $T$ 의 값이  $t_{0.05}(9)$ 보다 커서 귀무가설을 기각한다.
- ④  $T$ 의 값이  $t_{0.05}(9)$ 보다 작아서 귀무가설을 기각하지 못한다.

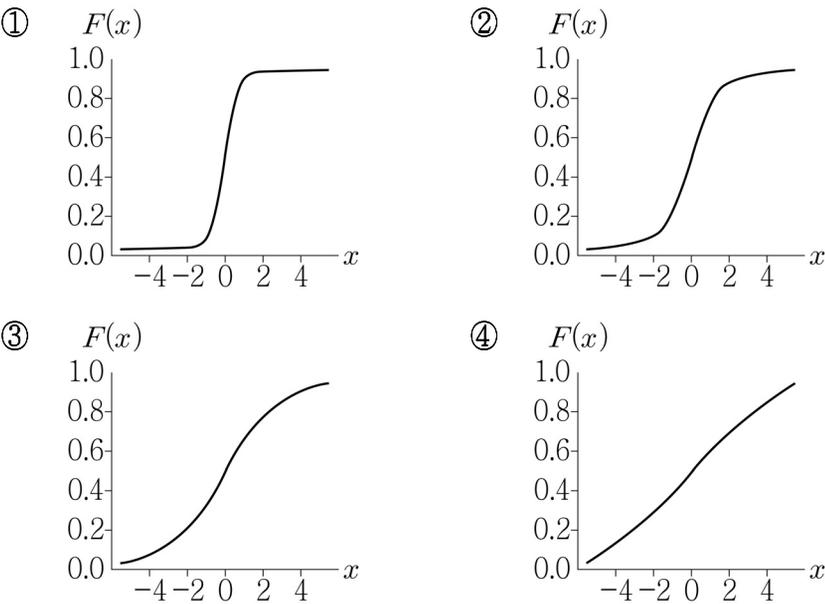
15. 수준 수가 4인 인자 A와 수준 수가 3인 인자 B의 각 수준조합에서 3회씩 반복 실험을 하였다. 다음 표는 인자 A와 B의 교호작용(interaction)이 유의하지 않아서 교호작용을 오차항에 포함한 후 작성한 분산분석표의 일부이다.

요인	제곱합	자유도	평균제곱	F-값
인자 A	54			
인자 B	72			
오차				
계	396			

유의수준 5%에서 검정할 때, 인자 A와 B의 유의성에 대한 설명으로 옳은 것은? (단,  $F_{\alpha}(k_1, k_2)$ 는 분자와 분모의 자유도가 각각  $k_1, k_2$ 인 F분포의 제 $100 \times (1 - \alpha)$  백분위수를 나타낼 때,  $F_{0.05}(2, 30) = 3.32$ ,  $F_{0.05}(3, 30) = 2.92$ 이다)

- ① 인자 A, B는 모두 유의하다.
- ② 인자 A는 유의하고, 인자 B는 유의하지 않다.
- ③ 인자 A는 유의하지 않고, 인자 B는 유의하다.
- ④ 인자 A, B는 모두 유의하지 않다.

16. 다음은 평균은 같고 분산이 서로 다른 정규분포의 누적분포함수  $F(x)$  ( $-\infty < x < \infty$ )의 그림이다. 분산이 가장 큰 것은?



17. 자료  $(x_i, y_i)$  ( $i = 1, 2, \dots, 10$ )로부터 다음 결과를 얻었을 때, 이 자료를 최소제곱법으로 적합한 단순선형회귀식은?

$$\bar{x} = 2, \bar{y} = 3, r = \frac{1}{3}, \sum_{i=1}^{10} (x_i - \bar{x})^2 = 1, \sum_{i=1}^{10} (y_i - \bar{y})^2 = 36$$

(단,  $r$ 는  $x$ 와  $y$ 의 표본상관계수이다)

- ①  $\hat{y} = -1 + 2x$
- ②  $\hat{y} = -2 + x$
- ③  $\hat{y} = -4 + 2x$
- ④  $\hat{y} = -4 + x$

18. 어느 사과 농장에서 수확한 사과의 평균 무게  $\mu$ 에 대한 가설  $H_0: \mu = 350$  대  $H_1: \mu \neq 350$ 을 검정하려고 한다. 이 농장에서 수확한 사과 49개를 임의추출하여 구한  $\mu$ 에 대한 95% 신뢰구간이 (349, 361)일 때, 이에 대한 설명으로 옳지 않은 것은?

- ① 유의수준 5%에서 검정할 때  $H_0$ 를 기각할 수 없다.
- ② 유의수준 2%에서 검정할 때  $H_0$ 를 기각할 수 없다.
- ③ 위 가설에 대한 유의확률은 0.01보다 크다.
- ④ 위 가설에 대한 유의확률은 0.03보다 작다.

19. 알코올 중독자와 비중독자 간에 질병 A의 유병률이 차이가 있는지 알아보기 위해 어느 지역의 알코올 중독자와 비중독자 중에서 각각 76명과 109명을 임의추출하여 질병 A의 유무에 대한 다음과 같은 분할표를 얻었다.

	중독 여부	중독자	비중독자	계
질병 A의 유무	있음	53	28	81
	없음	23	81	104
계		76	109	185

분석 목적과 표집 과정을 고려할 때, 이 지역에 대한 설명으로 옳은 것은?

- ① 알코올 중독자의 비율은  $\frac{76}{185}$ 으로 추정된다.
- ② 질병 A의 유병률은  $\frac{81}{185}$ 로 추정된다.
- ③ 알코올 중독자 중에서 질병 A의 유병률은  $\frac{53}{76}$ 으로 추정된다.
- ④ 질병 A의 유병자 중에서 알코올 중독자의 비율은  $\frac{53}{81}$ 으로 추정된다.

20. 어느 지역 남성의 흡연율  $p$ 를 추정하기 위해 이 지역 남성  $n$ 명을 임의추출하여 구한 흡연자의 표본비율을  $\hat{p}$ 라고 하자. 정규근사를 이용할 때, 모든  $p$ 에 대하여 부등식  $P(|\hat{p} - p| \leq 0.098) \geq 0.95$ 를 만족하기 위한  $n$ 의 최솟값은? (단, 표준정규분포를 따르는 확률변수  $Z$ 에 대하여  $P(Z \geq 1.96) = 0.025$ 이다)

- ① 64
- ② 100
- ③ 400
- ④ 625